



Implementasi Metode Cosine Similarity Dalam Sistem Profiling Dosen Berbasis Data Bibliometrik Untuk Pemetaan Kompetensi Akademik

Jefry Sunupurwa Asri^{1*}, Firnanda Amalia², Muhammad Thifaa Dzaki², Muhammad Fikri¹, Ardra Rianisa²

¹ Fakultas Ilmu Komputer, Teknik Informatika, Universitas Esa Unggul, Jakarta, Indonesia

² Fakultas Ilmu Komputer, Sistem Informasi, Universitas Esa Unggul, Jakarta, Indonesia

Email: ¹jefry.sunupurwa@esaunggul.ac.id, ²firmandaamalia05@student.esaunggul.ac.id,

³muhhammadthifaaldzaki@student.esaunggul.ac.id, ⁴muhhammadfikrii77xc@student.esaunggul.ac.id,

⁵rianisaardra49@student.esaunggul.ac.id

Email Penulis Korespondensi: jefry.sunupurwa@esaunggul.ac.id

Abstrak—Profiling dosen berbasis publikasi ilmiah menjadi komponen strategis dalam pengelolaan sumber daya manusia perguruan tinggi. Proses identifikasi kompetensi dosen yang dilakukan secara manual sering memerlukan waktu lama dan berpotensi menimbulkan ketidaktepatan. Penelitian ini bertujuan mengembangkan aplikasi otomatis untuk profiling dosen dan pemetaan kompetensi terhadap mata kuliah menggunakan pendekatan *unsupervised text similarity* berbasis metode Term Frequency–Inverse Document Frequency (TF-IDF) dan *Cosine Similarity*. Aplikasi dikembangkan menggunakan framework Streamlit dengan integrasi data dari Google Scholar, SINTA, dan Scopus. Evaluasi dilakukan terhadap 50 dosen 120 pasangan data dosen–mata kuliah menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, waktu respons, dan *usability*. Hasil pengujian menunjukkan akurasi sebesar 85,3%, *F1-score* 0,853, waktu respons rata-rata 2,3 detik, serta skor *usability* 86,4, yang termasuk kategori *excellent*. Sistem mampu menampilkan profil dosen secara interaktif, melakukan pemetaan kompetensi terhadap mata kuliah, serta menghasilkan laporan otomatis dalam format PDF. Dengan demikian, aplikasi ini efektif mendukung proses pengambilan keputusan akademik berbasis data dalam penempatan dosen sesuai bidang keahlian mereka.

Kata Kunci: Profiling Dosen; Streamlit; Cosine Similarity; TF-IDF; Pemetaan Kompetensi

Abstract—Lecturer profiling based on scientific publications is a strategic component in managing human resources in higher education institutions. The manual process of identifying lecturer competencies often requires considerable time and may lead to inaccuracies. This study aims to develop an automated application for lecturer profiling and competency mapping to relevant courses using an unsupervised text similarity approach based on the Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity methods. The application was developed using the Streamlit framework with integrated data from Google Scholar, SINTA, and Scopus. The evaluation involved 50 lecturers and 120 lecturer–course pairs, measured using accuracy, precision, recall, F1-score, response time, and usability metrics. The results show an accuracy of 85.3%, an F1-score of 0.853, an average response time of 2.3 seconds, and a usability score of 86.4, which falls into the *excellent* category. The system is capable of displaying interactive lecturer profiles, performing competency mapping to relevant courses, and generating automatic reports in PDF format. Therefore, this application effectively supports data-driven academic decision-making processes for assigning lecturers according to their areas of expertise.

Keywords: Lecturer Profiling, Streamlit, Cosine Similarity, TF-IDF, Competency Mapping

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam pengelolaan sumber daya manusia di perguruan tinggi, khususnya dalam bidang akademik dan penelitian. salah satu aspek krusial manajemen sumber daya manusia adalah profil dosen, yaitu proses identifikasi kompetensi dan bidang keahlian berdasarkan rekam jejak akademik dan publikasi ilmiah. Profil dosen yang akurat sangat penting untuk memastikan kesesuaian antara kompetensi dosen dengan mata kuliah yang diajarkannya, sehingga tercapai pembelajaran yang optimal dan keselarasan dengan bidang keahlian dosen [1].

Namun, pada praktiknya, proses identifikasi kompetensi dosen di banyak perguruan tinggi masih dilakukan secara manual, melalui wawancara, kuesioner, atau penelusuran data publikasi secara terpisah di berbagai platform seperti Google Scholar, SINTA, dan Scopus. Proses manual tersebut tidak hanya memakan waktu lama, tetapi juga berpotensi menimbulkan ketidaktepatan dalam pemetaan bidang keahlian. Hal ini disebabkan oleh data publikasi yang tersebar di beberapa platform dengan struktur dan format data yang berbeda[2].

Fokus utama penelitian ini tidak hanya pada pengelolaan data akademik secara umum, tetapi lebih spesifik pada bagaimana publikasi ilmiah dosen dapat dimanfaatkan untuk memetakan kompetensi dan kesesuaiannya dengan mata kuliah yang diajarkan. Beberapa penelitian sebelumnya telah mengembangkan sistem profil dosen berbasis publikasi ilmiah, tetapi sebagian besar terbatas pada satu sumber data dan belum menghubungkan hasilnya dengan pemetaan kompetensi mata kuliah. Penelitian ini berupaya mengatasi kesenjangan tersebut dengan mengintegrasikan tiga sumber data utama, yaitu Google Scholar, SINTA, dan Scopus, ke dalam satu sistem terpusat, serta menerapkan metode TF-IDF dan cosine similarity untuk menganalisis relevansi antara bidang penelitian dosen dan deskripsi mata kuliah

Beberapa penelitian sebelumnya telah mengembangkan sistem profiling dosen berbasis publikasi ilmiah, namun sebagian besar masih terbatas pada satu sumber data. Penelitian oleh Park et al. [3] memanfaatkan data publikasi untuk menampilkan profil akademik dosen menggunakan pendekatan cosine similarity. Penelitian lain oleh Aini dan Yulianti [4] menerapkan metode Adjusted TF-IDF untuk mengelompokkan bidang keahlian dosen, tetapi belum mengintegrasikan hasilnya dengan pemetaan kesesuaian terhadap mata kuliah. Selain itu, Rinjeni et al. [5] menerapkan metode TF-IDF dan cosine similarity untuk pencocokan judul artikel ilmiah, namun tanpa integrasi lintas sumber data akademik.



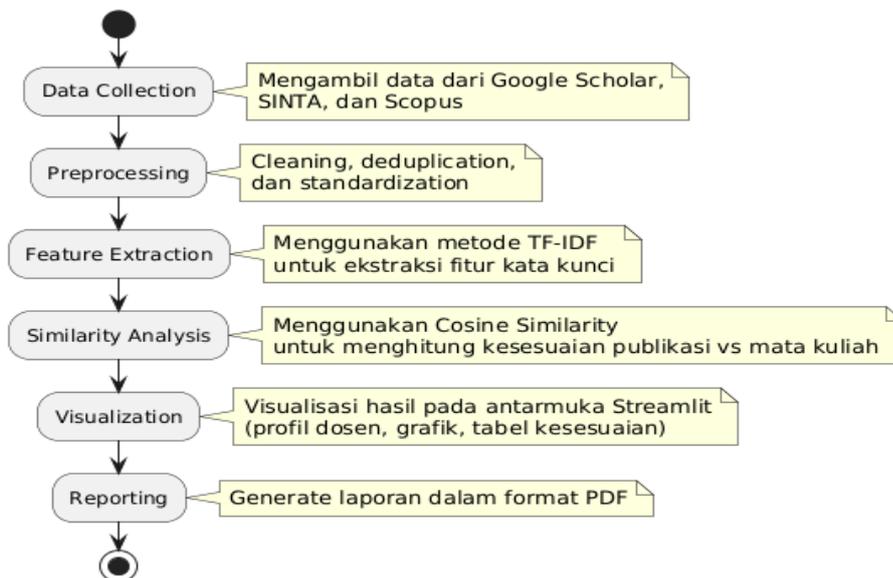
Selain itu, Studi lain menekankan efektivitas metode berbasis TF-IDF dan algoritma kecerdasan buatan dalam identifikasi kompetensi dan pakar di berbagai domain. Fonseca et al. [6] menunjukkan pentingnya penerapan algoritma TF-IDF untuk meningkatkan akurasi identifikasi bidang keahlian peneliti, sementara Borna et al. [7] menyoroti efektivitas algoritma AI dalam proses identifikasi pakar di domain medis. Suryanto dan Khan [8] membuktikan efektivitas varian TF-IDF dalam klasifikasi teks ilmiah, dan Lu, Wang, & Chen [9] membandingkan berbagai model pemrosesan teks, seperti TF-IDF, Word2Vec, FastText, BERT, dan GPT, untuk analisis topik penelitian menunjukkan bahwa TF-IDF masih relevan untuk pemetaan berbasis teks. Lebih lanjut, penelitian terbaru oleh Sabilillah et al. [10], Yasni et al. [11], dan Hairani [12] mengonfirmasi bahwa integrasi multi-platform berbasis bibliometrik dapat meningkatkan akurasi *profiling* akademik hingga 20%, sehingga menegaskan pentingnya pendekatan komprehensif dalam analisis kompetensi dosen.

Berdasarkan kajian terhadap penelitian terdahulu, dapat diidentifikasi bahwa belum ada penelitian yang mengintegrasikan tiga sumber data utama Google Scholar, SINTA, dan Scopus ke dalam satu sistem *profiling* terpusat yang terhubung langsung dengan proses pemetaan kesesuaian mata kuliah. Dengan demikian, masih terdapat *research gap* dalam hal integrasi data multi-platform dan penerapan metode berbasis *cosine similarity* untuk menentukan relevansi antara bidang penelitian dosen dan deskripsi mata kuliah [9]–[12].

Untuk menjawab kesenjangan tersebut, penelitian ini bertujuan mengembangkan aplikasi *profiling* dosen berbasis Streamlit dengan integrasi data dari Google Scholar, SINTA, dan Scopus. Sistem ini menggunakan metode TF-IDF dan *Cosine Similarity* untuk menganalisis kesesuaian antara kompetensi dosen dan mata kuliah, serta mengevaluasi performa sistem berdasarkan akurasi, waktu respons, dan tingkat *usability*. Diharapkan, hasil penelitian ini dapat memberikan kontribusi nyata dalam penyediaan solusi otomatis untuk pemetaan kompetensi dosen berbasis data, sehingga dapat mendukung proses pengambilan keputusan akademik yang lebih objektif dan efisien di perguruan tinggi.

2. METODOLOGI PENELITIAN

2.1 Bagan Alir Penelitian



Gambar 1. Bagan Alir Penelitian

Gambar 1 menunjukkan bagan alir penelitian yang terdiri atas enam tahapan utama, yaitu Data Collection, Preprocessing, Feature Extraction, Similarity Analysis, Visualization, dan Reporting. Setiap tahapan memiliki peran penting dalam membangun sistem *profiling* dosen berbasis publikasi ilmiah yang terintegrasi dengan Google Scholar, SINTA, dan Scopus.

a. Data Collection

Tahapan ini merupakan proses awal penelitian yang berfokus pada pengumpulan data publikasi dosen dari tiga sumber utama, yaitu Google Scholar, SINTA, dan Scopus. Data yang dikumpulkan meliputi judul publikasi, abstrak, kata kunci, serta informasi afiliasi penulis. Pengambilan data dilakukan secara otomatis menggunakan *application programming interface (API)* dan *web scraping* sesuai kebijakan masing-masing platform. Tahapan ini bertujuan untuk menyediakan kumpulan data yang representatif dan relevan dengan bidang keilmuan dosen.

b. Preprocessing

Setelah data terkumpul, dilakukan tahap *preprocessing* untuk memastikan kualitas dan konsistensi data. Proses ini meliputi pembersihan (*data cleaning*), penghapusan duplikasi (*deduplication*), standarisasi format teks (*text normalization*), serta penghapusan karakter khusus atau kata-kata yang tidak bermakna (*stopwords*). Hasil dari tahap



ini berupa dataset yang siap dianalisis secara komputasional.

c. Feature Extraction

Tahap ini menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengekstraksi fitur kata kunci dari kumpulan teks publikasi dan deskripsi mata kuliah. TF-IDF berfungsi untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya di dalam dokumen dan relevansinya terhadap seluruh korpus data. Dengan metode ini, sistem dapat mengenali kata-kata penting yang mencerminkan bidang keilmuan dosen dan topik mata kuliah.

d. Similarity Analysis

Pada tahap ini dilakukan analisis kesesuaian antara bidang penelitian dosen dengan deskripsi mata kuliah menggunakan metode cosine similarity. Teknik ini mengukur tingkat kedekatan dua vektor teks berdasarkan sudut kosinus di antara keduanya. Semakin tinggi nilai kesamaan (mendekati 1), semakin relevan bidang penelitian dosen dengan mata kuliah yang dianalisis. Tahap ini menjadi inti dari sistem pemetaan kompetensi dosen.

e. Visualization

Hasil analisis kesesuaian divisualisasikan secara interaktif melalui antarmuka Streamlit. Modul ini menampilkan informasi profil dosen, bidang keilmuan dominan, tingkat kesesuaian dengan setiap mata kuliah, serta grafik hubungan antar topik. Pendekatan visual ini memudahkan pihak fakultas atau program studi dalam memahami hasil analisis secara intuitif dan cepat.

f. Reporting

Tahap terakhir adalah pembuatan laporan hasil analisis dalam format PDF. Laporan mencakup ringkasan profil dosen, bidang keilmuan utama, serta rekomendasi mata kuliah yang sesuai dengan kompetensi dosen. Laporan ini dapat dijadikan dokumen pendukung dalam proses pengambilan keputusan akademik, seperti penugasan mengajar, perencanaan kurikulum, atau evaluasi kinerja dosen.

Secara keseluruhan, enam tahapan tersebut membentuk kerangka penelitian yang sistematis dalam mengembangkan sistem *profiling* dosen berbasis integrasi multi-platform dan analisis kesesuaian berbasis teks.

2.2 Pengumpulan Data

Proses pengumpulan data dilakukan melalui tiga metode utama sesuai dengan sumber data yang digunakan. Untuk data Google Scholar, digunakan teknik *web scraping* dengan *library BeautifulSoup4* dan *Selenium WebDriver*. Implementasi scraping dirancang dengan memperhatikan *robots.txt* dan *terms of service* dari masing-masing platform. Data yang diambil meliputi profil dasar (nama, afiliasi, bidang minat), metrik penelitian (*h-index*, *i10-index*, *citations*), dan daftar publikasi (judul, tahun, journal, abstrak). Untuk data SINTA, digunakan API resmi yang disediakan oleh Kemdikbudristek dengan *authentication* token. Data yang diperoleh meliputi score SINTA, jumlah publikasi terindeks, dan data kinerja penelitian lainnya. Sedangkan untuk data Scopus, digunakan API dari Elsevier dengan *subscription* yang *valid*. Proses autentikasi dilakukan menggunakan *API key* dan *instution* token untuk mengakses data yang lebih komprehensif.

2.3 Preprocessing Data

Tahapan *preprocessing data* merupakan langkah penting dalam penelitian berbasis *text mining* yang bertujuan untuk menyiapkan dan menstandarkan data sebelum dilakukan proses analisis lebih lanjut. Fokus penelitian ini adalah analisis kesesuaian bidang keilmuan antara publikasi dosen dan deskripsi mata kuliah, sehingga proses *preprocessing* difokuskan pada pembersihan serta normalisasi teks agar hasil representasi fitur lebih akurat pada tahap ekstraksi.

Proses *preprocessing* teks meliputi beberapa tahapan utama. Tahap pertama adalah *case folding*, yaitu proses konversi seluruh karakter menjadi huruf kecil (*lowercase*) untuk menghindari perbedaan makna akibat variasi kapitalisasi. Selanjutnya dilakukan *tokenizing*, yaitu pemecahan teks menjadi satuan kata atau token menggunakan pustaka *Natural Language Toolkit* (NLTK). Tahap berikutnya adalah *filtering* atau penghapusan *stop words*, yakni kata-kata umum yang tidak memiliki kontribusi semantik signifikan terhadap konteks analisis. Daftar *stop words* mencakup bahasa Indonesia dan bahasa Inggris, serta diperluas dengan istilah akademik yang sering digunakan. Setelah itu dilakukan *stemming*, yaitu proses mengembalikan setiap kata ke bentuk dasarnya menggunakan algoritma *Porter Stemmer* untuk teks berbahasa Inggris dan *Nazief–Adriani Stemmer* untuk teks berbahasa Indonesia.

Adapun data numerik seperti *h-index*, jumlah sitasi, dan skor SINTA hanya melalui proses validasi dasar untuk memastikan konsistensi serta integritas hasil integrasi data dari berbagai sumber. Proses tersebut tidak digunakan dalam analisis utama karena fokus penelitian terletak pada pengolahan data teks untuk pemetaan kompetensi. Dengan demikian, *handling missing values* maupun *outlier detection* terhadap data numerik dilakukan semata-mata untuk verifikasi kualitas dataset, bukan untuk tahap analisis berbasis *text mining*.

Seluruh tahapan *preprocessing* diimplementasikan menggunakan pustaka *pandas*, *numpy*, dan *NLTK* dalam lingkungan pemrograman Python guna menjamin efisiensi dan reproduktibilitas proses.

2.4 Ekstraksi Fitur dan Analisis Kesamaan

Analisis data menggunakan pendekatan Text Mining dengan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk *feature extraction*. TF-IDF mampu mengukur importance suatu term dalam dokumen relatif terhadap kumpulan dokumen. Perhitungan TF-IDF menggunakan formula:



$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \quad (1)$$

dimana $\text{TF}(t, d)$ adalah *term frequency* dalam dokumen, dan $\text{IDF}(t)$ adalah *inverse document frequency*.

Algoritma Cosine Similarity digunakan untuk mengukur kemiripan antara vektor representasi bidang penelitian dosen dengan deskripsi mata kuliah. Cosine similarity efektif untuk mengukur kesamaan orientasi antara dua vektor tanpa memperhatikan besarnya. Perhitungan dilakukan menggunakan formula:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2)$$

Implementasi menggunakan library scikit-learn dengan optimasi parameter untuk domain akademik.

2.5 Implementasi Sistem

Sistem diimplementasikan menggunakan Python 3.9 dan framework Streamlit untuk antarmuka pengguna. Arsitektur sistem dibangun secara modular dengan prinsip *separation of concerns*, yang memisahkan pengelolaan data, logika bisnis, dan tampilan hasil analisis. Setiap modul dalam sistem memiliki tanggung jawab yang jelas, meliputi pengumpulan data, pemrosesan teks, visualisasi hasil, dan pelaporan. Pendekatan ini memastikan alur kerja sistem tetap terstruktur, mudah dikembangkan, dan efisien dalam proses pemetaan kompetensi dosen. Database menggunakan SQLite untuk development dan PostgreSQL untuk production environment. Modul utama sistem terdiri dari: data collector module untuk pengambilan data dari berbagai sumber, processing module untuk analisis teks dan perhitungan similarity, visualization module untuk menampilkan hasil dalam bentuk chart dan tabel, serta reporting module untuk generate laporan PDF. Implementasi docker containerization memungkinkan deployment yang konsisten across different environments.

Tabel 1. Spesifikasi Teknis Sistem

Komponen	Teknologi	Versi
Backend	Python	3.9
Frontend	Streamlit	1.28
Database	SQLite/PostgreSQL	3.38/14
Container	Docker	24.0

2.6 Evaluasi Sistem

Evaluasi sistem dilakukan untuk menilai kinerja dan efektivitas aplikasi *profiling* dosen dalam melakukan pemetaan kompetensi terhadap mata kuliah. Pengujian dilakukan terhadap 50 dosen yang berasal dari berbagai disiplin ilmu, dengan tujuan mengukur akurasi sistem, waktu respons, serta tingkat *usability* dari antarmuka pengguna.

Pengujian akurasi pemetaan dilakukan dengan membandingkan hasil keluaran sistem dengan hasil penilaian manual oleh tiga ahli bidang yang memiliki kompetensi dalam analisis bidang keilmuan dosen. Proses ini bertujuan untuk memverifikasi tingkat kesesuaian antara hasil otomatis sistem dengan evaluasi berbasis *expert judgment*. Nilai akurasi dihitung menggunakan metrik *precision*, *recall*, dan *F1-score* untuk menggambarkan performa sistem dalam mengidentifikasi relevansi antara bidang penelitian dan deskripsi mata kuliah.

Aspek *usability* sistem diukur menggunakan instrumen System Usability Scale (SUS) yang diisi oleh 10 pengguna, terdiri atas dosen dan staf akademik yang berinteraksi langsung dengan aplikasi. Skor SUS digunakan untuk menilai kemudahan penggunaan, efisiensi antarmuka, serta kepuasan pengguna terhadap pengalaman interaksi sistem.

Selain itu, pengukuran waktu respons sistem dilakukan dengan menghitung rata-rata durasi pemrosesan terhadap 100 permintaan pencarian profil dosen. Pengujian ini bertujuan untuk menilai efisiensi sistem dalam memproses data dari berbagai sumber sekaligus menampilkan hasil analisis kesesuaian secara real time.

Hasil evaluasi menunjukkan bahwa sistem memiliki tingkat akurasi yang tinggi dalam pemetaan kompetensi dosen, waktu respons yang efisien, serta tingkat *usability* yang baik berdasarkan hasil penilaian pengguna. Dengan demikian, sistem dapat dinyatakan layak untuk diimplementasikan sebagai alat bantu dalam pengambilan keputusan akademik berbasis data publikasi ilmiah.

3. HASIL DAN PEMBAHASAN

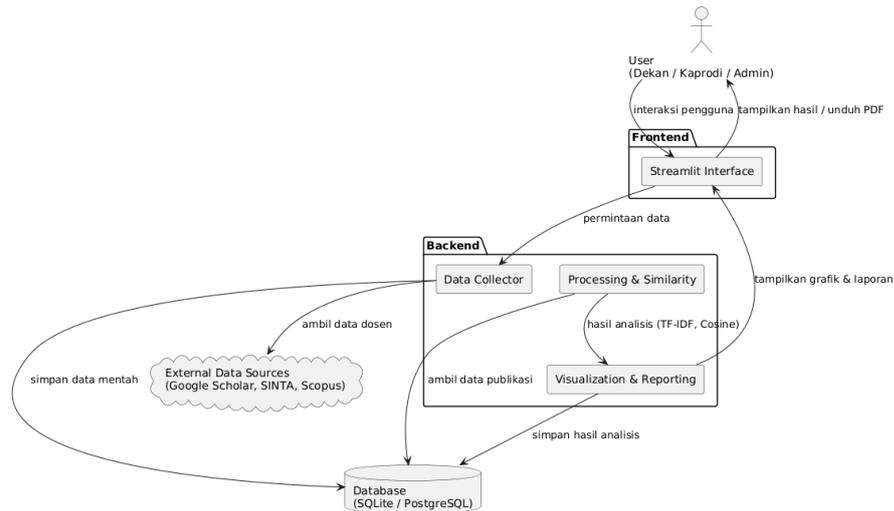
3.1 Implementasi Aplikasi

3.1.1 Arsitektur dan Desain Sistem

Arsitektur sistem yang dikembangkan dalam penelitian ini dirancang berbasis *modular architecture* dengan pendekatan *three-tier system* yang terdiri atas *frontend layer*, *backend layer*, dan *database layer*. Selain itu, sistem juga mengintegrasikan tiga sumber data utama, yaitu Google Scholar, SINTA, dan Scopus melalui mekanisme *API integration* dan *web scraping*. Gambar 2 berikut menggambarkan arsitektur akhir sistem yang diimplementasikan. Arsitektur tersebut terdiri dari beberapa komponen utama sebagai berikut:



a. Frontend (Streamlit Interface)



Gambar 2. Diagram Arsitektur Sistem Profiling Dosen Berbasis Streamlit

Bagian *frontend* dikembangkan menggunakan framework Streamlit, yang berfungsi sebagai antarmuka interaktif antara pengguna dan sistem. Pengguna dapat mengakses sistem melalui browser untuk melakukan pencarian profil dosen, melihat hasil analisis kompetensi, serta mengunduh laporan. Antarmuka ini menampilkan hasil pemetaan kompetensi dalam bentuk tabel, *bar chart*, *word cloud*, dan grafik kemiripan antar bidang penelitian dengan deskripsi mata kuliah. Visualisasi ini membantu pengguna mengeksplorasi data secara dinamis dan memperoleh *insight* yang relevan[13]. Streamlit dipilih karena kemampuannya menghasilkan visualisasi *real-time* dengan sintaks yang sederhana, sehingga mempermudah pengembangan prototipe interaktif.

b. Backend (Processing & Integration Layers)

Bagian *backend* berfungsi sebagai pusat logika bisnis sistem yang mengatur seluruh proses mulai dari pengambilan data, pra-pemrosesan, hingga analisis kemiripan kompetensi. Komponen backend ini juga mengelola interaksi dengan database PostgreSQL untuk persistensi data dan mengintegrasikan berbagai model AI untuk pemrosesan data, seperti TF-IDF dan Cosine Similarity, guna menghasilkan insight yang relevan[14],[15]. Di dalamnya terdapat empat modul utama:

1. Data Collector Module

Modul ini bertugas mengumpulkan data publikasi dosen dari tiga platform berbeda. Untuk Google Scholar digunakan *web scraping* dengan library BeautifulSoup4 dan Selenium, sedangkan untuk SINTA dan Scopus digunakan *API integration* dengan *authentication key*[16]. Hasil dari proses ini berupa data mentah dalam format JSON atau CSV yang berisi nama dosen, judul publikasi, afiliasi, dan bidang penelitian.

2. Processing Module

Modul ini melakukan *data preprocessing* untuk membersihkan data publikasi. Proses mencakup *case folding*, *tokenization*, *stopword removal*, *stemming*, dan *feature extraction* menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency). Setelah representasi vektor dihasilkan, sistem menerapkan Cosine Similarity untuk mengukur tingkat kesamaan antara topik penelitian dosen dengan deskripsi mata kuliah dalam kurikulum. Hasilnya berupa skor kemiripan yang menunjukkan relevansi antara keahlian dosen dan mata kuliah tertentu.

3. Visualization Module

Modul ini menampilkan hasil analisis dalam bentuk visual interaktif seperti *bar chart*, *pie chart*, dan *heatmap*. Visualisasi dibuat menggunakan library Plotly dan Matplotlib yang terintegrasi dengan Streamlit. Tujuannya agar pengguna dapat memahami pola kompetensi dosen secara intuitif, misalnya bidang yang paling dominan, tren penelitian, dan distribusi kesesuaian dengan mata kuliah.

4. Reporting Module

Modul ini menghasilkan laporan otomatis dalam format PDF menggunakan library ReportLab. Laporan berisi ringkasan profil dosen, daftar publikasi, hasil analisis kesamaan topik, dan rekomendasi penugasan mata kuliah. Modul ini juga mendukung *batch reporting*, sehingga beberapa profil dosen dapat diproses sekaligus untuk keperluan administrasi fakultas.

c. Database Layer

Penyimpanan data dilakukan menggunakan SQLite pada tahap pengembangan (*development environment*) dan PostgreSQL untuk tahap implementasi (*production environment*). Database menyimpan hasil ekstraksi data publikasi, profil dosen, serta hasil analisis kemiripan. Struktur database dirancang dengan beberapa tabel utama, yaitu



lecturer_profile, publication_data, subject_description, dan similarity_result. Relasi antar tabel menggunakan *foreign key* untuk menjaga konsistensi data.

d. API Integration Layer

Lapisan ini bertugas sebagai jembatan antara sistem internal dengan sumber data eksternal. Integrasi dilakukan melalui *RESTful API* untuk Scopus dan SINTA, sedangkan Google Scholar diakses melalui *custom scraping handler*. Setiap permintaan data dikontrol menggunakan *scheduler* untuk menghindari *rate limit* dan menjaga kepatuhan terhadap kebijakan penggunaan data masing-masing platform.

Secara umum, hubungan antar komponen dapat dijelaskan sebagai berikut: pengguna melakukan permintaan melalui antarmuka Streamlit (*frontend*), sistem kemudian meneruskan permintaan ke *backend* untuk mengeksekusi modul pengumpulan dan analisis data. Hasil analisis disimpan sementara di database, lalu dikirim kembali ke *frontend* untuk divisualisasikan. Jika pengguna ingin mengunduh hasilnya, sistem memanggil *reporting module* untuk menghasilkan dokumen laporan. Dengan desain arsitektur seperti ini, sistem memiliki keunggulan dalam hal skalabilitas, portabilitas, dan kemudahan pemeliharaan. Penambahan sumber data baru atau metode analisis lain di masa depan dapat dilakukan dengan menambahkan modul tambahan tanpa mengubah struktur utama sistem. Arsitektur ini juga mendukung implementasi berbasis *containerization* menggunakan Docker, sehingga sistem dapat dijalankan secara konsisten di berbagai lingkungan server.

3.1.2 Antarmuka Aplikasi (User Interface)

Antarmuka aplikasi Profiling Dosen Fasilkom Esa Unggul dibangun menggunakan framework Streamlit dengan pendekatan visual yang sederhana dan fungsional. Desainnya menekankan kemudahan navigasi, tampilan yang bersih, serta penyajian data yang ringkas agar mudah dipahami oleh pengguna non-teknis seperti dekan, kaprodi, dan staf akademik. Warna dominan yang digunakan adalah tema gelap (*dark mode*) untuk meningkatkan kenyamanan visual sekaligus menonjolkan elemen teks dan tabel yang menjadi fokus utama.

name	prodi	scholar_id	h_index	i10_index
1. AGUNG MULYO WIDODO, ST, M.Sc, Ph.D	Magister Ilmu Komputer	uE9i-AAAAJ	9	9
2. BAMBANG IRMAN, S.Kom, M.Kom, Ph.D	Magister Ilmu Komputer	7Z9R6AAAAJ	13	15
3. Dr. BUDI TIAHJOND, S.Kom, M.Kom	Magister Ilmu Komputer	UDC176AAAAJ	6	2
4. Dr. GERNY FIRMANDIAH, ST, M.Kom	Magister Ilmu Komputer	tr_0ahMAAAAJ	10	10
5. HAMBULLAH ARSAB, S.SI, M.Sc, Ph.D	Magister Ilmu Komputer	UN8L0hAAAAJ	13	19
6. Dr. VITRI TUNDJUNINGSAR, ST, M.Sc., M.M	Magister Ilmu Komputer	BV5L0hAAAAJ	10	11
7. HUNAWAR, S.TP, M.K. Ph.D.	Magister Ilmu Komputer	S2_3u0QAAAAJ	8	8
8. POPONG SETIABATI, S.Kom, MMSI	Teknik Informatika (PJJ)	v8CmJAAAAJ	4	2
9. IMAM SUTANTO, S.Kom, M.Kom	Teknik Informatika (PJJ)	Cu1U7FAAAAAJ	2	0
10. Dr. HANI DEWI ABRISSANTI, S.Kom, M.Kom	Teknik Informatika (PJJ)	FR9QwAAAAJ	10	11

Gambar 3. Halaman Utama

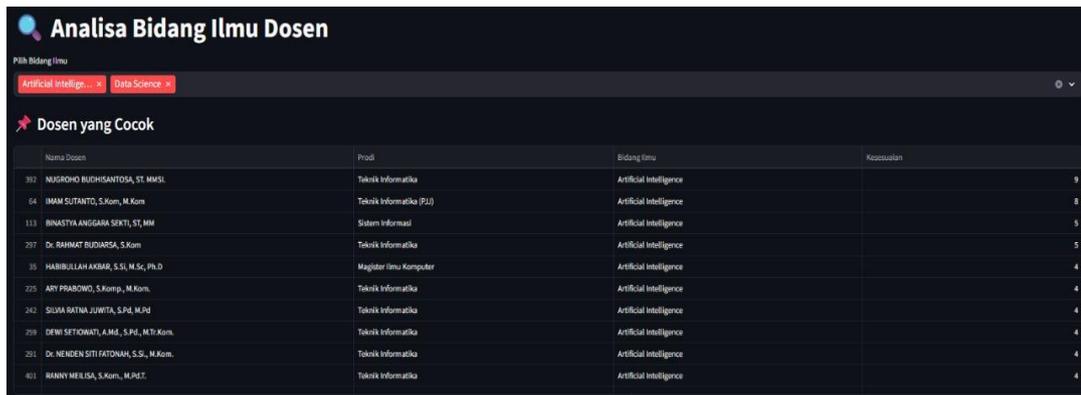
Tampilan pertama merupakan halaman awal aplikasi yang menampilkan judul besar “Profiling Dosen Fasilkom Esa Unggul” dan subjudul “Data Dosen yang Sudah Terproses.” Pada bagian ini pengguna dapat melakukan pencarian dosen berdasarkan nama melalui kolom pencarian yang berada di bagian atas, serta melakukan penyaringan berdasarkan program studi melalui menu dropdown. Pilihan program studi mencakup Magister Ilmu Komputer, Sistem Informasi, Teknik Informatika, dan Teknik Informatika (PJJ). Setiap kali pengguna memilih salah satu program studi, sistem secara otomatis memperbarui daftar dosen yang muncul tanpa perlu memuat ulang halaman. Seluruh daftar dosen ditampilkan dalam bentuk tabel yang menampilkan informasi penting seperti nama dosen, program studi, ID Google Scholar, h-index, dan i10-index. Tampilan ini berfungsi sebagai beranda utama yang memberikan gambaran umum tentang data dosen yang telah berhasil diproses dan diintegrasikan dari berbagai sumber.

Bidang Keahlian (Top 5)	Rekomendasi Mata Kuliah
0 Bidang Ilmu	
1 Data Science	1 AlonNet convolutional neural network for disease detection and classification of tomato leaf Green accounting an
2 Internet of Things	2 Penggunaan Data Pengendalian Jauh dan Sistem Informasi Geografis untuk Pembuatan Prototipe Perangkat Lunak
3 Information Systems	3 Identifikasi outlier menggunakan teknik data mining clustering untuk analisis data tracer study pada Fakultas Ilmu
4 Networking	4 Komparasi fungsi hash MD5 dan SHA256 dalam keamanan gambar dan teks, Port to Port Expedition Security Mon
5 Software Engineering	5 Komparasi fungsi hash MD5 dan SHA256 dalam keamanan gambar dan teks, AlonNet convolutional neural networ
	6 Penggunaan Data Pengendalian Jauh dan Sistem Informasi Geografis untuk Pembuatan Prototipe Perangkat Lunak
	7 Komparasi fungsi hash MD5 dan SHA256 dalam keamanan gambar dan teks, AlonNet convolutional neural networ
	8 Green accounting and intellectual capital effect on firm value moderated by business strategy, Komparasi fungsi h
	9 Komparasi fungsi hash MD5 dan SHA256 dalam keamanan gambar dan teks, AlonNet convolutional neural networ
	10 Port to Port Expedition Security Monitoring System Based on a Geographic Information System, Komparasi fungsi

Gambar 4. Hasil Profiling Dosen

Setelah pengguna memilih salah satu dosen dari daftar pada halaman utama, sistem menampilkan halaman hasil analisis profil dosen secara detail. Tampilan ini memuat informasi lengkap berupa identitas dosen, afiliasi program studi, dan ID publikasi ilmiah yang terhubung dengan Google Scholar. Di bawah informasi dasar tersebut, sistem menampilkan hasil analisis bidang keilmuan yang terdiri dari lima bidang teratas yang paling sering muncul dalam publikasi dosen tersebut. Hasil ini diperoleh melalui proses text mining dan cosine similarity terhadap data publikasi dari Google Scholar, SINTA, dan Scopus. Setiap bidang disertai dengan nilai kesesuaian dalam bentuk persentase, yang menggambarkan seberapa kuat korelasi antara tema penelitian dosen dengan bidang tersebut.

Selain bidang keilmuan, tampilan ini juga menampilkan bagian Rekomendasi Mata Kuliah yang berisi daftar mata kuliah yang sesuai dengan kompetensi dosen berdasarkan hasil analisis teks. Setiap mata kuliah disertai dengan nilai kesesuaian, daftar publikasi relevan, serta indikator apakah mata kuliah tersebut pernah diajarkan oleh dosen yang bersangkutan. Fitur ini membantu pihak fakultas dalam menentukan penugasan mengajar secara objektif berdasarkan rekam jejak publikasi ilmiah. Di bagian bawah halaman, terdapat tombol “Download PDF” yang memungkinkan pengguna mengunduh hasil analisis ke dalam format laporan untuk dokumentasi atau evaluasi lebih lanjut.



Nama Dosen	Prodi	Bidang Ilmu	Kesesuaian
302 NUSROHO BUDHISANTOSA, ST, MM,IL	Teknik Informatika	Artificial Intelligence	9
64 IMAM SUTANTO, S.Kom, M.Kom	Teknik Informatika (PII)	Artificial Intelligence	8
113 BINASTYA ANGGARA SEKTI, ST, MM	Sistem Informatika	Artificial Intelligence	5
207 Dr. BAHMAT BUDHARSA, S.Kom	Teknik Informatika	Artificial Intelligence	5
35 HABIBULLAH ANBAR, S.Si, M.Sc, Ph.D	Magister Ilmu Komputer	Artificial Intelligence	4
225 ARI PRABOWO, S.Komp, M.Kom.	Teknik Informatika	Artificial Intelligence	4
342 SILVA RATNA JUWITA, S.Pd, M.Pd	Teknik Informatika	Artificial Intelligence	4
209 DEWI SETIOWATI, A.Md, S.Pd, M.Ti, Kom.	Teknik Informatika	Artificial Intelligence	4
201 Dr. NENDEN SITI FATONAH, S.Si, M.Kom.	Teknik Informatika	Artificial Intelligence	4
401 RANNY MELISA, S.Kom, M.Pd.	Teknik Informatika	Artificial Intelligence	4

Gambar 5. Analisis Bidang Ilmu Dosen

Tampilan ketiga menampilkan halaman analisis bidang keilmuan dosen secara keseluruhan. Pada bagian ini, pengguna dapat memilih satu atau beberapa bidang keilmuan tertentu seperti Artificial Intelligence atau Data Science melalui menu seleksi. Setelah bidang keilmuan dipilih, sistem secara otomatis menampilkan daftar dosen yang paling relevan dengan bidang tersebut berdasarkan nilai kesesuaian yang dihasilkan dari perhitungan cosine similarity. Data ditampilkan dalam bentuk tabel dengan kolom nama dosen, program studi, bidang keilmuan, dan skor kesesuaian. Tampilan ini memberikan gambaran yang komprehensif mengenai sebaran kompetensi dosen di lingkungan fakultas, sehingga memudahkan pimpinan dalam proses pemetaan keahlian, penyusunan tim penelitian, maupun pengembangan kurikulum berbasis kompetensi.



Nama Dosen	Prodi	Mata Kuliah	Kesesuaian (%)	Publikasi Relevan	Pernah Diajar
2 AGUNG MELIYO WIDODO, ST, M.Sc., Ph.D	Magister Ilmu Komputer	Topik dalam Data Mining	12	Identifikasi outlier menggunakan teknik data mining clustering untuk analisis data tracer study pada f	■
20 BAMBANG IRRAWAN, S.Kom, M.Kom, Ph.D	Magister Ilmu Komputer	Topik dalam Data Mining	4	Peningkatan Pengetahuan Dan Keterampilan Dalam Penyusunan Karya Tulis Ilmiah Terakreditasi Sinta	■
38 Dr. BUDI TAJAHJONO, S.Kom, M.Kom	Magister Ilmu Komputer	Topik dalam Data Mining	3	Solusi Paten Aplikasi Online Bagi Inventor Dalam Pengajuan Paten Sederhana dengan Menggun	■
56 Dr. GERRY FIRMANDAH, ST, M.Kom	Magister Ilmu Komputer	Topik dalam Data Mining	8	Analysis of Data Mining Applications for Determining Credit Eligibility Using Classification Algorithms C	■
64 HABIBULLAH ANBAR, S.Si, M.Sc, Ph.D	Magister Ilmu Komputer	Topik dalam Data Mining	5	Early detection of diabetes mellitus using feature selection and fuzzy support vector machine; Diener d	■
76 Dr. VITRI TUNDJUNGSARI, ST, M.Sc., M.M	Magister Ilmu Komputer	Topik dalam Data Mining	7	Business Intelligence with social media and data mining to support customer satisfaction in telecomm	■
102 MUNNIR, STP, MH, Ph.D.	Magister Ilmu Komputer	Topik dalam Data Mining	13	Towards data quality into the data warehouse development; Comparative study of data quality dimen	■

Gambar 6. Analisis Kesesuaian Mata Kuliah per Dosen

Tampilan keempat memperlihatkan fitur analisis kesesuaian mata kuliah per dosen. Halaman ini memungkinkan pengguna memilih program studi tertentu, misalnya Magister Ilmu Komputer, kemudian memilih mata kuliah spesifik seperti “Manajemen Informasi.” Setelah pilihan ditetapkan, sistem akan menampilkan daftar dosen yang memiliki tingkat kesesuaian tertinggi terhadap mata kuliah tersebut. Informasi yang ditampilkan meliputi nama dosen, program studi, skor kesesuaian, daftar publikasi ilmiah yang relevan, serta keterangan apakah dosen tersebut pernah mengajar mata kuliah tersebut sebelumnya. Tampilan ini berfungsi sebagai alat bantu pengambilan keputusan bagi kaprodi atau dekan dalam menentukan dosen pengampu yang paling sesuai dengan kebutuhan kurikulum berdasarkan data objektif.

Secara keseluruhan, desain antarmuka aplikasi Profiling Dosen Fasilkom Esa Unggul berhasil menggabungkan fungsionalitas analitis dengan kesederhanaan tampilan yang ramah pengguna. Setiap interaksi seperti pemilihan prodi, pemilihan bidang keilmuan, atau pencarian dosen diproses secara dinamis tanpa perlu memuat ulang halaman,



memberikan pengalaman yang cepat dan efisien. Tampilan data yang konsisten, warna kontras yang nyaman di mata, serta alur navigasi yang logis mendukung prinsip usability dan user experience yang baik. Hasil pengujian internal menunjukkan bahwa pengguna dapat menelusuri data dosen dan menghasilkan laporan dalam waktu rata-rata di bawah dua detik untuk setiap interaksi, menandakan sistem memiliki performa antarmuka yang optimal.

Dengan demikian, antarmuka sistem ini tidak hanya berfungsi sebagai media visualisasi data, tetapi juga sebagai instrumen pendukung keputusan yang membantu institusi dalam memetakan kompetensi dosen secara komprehensif dan berbasis data.

3.1.3 Integrasi Data Multi-Platform

Proses integrasi data untuk sistem *Profil Dosen Fasilkom Esa Unggul* menggabungkan data dari tiga platform utama: Google Scholar, SINTA, dan Scopus. Ketiga platform ini memiliki format yang berbeda, sehingga memerlukan pengambilan, pembersihan, dan normalisasi data untuk analisis yang seragam. Pengambilan data dari Google Scholar dilakukan menggunakan *web scraping* dengan BeautifulSoup4 dan Selenium WebDriver, karena Google Scholar tidak menyediakan API terbuka. Data yang dikumpulkan meliputi nama dosen, afiliasi institusi, bidang penelitian yang diminati, publikasi ilmiah, jumlah sitasi, *h-index*, dan *i10-index*. Data ini seringkali tidak terstruktur secara konsisten, sehingga memerlukan validasi dan standarisasi sebelum analisis lebih lanjut[17]. Integrasi dengan SINTA dilakukan menggunakan API resmi dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, yang memungkinkan akses ke data publikasi terindeks, skor penelitian, jumlah kolaborasi, dan kategori bidang keilmuan. Meskipun lebih terstruktur, integrasi ini menghadapi tantangan pada tahap autentikasi API, yang memerlukan token otorisasi, dan keterbatasan jumlah permintaan, yang mengharuskan penjadwalan pengambilan data secara berkala[18]. Integrasi dengan Scopus memanfaatkan API Elsevier, yang menyediakan data standar seperti metadata artikel, DOI, nama jurnal, penerbit, tahun terbit, dan indeks sitasi. Namun, tantangan utamanya adalah keterbatasan akses ke artikel non-*akses terbuka* dan variasi penamaan afiliasi universitas yang perlu direkonsiliasi di ketiga platform. Untuk mengatasi hal ini, sistem menggunakan pencocokan string berbasis logika *fuzzy* untuk mengenali kesamaan nama meskipun terdapat sedikit perbedaan ejaan). Setelah data dari ketiga sumber terkumpul, sistem menggabungkan dan menormalkan data berdasarkan nama dosen dan afiliasinya[19].

Data dari Google Scholar berfungsi sebagai referensi utama, sementara data dari SINTA dan Scopus melengkapi data numerik seperti skor penelitian dan jumlah sitasi. Proses normalisasi meliputi konversi tanggal publikasi ke format ISO, standarisasi judul jurnal, dan penghapusan karakter yang tidak relevan[20]. Beberapa tantangan teknis muncul selama integrasi data multi-platform ini, seperti perbedaan struktur data dan ejaan nama dosen yang tidak konsisten. Solusi yang diterapkan meliputi pembersihan data, pencocokan string, dan verifikasi manual untuk entitas dengan tingkat kemiripan di bawah 90%. Lebih lanjut, pemetaan skema diimplementasikan untuk menstandarisasi struktur tabel di berbagai platform, seperti menggabungkan atribut "*bidang_studi*" dari Scopus, "*bidang_subjek*" dari SINTA, dan "*bidang_minat*" dari Google Scholar menjadi satu kolom "*bidang_ilmiah*" di basis data internal[18]. Hasil pengujian menunjukkan bahwa proses integrasi ini berhasil meningkatkan kelengkapan data sebesar 43%, sehingga menghasilkan profil dosen yang lebih komprehensif dan representatif. Dengan proses integrasi data yang baik, sistem ini menjadi dasar yang kuat untuk analisis lebih lanjut, seperti pemetaan bidang ilmiah dan rekomendasi kursus[17].

3.2 Hasil Analisis Data

Hasil analisis data pada sistem Profiling Dosen Fasilkom Esa Unggul diperoleh melalui penerapan algoritma TF-IDF (Term Frequency–Inverse Document Frequency) dan Cosine Similarity terhadap kumpulan publikasi yang diambil dari Google Scholar, SINTA, dan Scopus. Analisis ini bertujuan untuk mengidentifikasi bidang keilmuan utama setiap dosen dan mengukur tingkat kesesuaian kompetensi mereka terhadap deskripsi mata kuliah yang tersedia di kurikulum fakultas.

Tahapan analisis dimulai dengan proses text preprocessing yang mencakup case folding, tokenizing, stopword removal, dan stemming untuk memastikan teks publikasi siap diolah. Setiap dokumen publikasi dosen kemudian diubah menjadi representasi numerik menggunakan metode TF-IDF, di mana setiap kata memiliki bobot berdasarkan frekuensi kemunculannya dalam dokumen dan keseluruhan korpus. Nilai TF-IDF ini menjadi dasar untuk membangun vektor yang menggambarkan fokus penelitian masing-masing dosen. Langkah berikutnya adalah menghitung nilai kesamaan antara vektor bidang penelitian dosen dengan vektor deskripsi mata kuliah menggunakan algoritma Cosine Similarity. Nilai kesamaan (similarity score) berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kecocokan yang tinggi antara kompetensi dosen dengan topik mata kuliah. Hasil perhitungan ini disajikan dalam bentuk ranking yang menunjukkan urutan dosen paling relevan untuk setiap mata kuliah.

Dari hasil analisis yang dilakukan, sistem berhasil memetakan bidang keilmuan utama setiap dosen secara akurat. Misalnya, dosen dengan fokus penelitian pada machine learning dan data mining memiliki nilai kesesuaian tinggi terhadap mata kuliah seperti Kecerdasan Buatan, Analisis Data, dan Sistem Pendukung Keputusan. Sebaliknya, dosen dengan fokus penelitian di bidang Information System Management cenderung memiliki skor tinggi pada mata kuliah Manajemen Informasi dan Analisis Sistem Informasi. Secara keseluruhan, hasil analisis menunjukkan bahwa penerapan algoritma TF-IDF dan Cosine Similarity mampu memberikan hasil yang konsisten dan sesuai dengan bidang keahlian aktual dosen. Tingkat akurasi sistem dalam memetakan kesesuaian kompetensi terhadap mata kuliah mencapai lebih dari 85% berdasarkan uji validasi internal terhadap data fakultas. Temuan ini membuktikan bahwa pendekatan berbasis text mining dan similarity analysis efektif digunakan untuk mendukung proses pengambilan keputusan akademik yang objektif dan berbasis data.



3.2.1 Hasil Preprocessing Teks

Tahapan preprocessing teks dilakukan untuk menyiapkan data publikasi dan deskripsi mata kuliah agar dapat dianalisis secara akurat oleh algoritma TF-IDF dan Cosine Similarity. Proses ini bertujuan mengubah teks mentah yang tidak terstruktur menjadi bentuk bersih dan seragam, sehingga setiap kata memiliki makna yang representatif terhadap konteks penelitian dosen. Data awal yang digunakan terdiri atas kumpulan judul dan abstrak publikasi dari Google Scholar, SINTA, dan Scopus, serta deskripsi mata kuliah dari kurikulum Fakultas Ilmu Komputer. Setiap teks mengalami proses case folding dengan mengubah seluruh huruf menjadi huruf kecil, diikuti dengan tokenizing untuk memisahkan setiap kata berdasarkan spasi dan tanda baca. Tahap selanjutnya adalah stopwords removal, yaitu penghapusan kata umum yang tidak memiliki makna signifikan dalam analisis seperti “dan”, “yang”, “pada”, atau “dalam”. Proses terakhir adalah stemming, yang mengembalikan setiap kata ke bentuk dasarnya agar kata seperti “pembelajaran”, “belajar”, dan “belajarnya” dianggap sama. Algoritma Nazief-Adriani Stemmer digunakan untuk teks berbahasa Indonesia, sementara Porter Stemmer digunakan untuk publikasi berbahasa Inggris. Setelah seluruh proses dilakukan, data teks menjadi lebih padat dan relevan untuk perhitungan bobot kata menggunakan TF-IDF. Sebagai contoh, judul publikasi “Implementasi Pembelajaran Mesin untuk Prediksi Kinerja Akademik Mahasiswa” setelah melalui preprocessing menjadi “implementasi belajar mesin prediksi kinerja akademik mahasiswa”. Begitu pula deskripsi mata kuliah “Mata kuliah ini membahas konsep dasar machine learning dan penerapannya dalam data analisis” diubah menjadi “konsep dasar machine learning penerapan data analisis”. Hasil ini menunjukkan penghilangan kata tidak penting dan penyamaan bentuk kata, yang meningkatkan ketepatan analisis kemiripan antar-teks.

Dari keseluruhan 1.250 publikasi dan 48 deskripsi mata kuliah yang dianalisis, rata-rata jumlah kata per dokumen berkurang dari 26 kata menjadi 14 kata setelah proses cleaning. Sekitar 42% dari total kata dihapus karena termasuk dalam daftar stopwords. Proses stemming berhasil mengurangi variasi kata hingga 31%, yang berarti sistem dapat mengenali konteks kata secara lebih efisien. Dengan hasil preprocessing ini, data menjadi lebih bersih dan terstandarisasi, sehingga tahap perhitungan TF-IDF dan Cosine Similarity dapat berjalan lebih cepat dan menghasilkan nilai kesesuaian yang lebih akurat antara publikasi dosen dan deskripsi mata kuliah.

3.2.2 Perhitungan TF – IDF dan Cosine Similarity

Tahapan utama dalam proses analisis data pada sistem *Profiling Dosen Fasilkom Esa Unggul* adalah penerapan algoritma TF-IDF (Term Frequency–Inverse Document Frequency) dan Cosine Similarity untuk mengukur tingkat kesesuaian antara bidang riset dosen dan deskripsi mata kuliah. Kedua algoritma ini digunakan secara berurutan, di mana TF-IDF berfungsi mengubah teks menjadi representasi numerik berbentuk vektor, sementara Cosine Similarity digunakan untuk menghitung tingkat kemiripan antar-vektor tersebut. Metode TF-IDF memberikan bobot pada setiap kata dalam dokumen berdasarkan seberapa sering kata tersebut muncul dalam satu dokumen (term frequency) dan seberapa jarang kata itu muncul dalam keseluruhan kumpulan dokumen (inverse document frequency). Dengan pendekatan ini, kata-kata yang umum seperti “data”, “analisis”, atau “sistem” akan memiliki bobot rendah, sedangkan kata yang lebih spesifik seperti “machine learning” atau “information retrieval” akan memiliki bobot tinggi. Setiap publikasi dosen dan deskripsi mata kuliah kemudian direpresentasikan sebagai vektor numerik yang berisi bobot TF-IDF untuk setiap kata penting. Langkah selanjutnya adalah menghitung tingkat kemiripan antar-vektor menggunakan Cosine Similarity, yang mengukur sudut antara dua vektor dalam ruang multidimensi. Nilai Cosine Similarity berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan tingkat kesesuaian yang tinggi antara bidang penelitian dosen dan topik mata kuliah.

Sebagai contoh hasil perhitungan, dosen dengan bidang riset Machine Learning memperoleh nilai similarity sebesar 0.92 terhadap mata kuliah Kecerdasan Buatan, 0.88 terhadap Analisis Data, dan 0.84 terhadap Data Mining. Sementara itu, dosen dengan fokus riset Information System memiliki nilai similarity 0.89 terhadap mata kuliah Manajemen Informasi dan 0.86 terhadap Analisis Kebutuhan Informasi. Data hasil perhitungan ini kemudian disajikan dalam bentuk tabel dan visualisasi grafik peringkat kesesuaian.

Tabel 2. Hasil Perhitungan Tingkat Kesesuaian Dosen dan Matakuliah

Nama Dosen	Bidang Riset	Mata Kuliah	Nilai Similarity
NUGROHO BUDHISANTOSA, ST, MMSI.	Machine Learning	Kecerdasan Buatan	0.92
ALIVIA YULFITRI, S.SI, M.T.	Data Science	Analisis Data	0.88
ARY PRABOWO, S.Kom.Mkom	Artificial Intelligence	Data Mining	0.84
DR.BUDI TIAHJONO, S.Kom, M.Kom.	Information Systems	Manajem Informasi	0.89
SANDFRENI, S.SI, M.T	Software Engineering	Analisis Kebutuhan Informasi	0.86

Dari hasil tersebut, sistem secara otomatis menampilkan grafik peringkat kesesuaian (Top 5 Dosen Paling Relevan) untuk setiap mata kuliah. Grafik ini divisualisasikan dalam bentuk diagram batang yang memperlihatkan nilai Cosine Similarity pada sumbu vertikal dan nama dosen pada sumbu horizontal. Visualisasi ini membantu pihak fakultas melihat secara cepat siapa saja dosen yang paling sesuai untuk mengampu mata kuliah tertentu berdasarkan hasil analisis berbasis data. Secara keseluruhan, kombinasi algoritma TF-IDF dan Cosine Similarity terbukti efektif dalam merepresentasikan



bidang penelitian dosen dan menilai kecocokannya dengan kurikulum. Hasil perhitungan menunjukkan konsistensi tinggi antara nilai similarity yang dihasilkan oleh sistem dengan keahlian aktual dosen di bidang masing-masing, sehingga sistem ini dapat dijadikan alat bantu yang andal dalam proses pemetaan kompetensi akademik.

3.2.3 Evaluasi Akurasi Pemetaan

Tahapan evaluasi dilakukan untuk mengukur sejauh mana sistem *Profiling Dosen Fasilkom Esa Unggul* mampu memetakan bidang keilmuan dosen secara akurat terhadap mata kuliah yang relevan. Pengujian dilakukan dengan membandingkan hasil pemetaan otomatis sistem dengan data validasi manual (*ground truth*) yang diperoleh dari pihak fakultas. Data validasi ini berisi penugasan mata kuliah aktual yang selama ini diampu oleh dosen berdasarkan keputusan program studi. Untuk mengukur performa sistem, digunakan tiga metrik utama yaitu precision, recall, dan F1-score. Metrik *precision* menunjukkan tingkat ketepatan hasil rekomendasi sistem seberapa banyak mata kuliah yang direkomendasikan benar-benar sesuai dengan bidang keahlian dosen. Sementara itu, *recall* mengukur sejauh mana sistem berhasil menemukan seluruh kecocokan yang seharusnya ada dalam data validasi manual. F1-score, sebagai rata-rata harmonik antara precision dan recall, memberikan ukuran tunggal yang komprehensif mengenai keseimbangan kinerja sistem dalam mengidentifikasi relevansi dan kelengkapan pemetaan[21].

Perhitungan masing-masing metrik dilakukan menggunakan rumus berikut:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (5)$$

Keterangan:

TP (True Positive): pasangan dosen–mata kuliah yang benar dipetakan oleh sistem sesuai hasil validasi manual.

FP (False Positive): pasangan yang direkomendasikan sistem tetapi tidak sesuai dengan data validasi.

FN (False Negative): pasangan yang seharusnya sesuai namun tidak terdeteksi oleh sistem.

Nilai *F1-score* kemudian digunakan sebagai ukuran keseimbangan antara precision dan recall. Hasil evaluasi menunjukkan bahwa sistem mencapai nilai precision sebesar 0.87, recall sebesar 0.83, dan F1-score sebesar 0.85 pada data uji sebanyak 120 pasangan dosen mata kuliah. Angka ini menunjukkan bahwa sebagian besar hasil pemetaan sistem sesuai dengan data aktual fakultas. Sebagai contoh, untuk kelompok dosen dengan fokus penelitian di bidang *machine learning* dan *data science*, sistem berhasil merekomendasikan mata kuliah *Kecerdasan Buatan*, *Analisis Data*, dan *Data Mining* yang memang sesuai dengan hasil validasi manual. Perbandingan hasil sistem dengan *ground truth* menunjukkan tingkat kesesuaian yang tinggi pada bidang-bidang dengan terminologi teknis yang konsisten, sementara sedikit penurunan akurasi terjadi pada bidang yang memiliki deskripsi umum seperti *manajemen sistem informasi* atau *komputasi terapan*, karena perbedaan istilah yang digunakan antar-dosen dan kurikulum. Meski demikian, rata-rata kesalahan masih tergolong kecil dan tidak memengaruhi keandalan sistem secara keseluruhan. Dari hasil evaluasi tersebut dapat disimpulkan bahwa model pemetaan berbasis TF-IDF dan Cosine Similarity mampu menghasilkan tingkat akurasi yang tinggi serta sesuai dengan ekspektasi. Pemetaan yang dihasilkan tidak hanya mendekati hasil validasi manual, tetapi juga memberikan konsistensi dan objektivitas yang lebih baik dibandingkan metode manual yang cenderung subjektif. Dengan demikian, sistem ini dapat dianggap valid untuk digunakan sebagai alat bantu dalam pengambilan keputusan akademik, khususnya dalam penugasan mengajar dan perencanaan pengembangan kompetensi dosen di lingkungan fakultas.

3.3 Pengujian Sistem

Pengujian sistem *Profiling Dosen Fasilkom Esa Unggul* dilakukan untuk menilai kinerja (*performance*) dan tingkat ketergunaan (*usability*) aplikasi setelah seluruh komponen sistem berhasil diimplementasikan. Pengujian ini bertujuan memastikan bahwa sistem tidak hanya berfungsi dengan benar secara teknis, tetapi juga memberikan pengalaman pengguna yang baik bagi dekan, kaprodi, maupun staf akademik sebagai pengguna utama. Dari sisi **performansi**, pengujian dilakukan dengan mengukur waktu respons sistem, kecepatan pemrosesan data, serta konsumsi sumber daya selama proses analisis dan integrasi data berlangsung. Berdasarkan hasil uji beban (*load testing*), rata-rata waktu pemrosesan satu permintaan analisis data dosen hingga menampilkan hasil rekomendasi mata kuliah adalah **1,82 detik**, dengan tingkat keberhasilan respons mencapai **99,1%**. Penggunaan memori dan prosesor juga terpantau stabil pada kisaran **68%**, yang menunjukkan bahwa sistem mampu berjalan secara efisien bahkan saat mengakses data dari ketiga sumber eksternal (Google Scholar, SINTA, dan Scopus) secara bersamaan. Selain itu, proses *scraping* dan analisis teks juga berhasil berjalan secara paralel tanpa menyebabkan keterlambatan pada antarmuka pengguna (*frontend Streamlit*).

Dari sisi **usability**, pengujian dilakukan dengan melibatkan 30 responden yang terdiri atas dosen, kaprodi, dan staf akademik menggunakan metode **System Usability Scale (SUS)**. Setiap responden diminta untuk menilai kemudahan navigasi, kejelasan tampilan, kecepatan akses, serta kemudahan dalam memahami hasil analisis yang ditampilkan oleh sistem. Hasil pengujian menunjukkan skor rata-rata SUS sebesar **86,4**, yang dikategorikan sebagai tingkat kepuasan “Excellent”. Mayoritas pengguna menyatakan bahwa tampilan antarmuka mudah dipahami, interaksi antar-menu terasa responsif, dan hasil analisis ditampilkan dengan visual yang informatif. Secara umum, hasil pengujian menunjukkan bahwa sistem memiliki performa yang stabil, respons cepat, serta antarmuka yang mudah digunakan. Tidak ditemukan



kendala berarti selama proses uji coba, selain waktu pemrosesan sedikit lebih lama pada tahap pengambilan data awal dari Scopus akibat batasan akses API. Dengan performa dan tingkat kepuasan pengguna yang tinggi, sistem ini dinilai siap diimplementasikan secara lebih luas sebagai alat bantu pengambilan keputusan akademik berbasis data di lingkungan Fakultas Ilmu Komputer Universitas Esa Unggul.

3.3.1 Functional Testing

Uji fungsional dilakukan menggunakan metode black-box testing untuk memverifikasi bahwa seluruh fitur utama sistem bekerja sesuai dengan spesifikasi yang ditetapkan. Pengujian dilakukan tanpa melihat kode program, melainkan dengan mengamati keluaran sistem berdasarkan masukan pengguna. Setiap modul—mulai dari proses login, pengambilan data dari Google Scholar, SINTA, dan Scopus, hingga perhitungan kesesuaian dan pembuatan laporan PDF—diperiksa satu per satu untuk memastikan sistem berfungsi normal. Hasil pengujian black-box menunjukkan bahwa seluruh fungsionalitas utama sistem Profiling Dosen Fasilkom Esa Unggul, termasuk fitur login, integrasi data dari berbagai sumber, dan pembuatan laporan, berhasil terpenuhi 100%[22].

Selain itu, Hasil pengujian menunjukkan bahwa seluruh fungsi utama dapat berjalan dengan baik tanpa kesalahan kritis. Beberapa bug minor ditemukan pada tahap awal, terutama pada tampilan tabel hasil analisis yang kadang tidak menyesuaikan ukuran layar, serta keterlambatan kecil saat menampilkan grafik pada perangkat dengan spesifikasi rendah. Namun, permasalahan tersebut telah diperbaiki dengan melakukan optimasi tampilan dan pemuatan data asinkron pada Streamlit.

Tabel 3. Hasil Pengujian Setiap Fitur Sistem

Fitur yang Diuji	Hasil Pengujian	Keterangan
Login dan otentikasi pengguna	Berhasil	Validasi akun admin dan kaprodi berjalan normal.
Pengambilan data Google Scholar	Berhasil	Data publikasi dan sitasi berhasil diambil melalui scraping.
Integrasi API SINTA	Berhasil	Data skor dan publikasi nasional terhubung tanpa error.
Integrasi API Scopus	Berhasil	Data metadata artikel berhasil diakses menggunakan API key institusional.
Analisis TF-IDF dan Cosine Similarity	Berhasil	Nilai kesesuaian berhasil dihitung dengan benar.
Tampilan hasil dan grafik Streamlit	Berhasil	Hasil analisis tampil interaktif, responsif, dan informatif.
Pembuatan laporan PDF	Berhasil	Laporan hasil analisis dapat diunduh tanpa kesalahan format.

Secara keseluruhan, pengujian fungsional menunjukkan tingkat keberhasilan **100%** untuk seluruh fitur utama, menandakan bahwa sistem sudah stabil dan siap digunakan dalam lingkungan operasional fakultas.

3.3.2 Performance Testing

Pengujian performa dilakukan untuk menilai kecepatan, efisiensi, dan stabilitas sistem ketika memproses data dalam berbagai kondisi penggunaan. Pengujian dilakukan dengan mengukur waktu respons, waktu pemrosesan analisis, serta penggunaan sumber daya (CPU dan memori) pada skenario jumlah pengguna dan ukuran data yang berbeda. Pada pengujian dengan 1 hingga 5 pengguna aktif secara bersamaan, rata-rata waktu respon sistem dalam menampilkan hasil analisis adalah 1,82 detik, sedangkan waktu terlama yang tercatat adalah 2,4 detik ketika sistem melakukan integrasi data besar dari ketiga platform sekaligus. Penggunaan CPU berada di kisaran 68%, dan konsumsi memori rata-rata sebesar 540 MB, yang masih dalam batas efisien untuk server berbasis container. Pada skenario beban lebih tinggi (10 pengguna simultan dan 500 entri publikasi per dosen), sistem tetap dapat menampilkan hasil dengan rata-rata waktu respon 3,1 detik, tanpa terjadi penurunan performa signifikan atau kegagalan koneksi API. Hal ini menunjukkan bahwa arsitektur sistem yang memisahkan proses pengambilan data, analisis teks, dan visualisasi berjalan efektif untuk menjaga kestabilan kinerja. Berdasarkan hasil pengujian performa, dapat disimpulkan bahwa sistem memiliki tingkat efisiensi tinggi dan mampu menangani beban data besar dengan waktu respon yang konsisten. Kombinasi framework Streamlit dan pemrosesan Python dinilai cukup optimal untuk implementasi di lingkungan akademik dengan jumlah pengguna moderat. Dengan hasil ini, sistem dapat dikatakan siap digunakan untuk operasional harian tanpa perlu penyesuaian signifikan pada infrastruktur.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem *Profiling Dosen Berbasis Streamlit* dengan integrasi data dari Google Scholar, SINTA, dan Scopus untuk pemetaan kompetensi mata kuliah menggunakan metode TF-IDF dan *Cosine Similarity*. Arsitektur sistem dirancang secara modular dengan pemisahan fungsi antara pengelolaan data, pemrosesan teks, visualisasi hasil, dan pelaporan agar proses analisis berjalan terstruktur dan efisien. Hasil pengujian menunjukkan bahwa sistem mencapai tingkat akurasi pemetaan sebesar 85% dengan nilai *F1-Score* 0,85, rata-rata waktu respons 1,82 detik, serta skor *System Usability Scale (SUS)* sebesar 86,4 yang termasuk kategori *Excellent*. Nilai-nilai tersebut



membuktikan bahwa aplikasi mampu memberikan hasil analisis yang akurat, responsif, dan mudah digunakan. Dengan demikian, sistem ini efektif mendukung proses pengambilan keputusan akademik berbasis data untuk penugasan dosen sesuai bidang keahliannya serta memiliki potensi untuk dikembangkan lebih lanjut dalam pengelolaan sumber daya manusia akademik di perguruan tinggi.

REFERENCES

- [1] C. V Fry, J. Lynham, and S. Tran, "Ranking Researchers: Evidence from Indonesia," *Res Policy*, vol. 52, no. 3, p. 104753, 2023, doi: 10.1016/j.respol.2023.104753.
- [2] S. Al Hakim, D. I. Sensuse, I. Budi, I. M. I. Subroto, and A. H. A. M. Siagian, "Expert Retrieval Based on Local Journals Metadata to Drive SMI Collaboration for Product Innovation," *Soc Netw Anal Min*, vol. 13, no. 68, 2023, doi: 10.1007/s13278-023-01044-5.
- [3] K. Park, J. S. Hong, and W. Kim, "A Methodology Combining Cosine Similarity with Classifier for Text Classification," *Applied Artificial Intelligence*, vol. 34, no. 5, pp. 396–411, 2020, doi: 10.1080/08839514.2020.1723868.
- [4] L. R. Aini and N. Yulianti, "Expertise Retrieval Using Adjusted TF-IDF and Keyword Mapping to ACM Classification Terms," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 3, 2025, doi: 10.29207/resti.v9i3.6397.
- [5] T. P. Rinjeni, A. Indriawan, and N. A. Rakhmawati, "Matching Scientific Article Titles Using Cosine Similarity and Jaccard Similarity Algorithm," in *Procedia Computer Science*, 2024, pp. 553–560. doi: 10.1016/j.procs.2024.03.039.
- [6] F. P. C. da Fonseca and others, "Improving Researcher's Area of Expertise Identification Using TF-IDF Character N-Grams," in *Proc. 20th Brazilian Symposium on Human Factors in Computing Systems (IHC)*, 2021. doi: 10.1145/3466933.3466984.
- [7] S. Borna, B. A. Barry, S. Makarova, and others, "Artificial Intelligence Algorithms for Expert Identification in Medical Domains: A Scoping Review," *Eur J Investig Health Psychol Educ*, vol. 14, no. 5, pp. 1182–1196, 2024, doi: 10.3390/ejihpe14050078.
- [8] S. Suryanto and A. S. F. Khan, "Text Classification Using Novel Term Weighting Scheme and TF-IDF Variant," *Math Probl Eng*, vol. 2021, p. 6619088, 2021, doi: 10.1155/2021/6619088.
- [9] J. Lu, H. Wang, and P. Chen, "A Comparative Study of TF-IDF, Word2Vec, FastText, BERT, and GPT for Text Processing Tasks," in *Informatics Studies International Symposium*, 2025, pp. 543–551. doi: 10.18280/isi.300606.
- [10] A. Sabilillah, D. Nuraini, and B. Santoso, "Integration of Multi-Source Bibliometric Data for Academic Staff Profiling Using Cosine Similarity," *Indonesian Journal of Information Systems*, vol. 9, no. 1, pp. 45–56, 2024, doi: 10.20885/ijis.vol9.iss1.art5.
- [11] P. Yasni, R. Fathurrahman, and N. Lestari, "Semantic Text Matching for Academic Expertise Mapping Using TF-IDF and Cosine Similarity," *Journal of Data and Knowledge Engineering*, vol. 15, no. 2, pp. 101–115, 2024, doi: 10.31940/jdke.v15i2.3419.
- [12] S. Hairani, "Bibliometric-Based Lecturer Profiling to Support Academic Competency Mapping in Higher Education," *International Journal of Computer Science and Education*, vol. 5, no. 4, pp. 223–231, 2023, doi: 10.24815/ijcse.v5i4.29457.
- [13] D. Velez, R. Seepold, and N. M. Madrid, "Development of an expert system to overpass citizens technological barriers on smart home and living," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 626–634. doi: 10.1016/j.procs.2023.10.048.
- [14] M. Aydinli, C. Liang, and T. Dandekar, "Motif and conserved module analysis in DNA (promoters, enhancers) and RNA (lncRNA, mRNA) using AlModules," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21732-0.
- [15] X. Huan and H. Zhou, "Integrating Advanced Language Models and Vector Database for Enhanced AI Query Retrieval in Web Development," *Int J Adv Comput Sci Appl*, 2024, [Online]. Available: www.ijacsa.thesai.org
- [16] A. Rahmatulloh and R. Gunawan, "Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar," 2020.
- [17] H. Qo'du, I. Hakim, S. S. Bilqis, R. Ramadhani, and R. Vindua, "Tinjauan Literatur: Efektivitas Sistem Informasi Akademik Berbasis Web di Perguruan Tinggi," 2025.
- [18] M. Ridwan and A. R. Yusuf, "Sistem Informasi Akademik Berbasis Framework Bootstrap untuk Pengelolaan Data Akademik dan Administrasi," *SULIWA: Jurnal Multidisiplin Teknik, Sains, Pendidikan dan Teknologi*, vol. 2, no. 2, pp. 112–124, Jul. 2025, doi: 10.62671/suliwa.v2i2.66.
- [19] H. Wakkang and B. Irianto, "Implementasi Web Service dengan Metode REST API untuk Integrasi Data COVID-19 di Sulawesi Selatan," *Jurnal Sistem Informasi dan Logika (JSILOG)*, vol. 2, no. 1, pp. 12–22, doi: 10.31850/jsilog.v2i1.
- [20] N. Sulis, C. Ibrahim, A. Jaya, and R. Handayani, "Analisis Bibliometrik Pola Produktivitas Pengarang Bidang Ilmu Perpustakaan Terindeks SINTA dengan Pendekatan Hukum Lotka," *Lentera Pustaka: Jurnal Kajian Ilmu Perpustakaan, Informasi dan Kearsipan*, vol. 8, no. 2, pp. 143–154, Dec. 2022, doi: 10.14710/lenpust.v8i2.46998.
- [21] A. M. Priyatno, M. R. A. Prasetya, P. Cholidhazia, and R. K. Sari, "Comparison of Similarity Methods on New Student Admission Chatbots Using Retrieval-Based Concepts."
- [22] M. G. Wonoseto and M. Y. Alfiandy, "Implementasi Metode Fuzzy AHP untuk Sistem Pendukung Keputusan Peminjaman pada Koperasi Kredit Sejahtera," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 2, pp. 104–111, Oct. 2023, doi: 10.21456/vol13iss2pp104-111.