

Sistem Rekomendasi Dosen Pembimbing Berbasis SBERT dan Cosine Similarity

Rafif Fairuz Ramadani*, Ika Nur Fajri

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Sleman, Indonesia

Email: ¹*rafiffr@students.amikom.ac.id, ²fajri@amikom.ac.id,

Email Penulis Korespondensi: rafiffr@students.amikom.ac.id*

Submitted: 17/05/2026; Accepted: 03/07/2026; Published: 30/09/2026

Abstrak— Ketidaksesuaian antara topik tugas akhir mahasiswa dan keahlian dosen pembimbing menjadi salah satu penyebab utama keterlambatan kelulusan di perguruan tinggi. Penelitian ini mengimplementasikan Sentence-BERT (SBERT) model *paraphrase-multilingual-MiniLM-L12-v2* dan *cosine similarity* untuk membangun sistem rekomendasi dosen pembimbing berbasis kemiripan semantik dari judul tugas akhir mahasiswa. Dataset penelitian mencakup 1.420 judul penelitian dari 43 dosen Program Studi Sistem Informasi, Universitas Amikom Yogyakarta, yang dikumpulkan melalui *web scraping* Google Scholar menggunakan SerpAPI. Sebagai pembaruan dari pendekatan BERT standar yang umum digunakan penelitian terdahulu, penelitian ini mengadopsi SBERT varian yang secara arsitektur dioptimalkan untuk tugas *sentence similarity* dan mendukung lebih dari 50 bahasa termasuk Indonesia dan Inggris, sehingga lebih sesuai untuk dataset bilingual. Tahap *preprocessing* hanya meliputi pembersihan karakter khusus berbasis *regex* dan normalisasi spasi, tanpa tahapan *stemming* maupun penghapusan *stopword*, karena SBERT mampu memahami konteks semantik secara langsung. Evaluasi dilakukan menggunakan 10 data uji berupa judul tugas akhir nyata dari repositori EPrints AMIKOM beserta dosen pembimbing aktual sebagai *ground truth*, dengan metrik *Hit Rate* dan *Mean Reciprocal Rank* (MRR) pada variasi Top-1, Top-2, dan Top-3. Hasil menunjukkan *Hit Rate* 40% pada Top-1, 70% pada Top-2, dan 90% pada Top-3, dengan MRR 0,400, 0,550, dan 0,617. Penggunaan data uji nyata berbasis repositori EPrints AMIKOM menghasilkan evaluasi yang lebih objektif dan dapat direplikasi. Temuan ini mengindikasikan bahwa SBERT multilingual berpotensi merekomendasikan dosen pembimbing yang relevan secara semantik pada dataset bilingual, meskipun evaluasi lebih lanjut dengan data yang lebih besar masih diperlukan.

Kata Kunci: Sentence-BERT; Cosine Similarity; Rekomendasi Dosen Pembimbing; Tugas Akhir; Natural Language Processing

Abstract— The mismatch between students' final project topics and supervisors' expertise is one of the main causes of graduation delays in higher education. This study implements Sentence-BERT (SBERT) with the *paraphrase-multilingual-MiniLM-L12-v2* model and *cosine similarity* to build a supervisor recommendation system based on the semantic similarity of students' final project titles. The dataset includes 1,420 research titles from 43 lecturers in the Information Systems Program at Universitas Amikom Yogyakarta, collected through *web scraping* from Google Scholar using SerpAPI. Unlike previous studies using standard BERT, this study employs SBERT a variant architecturally optimized for *sentence similarity* tasks and supporting over 50 languages including Indonesian and English, making it more suitable for bilingual datasets. *Preprocessing* consists solely of special character cleaning using *regex* and whitespace normalization, without *stemming* or *stopword* removal, as SBERT inherently understands semantic context. Evaluation using 10 real thesis titles from the EPrints AMIKOM repository with actual supervisors as *ground truth* with *Hit Rate* and *Mean Reciprocal Rank* (MRR) metrics across Top-1, Top-2, and Top-3 variations shows *Hit Rate* of 40% at Top-1, 70% at Top-2, and 90% at Top-3, with MRR of 0.400, 0.550, and 0.617 respectively. The use of real thesis data from the EPrints AMIKOM repository as ground truth provides a more objective and reproducible evaluation. These results suggest that multilingual SBERT shows promise in providing semantically relevant supervisor recommendations for bilingual academic datasets, though further evaluation on a larger dataset is warranted.

Keywords: Sentence-BERT; Cosine Similarity; Supervisor Recommendation; Final Project; Natural Language Processing

1. PENDAHULUAN

Menyelesaikan tugas akhir merupakan kewajiban yang harus dipenuhi oleh setiap mahasiswa sebagai salah satu persyaratan kelulusan pada jenjang pendidikan sarjana [1]. Namun, proses ini kerap menimbulkan berbagai kendala, terutama dalam hal penentuan topik penelitian dan pemilihan dosen pembimbing yang memiliki relevansi dengan bidang yang dikaji. Ketidakcocokan antara topik yang diangkat mahasiswa dengan keahlian dosen pembimbing yang ditunjuk merupakan salah satu penyebab utama terhambatnya penyelesaian tugas akhir, yang pada akhirnya berdampak pada mundurnya waktu kelulusan mahasiswa [2]. Dampak tersebut tidak hanya dirasakan dalam hal durasi studi, tetapi juga berpengaruh terhadap meningkatnya tekanan psikologis akademik, melemahnya semangat belajar, dan menurunnya kualitas proses bimbingan [2]. Dalam praktiknya, mahasiswa sering kesulitan menemukan dosen pembimbing yang tepat karena minimnya informasi yang tersedia mengenai bidang penelitian dan rekam jejak publikasi ilmiah para dosen.

Hingga saat ini, proses pemilihan dosen pembimbing di berbagai perguruan tinggi masih banyak yang dilakukan secara konvensional dan bersifat subjektif, yakni bergantung pada keputusan koordinator program studi tanpa didukung oleh sistem rekomendasi yang terstandar [3]. Kondisi ini mengakibatkan dosen yang ditugaskan sebagai pembimbing tidak selalu memiliki kompetensi yang relevan dengan topik penelitian mahasiswa, sehingga proses bimbingan pun menjadi tidak maksimal [4]. Di sisi lain, ketimpangan jumlah mahasiswa dengan jumlah dosen yang tersedia turut menyebabkan distribusi beban pembimbingan menjadi tidak seimbang. Persoalan ini semakin diperparah oleh terbatasnya komunikasi antara mahasiswa dan dosen sebelum proses bimbingan berlangsung, serta kurangnya pemahaman mahasiswa terhadap latar belakang dan fokus penelitian dosen [1], [4]. Oleh karena itu, diperlukan sebuah sistem yang mampu memfasilitasi proses penentuan dosen pembimbing secara lebih objektif dan terstruktur berdasarkan kesesuaian bidang penelitian.

Kemajuan teknologi kecerdasan buatan, khususnya di bidang *Natural Language Processing* (NLP), membuka peluang besar untuk mengatasi permasalahan tersebut [5], [6]. Melalui NLP, komputer dapat memahami, menganalisis, dan mengolah data berbentuk teks secara otomatis, sehingga memungkinkan pengukuran tingkat kesesuaian antara topik tugas akhir mahasiswa dan keahlian dosen berdasarkan publikasi ilmiah yang dimilikinya [5]. Salah satu pendekatan yang umum digunakan dalam NLP adalah *semantic similarity*, yaitu metode yang digunakan untuk mengukur kedekatan makna antara dua teks. Pendekatan ini dianggap lebih unggul dibandingkan pencocokan kata kunci secara harfiah karena mampu memahami konteks dan relasi semantik antar kalimat secara lebih mendalam. Dalam beberapa tahun terakhir, model berbasis arsitektur *transformer*, seperti *Bidirectional Encoder Representations from Transformers* (BERT), telah memperlihatkan performa yang sangat kompetitif dalam berbagai tugas NLP, termasuk *text similarity* dan *text classification* [7], [8].

Berbagai upaya telah dilakukan oleh peneliti sebelumnya untuk membangun sistem yang mampu merekomendasikan dosen pembimbing secara otomatis, dengan memanfaatkan teknik NLP dan *text mining* sebagai landasannya. Salah satunya adalah penelitian yang dilakukan oleh Hairani dan Mujahid [3], yang membangun sistem rekomendasi menggunakan kombinasi *cosine similarity* dan TF-IDF. Sistem yang dihasilkan mampu mencocokkan topik berdasarkan kesamaan leksikal, namun pendekatan TF-IDF secara inheren hanya mempertimbangkan frekuensi kata tanpa memperhitungkan makna di balik konteks kalimat secara utuh. Penelitian lainnya oleh Budiono dan Eniyati [4] menggunakan strategi *content-based filtering* yang mengandalkan riwayat penelitian dosen sebagai basis rekomendasi. Meski efektif dalam mencocokkan profil, pendekatan ini belum mengadopsi teknik representasi semantik berbasis *deep learning*, sehingga kemampuannya dalam mengenali kemiripan makna yang lebih abstrak masih terbatas.

Sabilillah et al. [7] mengembangkan sistem rekomendasi dosen pembimbing di Universitas Dian Nuswantoro dengan memadukan BERT dan *cosine similarity*, dan mencatatkan akurasi sebesar 90%. Namun demikian, model *bert-base-uncased* yang digunakan dalam penelitian tersebut pada dasarnya dioptimalkan untuk korpus berbahasa Inggris, sehingga kinerjanya berpotensi menurun ketika berhadapan dengan teks yang memadukan dua bahasa sekaligus, seperti bahasa Indonesia dan Inggris. Penelitian lain oleh Syabilla et al [9], mengintegrasikan LDA, TF-IDF, dan BERT dengan data publikasi dosen dari Google Scholar, memperoleh akurasi 80% meskipun masih terdapat keterbatasan pada nilai recall yang rendah. Di sisi lain, Hanif et al. [5] mengambil pendekatan berbeda dengan menggunakan NLP untuk mengkategorikan minat penelitian mahasiswa melalui teknik *text classification*, sementara Azhari dan Buulolo [10] memilih kombinasi *TextRank* dan *cosine similarity* sebagai fondasi sistem rekomendasi berbasis ekstraksi kata kunci. Meskipun sejumlah penelitian mutakhir mulai melirik penggunaan *sentence embedding* berbasis *transformer* untuk mengukur kemiripan semantik, penerapannya secara spesifik pada konteks pemilihan dosen pembimbing di perguruan tinggi dalam negeri masih belum banyak dieksplorasi.

Merujuk pada kajian penelitian terdahulu, terdapat beberapa kesenjangan penelitian yang belum sepenuhnya terselesaikan. Mayoritas penelitian sebelumnya masih mengandalkan pendekatan berbasis pencocokan kata kunci seperti TF-IDF, *TextRank*, dan *content-based filtering* konvensional yang kurang mampu menangkap hubungan semantik antar teks secara mendalam. Penelitian yang menggunakan BERT standar pun umumnya lebih berfokus pada teks berbahasa Inggris, sehingga performanya kurang memadai dalam mengolah dataset bilingual seperti publikasi ilmiah dosen di Indonesia yang umumnya menggunakan kombinasi bahasa Indonesia dan Inggris. Selain itu, sebagian besar penelitian hanya menitikberatkan pada evaluasi akurasi klasifikasi tanpa mengukur performa sistem rekomendasi dalam skenario *Top-N recommendation* yang lebih sesuai dengan kebutuhan nyata proses pemilihan dosen pembimbing.

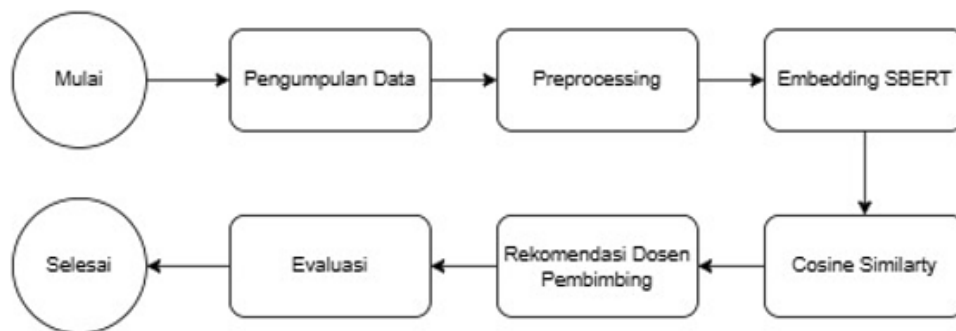
Penelitian ini berupaya menjawab kesenjangan tersebut dengan mengimplementasikan Sentence-BERT (SBERT) menggunakan model *paraphrase-multilingual-MiniLM-L12-v2* [11], [12]. Model ini dirancang secara khusus untuk tugas *semantic similarity* dan mendukung lebih dari 50 bahasa, termasuk bahasa Indonesia dan Inggris, sehingga lebih sesuai untuk menangani karakteristik dataset bilingual publikasi dosen di Indonesia. Berbeda dengan BERT standar yang memanfaatkan representasi token [CLS], SBERT menghasilkan *sentence embedding* melalui mekanisme *mean pooling* yang terbukti lebih representatif dalam mengukur tingkat kemiripan antar kalimat [11], [13]. Dataset yang digunakan dalam penelitian ini terdiri dari 1.420 judul penelitian yang berasal dari 43 dosen Program Studi Sistem Informasi, Universitas Amikom Yogyakarta, yang diperoleh melalui proses *web scraping* Google Scholar menggunakan SerpAPI [14].

Kontribusi utama penelitian ini terletak pada penerapan model SBERT multilingual sebagai dasar sistem rekomendasi dosen pembimbing berbasis *semantic similarity* pada dataset bilingual publikasi dosen di lingkungan perguruan tinggi Indonesia. Selain itu, penelitian ini turut mengevaluasi kinerja sistem menggunakan metrik *Hit Rate* dan *Mean Reciprocal Rank* (MRR) dalam skenario *Top-1*, *Top-2*, dan *Top-3 recommendation*, sehingga menghasilkan gambaran yang lebih menyeluruh mengenai kualitas rekomendasi yang dihasilkan oleh sistem. Secara keseluruhan, tujuan penelitian ini adalah membangun sistem rekomendasi dosen pembimbing di Universitas Amikom Yogyakarta berbasis metode SBERT dan *cosine similarity*, sekaligus mengevaluasi performanya berdasarkan relevansi rekomendasi yang dihasilkan. Penelitian ini diharapkan dapat menjadi solusi praktis bagi mahasiswa dalam menentukan dosen pembimbing yang sesuai secara lebih objektif, efektif, dan efisien, serta dapat menjadi referensi bagi pengembangan sistem rekomendasi akademik berbasis NLP di perguruan tinggi Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Secara keseluruhan, alur penelitian ini terdiri atas beberapa tahap yang disusun secara sistematis dan berurutan, seperti yang diilustrasikan pada Gambar 1. Setiap tahapan dirancang secara berurutan untuk memastikan proses pengembangan sistem rekomendasi dosen pembimbing berbasis *Sentence-BERT* (SBERT) dan *cosine similarity* berjalan secara terstruktur dan dapat direplikasi.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data diperoleh melalui proses *web scraping* pada Google Scholar dengan memanfaatkan layanan SerpAPI [14]. SerpAPI dipilih lantaran menawarkan antarmuka yang terstruktur dan handal, serta mampu mengekstrak data akademik secara otomatis tanpa membutuhkan autentikasi manual. Total data yang berhasil dihimpun adalah 1.420 judul publikasi dari 43 dosen Program Studi Sistem Informasi, Universitas Amikom Yogyakarta. Setiap entri dataset memuat dua atribut utama, yaitu nama dosen dan judul penelitian. Analisis distribusi bahasa menunjukkan bahwa dataset bersifat bilingual dengan komposisi sekitar 49% berbahasa Indonesia dan 42% berbahasa Inggris, serta 9% campuran atau tidak teridentifikasi. Karakteristik bilingual ini menjadi dasar pertimbangan pemilihan model SBERT multilingual pada tahap selanjutnya [12], [13].

2.3 Preprocessing

Tahap *preprocessing* bertujuan untuk membersihkan teks dari karakter yang tidak relevan sebelum diproses oleh model SBERT. *Preprocessing* yang diterapkan meliputi penghapusan teks di dalam kurung kurawal menggunakan *regular expression* (regex) dan normalisasi spasi berlebih. Tahapan seperti *case folding*, tokenisasi, *stopword removal*, *stemming*, dan *lemmatization* tidak diterapkan karena SBERT memproses kalimat secara utuh sehingga perubahan tersebut dapat memengaruhi representasi semantik kalimat [11].

2.4 Embedding SBERT

Model yang digunakan pada tahap ini adalah *Sentence-BERT* (SBERT) dengan model *paraphrase-multilingual-MiniLM-L12-v2* [11][12]. SBERT merupakan pengembangan dari arsitektur BERT yang menggunakan *Siamese network* dengan mekanisme *mean pooling* untuk menghasilkan representasi kalimat secara langsung [11]. Berbeda dengan BERT standar yang mengambil representasi dari token [CLS] pada *last hidden*

state, SBERT menghasilkan *sentence embedding* dengan merata-ratakan representasi seluruh token dalam kalimat sehingga menghasilkan vektor yang lebih representatif untuk tugas *sentence similarity* [13].

Model *paraphrase-multilingual-MiniLM-L12-v2* dipilih berdasarkan tiga pertimbangan utama. Pertama, model ini dirancang khusus untuk tugas *sentence similarity* sehingga lebih optimal dibandingkan BERT standar untuk kasus rekomendasi berbasis kemiripan teks [11]. Kedua, model ini mendukung lebih dari 50 bahasa termasuk Indonesia dan Inggris melalui teknik *knowledge distillation* sehingga sesuai untuk dataset bilingual [12]. Ketiga, model ini menghasilkan *sentence embedding* berukuran 384 dimensi yang lebih efisien dibandingkan BERT standar (768 dimensi) tanpa mengorbankan kualitas representasi semantik [13]. Proses *embedding* dilakukan secara *batch* menggunakan fungsi *encode()* yang disediakan oleh library *sentence-transformers* [12], [15].

2.5 Cosine Similarity

Setelah seluruh vektor *embedding* judul penelitian dosen terbentuk, proses dilanjutkan dengan penghitungan *cosine similarity* antara *embedding* judul tugas akhir mahasiswa dan *embedding* masing-masing judul penelitian dosen [16], [17]. *Cosine similarity* digunakan untuk mengukur kedekatan dua vektor berdasarkan nilai kosinus sudut yang terbentuk di antara keduanya, sebagaimana diformulasikan pada Persamaan (1) [16].

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

Nilai yang dihasilkan berada pada rentang 0 hingga 1, di mana semakin mendekati angka 1 maka semakin tinggi pula kemiripan semantik antara judul tugas akhir mahasiswa dengan penelitian dosen yang bersangkutan [17]. Untuk setiap dosen, diambil nilai *cosine similarity* tertinggi dari seluruh judulnya sebagai skor representatif, kemudian diurutkan secara menurun untuk menghasilkan Top-N rekomendasi [3], [18].

2.6 Evaluasi

Evaluasi sistem dilakukan menggunakan 10 data uji berupa judul tugas akhir nyata yang diperoleh dari repositori EPrints AMIKOM. Setiap data uji memiliki dosen pembimbing aktual yang digunakan sebagai ground truth untuk mengevaluasi hasil rekomendasi sistem.

Metrik evaluasi yang digunakan dalam penelitian ini adalah Hit Rate dan Mean Reciprocal Rank (MRR). Hit Rate digunakan untuk mengukur tingkat keberhasilan sistem dalam menemukan dosen yang sesuai pada daftar rekomendasi Top-N, sedangkan MRR digunakan untuk mengevaluasi kualitas peringkat rekomendasi berdasarkan posisi dosen yang relevan dalam daftar rekomendasi [17].

Rumus Hit Rate ditunjukkan pada Persamaan (2)

$$\text{Hit Rate}@K = \frac{\text{Jumlah rekomendasi berhasil}}{\text{Jumlah data uji}} \quad (2)$$

Sedangkan rumus Mean Reciprocal Rank (MRR) ditunjukkan pada Persamaan (3)

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \quad (3)$$

Keterangan:

Q : jumlah seluruh data uji

rank_i : posisi dosen aktual pada daftar rekomendasi untuk data uji ke- i

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Pengumpulan data dilakukan melalui *web scraping*. Proses akuisisi data menghasilkan total 1.420 judul publikasi ilmiah dari 43 dosen Program Studi Sistem Informasi, Universitas Amikom Yogyakarta. Setiap entri memuat nama dosen beserta judul penelitian yang selanjutnya dijadikan basis pengetahuan sistem rekomendasi. Sebagian contoh data hasil akuisisi ditunjukkan pada Tabel 1.

Tabel 1. Contoh Dataset Hasil *Web Scraping*

Nama Dosen	Hasil
Acihmah Sidauruk	Penerapan Teknologi Smart Security IoT untuk Peningkatan Lingkungan di Padukuhan Mlati, Yogyakarta, Optimization and Collaboration...
Aditya Rizki Yudiantika	The development of mobile augmented reality quiz visualization methods based on markerless tracking for museum...
Afrig Aminuddin	Benchmark of Machine Learning Classifiers for Mental Health Prediction in the Tech Sector, Machine Learning...
Agung Nugroho	Sistem Pelaporan Data Capaian Standar Pelayanan Minimal Bidang Kesehatan Berbasis Website, Sensitivity Testing in Decision...
Agus Fatkhurohman	Virtual Reality Exposure Therapy as a Novel Approach to Acrophobia Treatment, The Effect of Augmentation...

Berdasarkan analisis distribusi bahasa menggunakan library *langdetect*, dataset yang terkumpul bersifat bilingual dengan komposisi sekitar 49% berbahasa Indonesia dan 42% berbahasa Inggris, serta 9% campuran atau tidak teridentifikasi. Karakteristik bilingual ini menjadi dasar utama pemilihan model SBERT multilingual dalam penelitian ini.

3.2 Hasil *Preprocessing*

Tahap *preprocessing* dilakukan untuk membersihkan karakter tidak relevan dari teks sebelum diproses oleh model SBERT. Pendekatan yang diterapkan bersifat minimal, yaitu penghapusan karakter khusus berbasis *regex* dan normalisasi spasi berlebih. Contoh hasil *preprocessing* ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil *Preprocessing* Step by Step

Tahap	Hasil
Awal	{JALUR MAGANG IT} PERANCANGAN USER INTERFACE DAN USER EXPERIENCE APLIKASI MOBILE DAUR ULANG SAMPAH MENGGUNAKAN METODE DESIGN THINKING
Regex	PERANCANGAN USER INTERFACE DAN USER EXPERIENCE APLIKASI MOBILE DAUR ULANG SAMPAH MENGGUNAKAN METODE DESIGN THINKING
Whitespace Normalization	PERANCANGAN USER INTERFACE DAN USER EXPERIENCE APLIKASI MOBILE DAUR ULANG SAMPAH MENGGUNAKAN METODE DESIGN THINKING
Final	PERANCANGAN USER INTERFACE DAN USER EXPERIENCE APLIKASI MOBILE DAUR ULANG SAMPAH MENGGUNAKAN METODE DESIGN THINKING

Hasil *preprocessing* menunjukkan bahwa proses pembersihan data berjalan dengan baik. Pendekatan *preprocessing* minimal berbasis *regex* ini sesuai dengan karakteristik SBERT yang memproses kalimat secara

utuh, sehingga normalisasi leksikal lanjutan seperti tokenisasi, penghapusan *stopword*, *stemming*, maupun *lemmatization* tidak diperlukan. Pendekatan ini juga mempertahankan bentuk asli istilah teknis seperti nama metode (*naive bayes*, *cosine similarity*, *neural network*) yang menjadi penanda semantik penting dalam pengukuran kemiripan antar judul penelitian [7], [18].

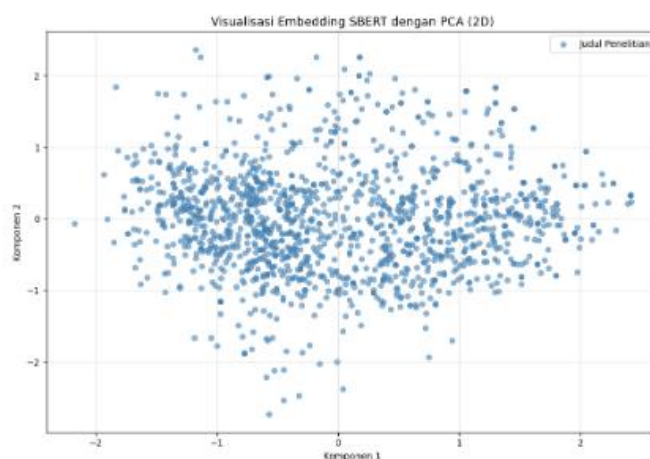
3.3 Hasil Visualisasi

Visualisasi *word cloud* dilakukan untuk memetakan kata-kata yang paling sering muncul dalam judul penelitian dosen setelah tahap *preprocessing*. Visualisasi ini bertujuan memberikan gambaran umum mengenai topik-topik penelitian yang paling banyak ditekuni oleh dosen di program studi ini. Selain itu, dilakukan pula reduksi dimensi terhadap vektor *embedding* SBERT menggunakan metode PCA 2D untuk memperlihatkan pola distribusi dan pengelompokan judul penelitian berdasarkan kemiripan semantiknya.



Gambar 2. Visualisasi Word Cloud Judul Penelitian Dosen

Hasil *word cloud* pada Gambar 2 memperlihatkan kata-kata yang paling sering muncul dalam judul penelitian dosen setelah tahap *preprocessing*. Istilah-istilah seperti "using", "based", "sistem informasi", "machine learning", dan "aplikasi" tampak mendominasi dengan ukuran terbesar, yang mencerminkan bahwa sebagian besar penelitian dosen berfokus pada pengembangan sistem dan penerapan algoritma berbasis teknologi informasi.



Gambar 3. Visualisasi Embedding SBERT dengan PCA (2D)

Gambar 3 menampilkan hasil visualisasi *embedding* SBERT menggunakan reduksi dimensi PCA 2D. Sebaran titik-titik pada grafik memperlihatkan distribusi judul penelitian yang cukup merata di seluruh ruang vektor berdimensi rendah, tanpa pembentukan kluster yang tegas. Hal ini mengindikasikan bahwa topik-topik penelitian dosen cukup beragam, meskipun sebagian titik yang berdekatan masih mencerminkan adanya kemiripan semantik antar judul penelitian tertentu.

3.4 Hasil Pengujian Rekomendasi

Sebagai ilustrasi cara kerja sistem, apabila mahasiswa memasukkan judul tugas akhir "*Implementasi Deep Learning untuk Deteksi Objek pada Citra Digital*".

Tabel 3. Contoh Hasil Rekomendasi Sistem

No	Nama Dosen	Cosine Similarity
1	Anggit Dwi Hartanto	0.710
2	Norhikmah	0.674
3	Sumarni Adi	0.620

Hasil pada Tabel 3 menunjukkan bahwa sistem merekomendasikan Anggit Dwi Hartanto sebagai dosen pembimbing utama dengan nilai *cosine similarity* tertinggi sebesar 0.710. Nilai *cosine similarity* yang relatif berdekatan antar rekomendasi menunjukkan bahwa beberapa dosen memiliki topik penelitian yang serupa. Meskipun demikian, model SBERT tetap mampu menempatkan dosen dengan relevansi semantik tertinggi pada posisi pertama rekomendasi. Hasil pengujian Top-3 pada 10 data uji ditunjukkan pada Tabel 4. Data uji menggunakan judul tugas akhir nyata dari repositori EPrints AMIKOM beserta dosen pembimbing aktual sebagai ground truth. Sistem berhasil merekomendasikan dosen aktual pada 9 dari 10 data uji sehingga memperoleh Hit Rate sebesar 90% pada skenario Top-3.

Tabel 4. Hasil Pengujian Data (Top-3)

No	Judul	Rekomendasi	Cosine	Aktual	Ket
1	ANALISIS SENTIMEN MASYARAKAT INDONESIA TER..	1. Mei Parwanto Kurniawan 2. Anggit Dwi Hartanto 3. Arif Nur Rohman	0.642 0.638 0.638	Anggit Dwi Hartanto	hit
2	EVALUASI MODEL MACHINE LEARNING BERBASIS E..	1. Norhikmah 2. Arif Nur Rohman 3. Yoga Pristyanto	0.734 0.647 0.634	Yoga Pristyanto	hit
3	ANALISIS KINERJA ALGORITMA SUPPORT VECTOR	1. Norhikmah 2. Yoga Pristyanto 3. Acihmah Sidaauruk	0.707 0.702 0.702	Yoga Pristyanto	hit
4	COMPARISON OF ENSEMBLE METHODS FOR DECISIO...	1. Yoga Pristyanto 2. Marwan Noor Fauzy 3. Afrig Aminuddin	0.650 0.560 0.530	Yoga Pristyanto	hit
5	PENGEMBANGAN APLIKASI VIRTUAL REALITY UNTU...	1. Kardilah Rohmat Hidayat 2. Ika Asti Astuti 3. Norhikmah	0.751 0.742 0.704	Ika Asti Astuti	hit
6	PERANCANGAN DAN PENGEMBANGAN APLIKASI ANDR..	1. Bety Wulan Sari 2. Ninik Tri Hartanti 3. Ika Asti Astuti	0.538 0.515 0.511	Ika Asti Astuti	hit
7	IMPLEMENTASI METODE NAIVE BAYES UNTUK KLAS...	1. Ika Nur Fajri 2. Yoga Pristyanto 3. Bety Wulan Sari	0.655 0.596 0.596	Ika Nur Fajri	hit
8	PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KN...	1. Ika Nur Fajri 2. Anggit Dwi Hartanto 3. Ninik Tri Hartanti	0.761 0.647 0.642	Ika Nur Fajri	hit
9	PEMBUATAN ANIMASI 3D IKLAN PRODUK TWS VH-S..	1. Mei Parwanto Kurniawan 2. Norhikmah 3. Alfie Nur Rahmi	0.647 0.612 0.586	Ika Nur Fajri	miss
10	SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN KARYA..	1. Sharazita Dyah Anggita 2. Arif Nur Rohman 3. Agus Fatkhurohman	0.601 0.587 0.502	Sharazita Dyah Anggita	hit

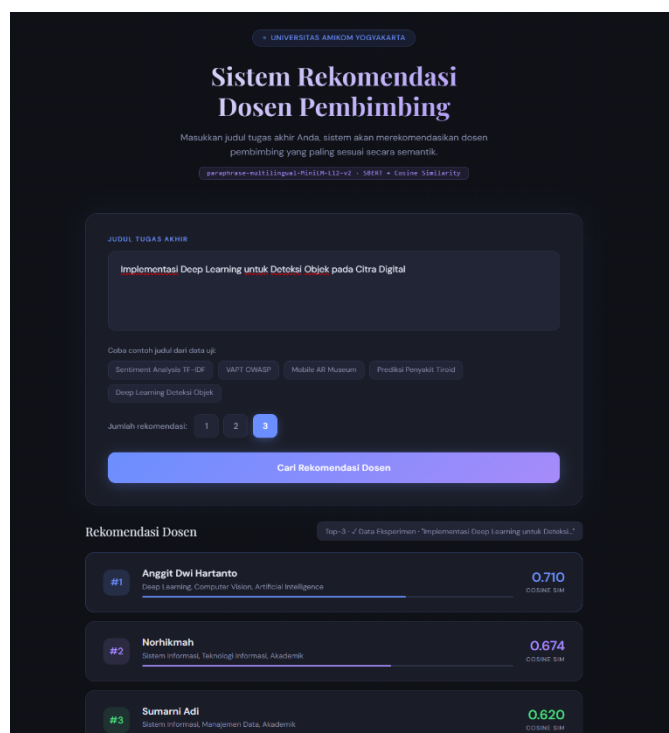
Nilai cosine similarity yang diperoleh berada pada rentang 0.511 hingga 0.761, menunjukkan bahwa model SBERT mampu mengukur kemiripan semantik secara realistis tanpa pencocokan kata secara literal. Pada data uji

No. 6, meskipun selisih nilai antar rekomendasi relatif tipis (0.538, 0.515, 0.511), sistem tetap berhasil menempatkan dosen aktual dalam Top-3.

Dibandingkan dengan penelitian Sabilillah *et al.* [7] yang menggunakan BERT standar (*bert-base-uncased*) dengan akurasi 90%, penelitian ini mencapai *Hit Rate* 90% pada Top-3 menggunakan data uji *ground truth* nyata dari repositori EPrints AMIKOM. Perlu dicatat bahwa perbandingan ini bersifat kontekstual karena metrik yang digunakan berbeda — akurasi klasifikasi vs. *Hit Rate* dan MRR dalam skenario *Top-N recommendation*. Nilai MRR 0.617 mengindikasikan ruang peningkatan pada kualitas pemeringkatan, yang dapat dieksplorasi melalui *threshold cosine similarity* atau metode *re-ranking*.

3.5 Hasil Implementasi Web

Implementasi sistem dilakukan dalam bentuk aplikasi berbasis web yang bertujuan memudahkan mahasiswa memperoleh rekomendasi dosen pembimbing secara otomatis berdasarkan judul tugas akhir yang dimasukkan. Sistem memanfaatkan model *Sentence-BERT* dan *cosine similarity* untuk menghitung tingkat kemiripan semantik antara judul tugas akhir mahasiswa dan bidang keahlian dosen.



Gambar 4. Tampilan Antarmuka Sistem Rekomendasi Dosen Pembimbing

Gambar 4 menunjukkan tampilan antarmuka sistem rekomendasi dosen pembimbing. Pada halaman utama tersedia kolom input judul tugas akhir, pilihan jumlah rekomendasi dosen, serta tombol untuk menjalankan proses rekomendasi.

Setelah proses dilakukan, sistem menampilkan hasil rekomendasi berupa nama dosen, bidang keahlian, nilai *cosine similarity*, dan progress bar visual untuk mempermudah pengguna membandingkan tingkat relevansi antar dosen. Antarmuka dirancang sederhana dan interaktif sehingga memudahkan mahasiswa dalam menggunakan sistem secara cepat dan informatif.

Sistem dirancang dengan alur penggunaan yang sederhana agar mudah diakses oleh mahasiswa tanpa memerlukan pemahaman teknis terkait metode NLP maupun *machine learning*. Pengguna hanya perlu memasukkan judul tugas akhir pada kolom input, menentukan jumlah rekomendasi yang diinginkan, kemudian menekan tombol proses untuk memperoleh hasil rekomendasi secara *real-time*. Seluruh proses perhitungan *embedding* menggunakan model SBERT dan *cosine similarity* dilakukan di sisi backend sehingga pengguna hanya berinteraksi dengan antarmuka yang sederhana dan responsif. Implementasi berbasis web ini juga memberikan fleksibilitas karena sistem dapat diakses melalui berbagai perangkat seperti laptop maupun smartphone selama terhubung dengan internet. Dengan demikian, sistem tidak hanya berfungsi sebagai prototipe penelitian, tetapi juga memiliki potensi untuk diintegrasikan ke dalam layanan akademik program studi sebagai alat bantu objektif dalam proses penentuan dosen pembimbing.

Selain sebagai media implementasi model, aplikasi web ini berperan sebagai sarana visualisasi hasil rekomendasi agar proses pencocokan dosen pembimbing menjadi lebih transparan. Mahasiswa dapat melihat daftar dosen yang direkomendasikan beserta tingkat kemiripan masing-masing sehingga proses pemilihan pembimbing tidak lagi hanya berdasarkan preferensi subjektif atau rekomendasi informal. Dengan adanya

visualisasi skor similarity, pengguna memperoleh gambaran yang lebih objektif mengenai relevansi topik penelitian dengan bidang keahlian dosen yang tersedia.

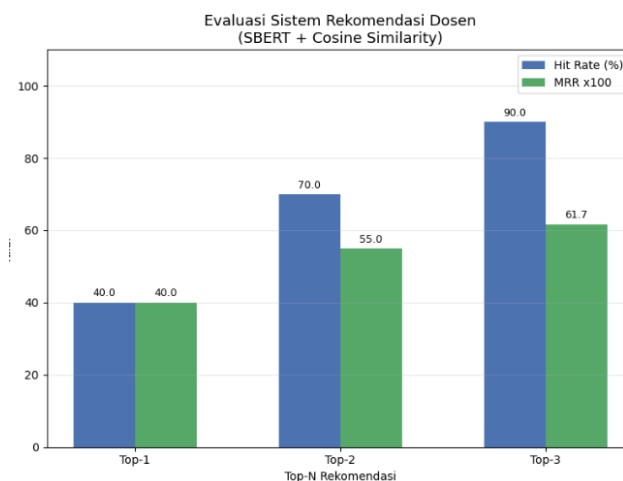
3.6 Hasil Evaluasi

Evaluasi sistem dilakukan pada tiga variasi rekomendasi yaitu Top-1, Top-2, dan Top-3 menggunakan metrik Hit Rate dan Mean Reciprocal Rank (MRR). Ringkasan hasil evaluasi ditunjukkan pada Tabel 5 dan divisualisasikan pada Gambar 5.

Tabel 5. Ringkasan Hasil Evaluasi

TOP N	Benar/total	Hit Rate	MRR
Top-1	4/10	40%	0.400
Top-2	7/10	70%	0.550
Top-3	9/10	90%	0.617

Berdasarkan Tabel 5, performa sistem meningkat seiring bertambahnya nilai Top-*N*. Pada Top-1 diperoleh *Hit Rate* sebesar 40% dan *MRR* 0.400, sedangkan pada Top-2 meningkat menjadi 70% dengan *MRR* 0.550. Performa terbaik diperoleh pada Top-3 dengan *Hit Rate* 90% dan *MRR* 0.617, yang menunjukkan bahwa 9 dari 10 dosen aktual berhasil direkomendasikan dalam tiga peringkat teratas, meskipun belum selalu berada pada posisi pertama. Satu data uji yang belum berhasil ditemukan pada Top-3 menunjukkan bahwa sistem masih memiliki keterbatasan dalam merekomendasikan dosen untuk topik penelitian tertentu. Hasil ini menunjukkan bahwa model SBERT mampu menangkap kemiripan semantik antar judul penelitian, namun kualitas pemeringkatan masih dapat ditingkatkan.



Gambar 5. Grafik Evaluasi Sistem Rekomendasi Dosen (SBERT + Cosine Similarity)

Gambar 5 memperlihatkan tren peningkatan performa sistem secara konsisten pada kedua metrik evaluasi seiring bertambahnya jumlah rekomendasi (Top-*N*). Pada skenario Top-1, sistem memperoleh Hit Rate sebesar 40% dan MRR sebesar 0.400. Nilai yang sama antara Hit Rate dan MRR pada Top-1 menunjukkan bahwa seluruh rekomendasi yang berhasil ditemukan langsung berada pada posisi pertama, sehingga reciprocal rank bernilai maksimal. Ketika jumlah rekomendasi ditingkatkan menjadi Top-2, Hit Rate meningkat menjadi 70% dan MRR menjadi 0.550. Peningkatan ini menunjukkan bahwa sistem mulai mampu menemukan dosen aktual yang sebelumnya tidak muncul pada Top-1, meskipun sebagian berada pada posisi kedua sehingga nilai MRR tidak meningkat secara proporsional dengan Hit Rate.

Pada skenario Top-3, sistem mencapai Hit Rate 90% dengan MRR sebesar 0.617. Hasil ini menunjukkan bahwa 9 dari 10 dosen aktual berhasil ditemukan dalam tiga rekomendasi teratas, namun belum seluruhnya menempati posisi pertama. Selisih antara Hit Rate dan MRR pada Top-3 mengindikasikan bahwa kualitas ranking sistem masih dapat ditingkatkan, terutama untuk mengoptimalkan penempatan dosen yang paling relevan pada urutan pertama rekomendasi. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa kombinasi SBERT multilingual dan *cosine similarity* cukup efektif digunakan untuk sistem rekomendasi dosen pembimbing pada dataset bilingual Indonesia-Inggris, khususnya dalam skenario pemberian beberapa alternatif rekomendasi kepada mahasiswa.

4. KESIMPULAN

Penelitian ini berhasil membangun sistem rekomendasi dosen pembimbing berbasis kemiripan semantik menggunakan model *Sentence-BERT (SBERT) paraphrase-multilingual-MiniLM-L12-v2* dikombinasikan dengan *cosine similarity* di lingkungan Universitas Amikom Yogyakarta. Sistem dibangun dari 1.420 judul publikasi ilmiah milik 43 dosen Program Studi Sistem Informasi yang diperoleh melalui *web scraping* Google Scholar menggunakan SerpAPI. Evaluasi menggunakan 10 data uji berupa judul tugas akhir nyata dari repositori EPrints AMIKOM dengan dosen pembimbing aktual sebagai *ground truth* menghasilkan *Hit Rate* 90% pada Top-3 dan MRR 0.617, menunjukkan bahwa sistem mampu merekomendasikan dosen yang relevan secara semantik. Kontribusi utama penelitian ini adalah penerapan SBERT multilingual yang mampu memproses dataset bilingual Indonesia–Inggris secara langsung tanpa translasi, serta adopsi evaluasi berlapis (*Hit Rate* dan MRR pada tiga variasi Top-N) yang menghasilkan penilaian lebih menyeluruh dibandingkan penelitian terdahulu yang umumnya hanya melaporkan satu metrik. Keterbatasan penelitian ini terletak pada jumlah data uji yang terbatas (10 sampel) sehingga generalisasi hasil perlu dikonfirmasi dengan data yang lebih besar. Selain itu, nilai *cosine similarity* yang rendah pada beberapa posisi rekomendasi (misalnya 0.515 dan 0.511 pada data uji No. 6) mengindikasikan perlunya penerapan *threshold* minimum agar rekomendasi yang dihasilkan benar-benar bermakna secara semantik. Untuk penelitian selanjutnya disarankan memperluas dataset dari sumber SINTA, mengeksplorasi model *fine-tuned* SBERT pada domain akademik Indonesia, serta mengintegrasikan sistem dengan platform akademik perguruan tinggi agar dapat dimanfaatkan langsung oleh mahasiswa dan program studi.

REFERENCES

- [1] R. Megawati dan M. Damayanti, “Peran Dosen Pembimbing Skripsi dalam Proses Penyelesaian Tugas Akhir Mahasiswa,” *J-HEST Journal of Health Education Economics Science and Technology*, vol. 4, no. 1, hlm. 33–39, Agu 2022, doi: 10.36339/jhest.v4i1.63.
- [2] R. T. Sugiharno, W. H. Ari Susanto, dan F. Wospakrik, “Faktor-Faktor yang Mempengaruhi Kecemasan Mahasiswa dalam Menghadapi Tugas Akhir,” *Jurnal Keperawatan Silampari*, vol. 5, no. 2, hlm. 1189–1197, Jun 2022, doi: 10.31539/jks.v5i2.3760.
- [3] H. Hairani dan M. Mujahid, “Recommendations of Thesis Supervisor using the Cosine Similarity Method,” *SISTEMASI*, vol. 11, no. 3, hlm. 646, Sep 2022, doi: 10.32520/stmsi.v11i3.2003.
- [4] A. Budiono dan S. Eniyati, “Sistem Rekomendasi Dosen Pembimbing Tugas Akhir Menggunakan Content Based Filtering,” *Jurnal Elektronik dan Komputer*, vol. 16, no. 1, hlm. 64–71, Jul 2023, doi: 10.51903/elkom.v16i1.1002.
- [5] A. Jimly Hanif, M. Nur Farid, dan B. Hasanah, “Penerapan Natural Language Processing untuk Klasifikasi Bidang Minat berdasarkan Judul Tugas Akhir,” *Jurnal Sistim Informasi dan Teknologi*, hlm. 41–49, Feb 2023, doi: 10.37034/jsisfotek.v5i1.196.
- [6] I. R. Hendrawan dan E. Utami, *Natural Language Processing*. Yogyakarta: Penerbit Andi, 2023.
- [7] F. T. Sabilillah, S. Winarno, dan R. B. Abiyyi, “Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, hlm. 585–594, Des 2024, doi: 10.29408/edumatic.v8i2.27791.
- [8] B. Kurniawan, A. A. Aldino, dan A. R. Isnain, “Sentimen analisis terhadap kebijakan penyelenggara sistem elektronik (Pse) menggunakan algoritma bidirectional encoder representations from transformers (Bert),” *J. Teknol. dan Sist. Inf*, vol. 3, no. 4, hlm. 98–106, Des 2022.
- [9] M. Syabilla, J. Zeniarja, dan Q. Nabila, “Implementasi LDA, TF-IDF, dan BERT dalam Sistem Rekomendasi Dosen Pembimbing untuk Mahasiswa,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, Mar 2025, doi: 10.47065/bits.v6i4.6499.
- [10] A. A. B. E. dan S. N., “Sistem Rekomendasi Dosen Pendamping Skripsi Berbasis Text Rank menggunakan Metode Cosine Similarity,” *Pelita Informatika : Informasi dan Informatika*, vol. 10, no. 3, hlm. 119–122, 2022.
- [11] J. Seo, S. Lee, L. Liu, dan W. Choi, “TA-SBERT: Token Attention Sentence-BERT for Improving Sentence Representation,” *IEEE Access*, vol. 10, hlm. 39119–39128, 2022, doi: 10.1109/ACCESS.2022.3164769.
- [12] R. M. Holis, P. E. P. Utomo, dan B. F. Hutabarat, “Semantic FAQ Chatbot Using SBERT (Sentence-BERT) and Cosine Similarity for Academic Services,” *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, hlm. 915–922, Okt 2025, doi: 10.47709/brilliance.v5i2.7027.
- [13] K. Venkatesh Sharma, P. R. Ayiluri, R. Betala, P. Jagdish Kumar, dan K. Shirisha Reddy, “Enhancing query relevance: leveraging SBERT and cosine similarity for optimal information retrieval,” *Int. J. Speech Technol.*, vol. 27, no. 3, hlm. 753–763, Sep 2024, doi: 10.1007/s10772-024-10133-5.
- [14] SerpApi, “Google Scholar API,” SerpApi. Diakses: 16 Mei 2026. [Daring]. Tersedia pada: <https://serpapi.com/google-scholar-api>
- [15] B. Raharjo, *Deep Learning dengan Python*. Semarang: Penerbit Yayasan Prima Agus Teknik, 2022.
- [16] A. Widiyanto, E. Pebriyanto, F. Fitriyanti, dan M. Marna, “Document Similarity Using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity,” *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, vol. 4, no. 2, hlm. 149–153, Agu 2024, doi: 10.20895/dinda.v4i2.1589.
- [17] Ylber Januzaj dan Artan Luma, “Cosine Similarity – A Computing Approach to Match Similarity Between Higher Education Programs and Job Market Demands Based on Maximum Number of Common Words,” *International*



- Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 12, hlm. 258–268, Jun 2022, doi: 10.3991/ijet.v17i12.30375.
- [18] A. Apriani, H. Zakiyudin, dan K. Marzuki, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta,” *Jurnal Bumigora Information Technology (BITE)*, vol. 3, no. 1, hlm. 19–27, Jul 2021, doi: 10.30812/bite.v3i1.1110.