

Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir

Ferris Tita Sabilillah^{1,*}, Sri Winarno¹, Ryandhika Bintang Abiyyi¹

¹ Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Indonesia

* Correspondence: ferristita1606@gmail.com

Copyright: © 2024 by the authors

Received: 13 Oktober 2024 | Revised: 19 Oktober 2024 | Accepted: 11 November 2024 | Published: 19 Desember 2024

Abstrak

Permasalahan dalam penyelesaian tugas akhir, yang sering kali berkontribusi terhadap keterlambatan kelulusan mahasiswa, kerap diakibatkan oleh ketidaksesuaian antara topik penelitian mahasiswa dan keahlian dosen pembimbing. Oleh karena itu diperlukan metode untuk mengatasi ketidaksesuaian tersebut dalam menyelesaikan tugas akhir. Penelitian ini bertujuan untuk mengimplementasikan model *Bidirectional Encoder Representations from Transformers* (BERT) dan *cosine similarity* guna merekomendasikan dosen pembimbing berdasarkan judul tugas akhir mahasiswa. Dataset penelitian mencakup 3.723 judul penelitian yang diperoleh melalui web scraping dari Google Scholar dan ResearchGate, mewakili keahlian 63 dosen di Program Studi Teknik Informatika, Universitas Dian Nuswantoro. Tahap pengolahan data meliputi preprocessing untuk menghasilkan embedding vektor dari judul penelitian dosen, yang kemudian dibandingkan dengan judul tugas akhir mahasiswa. Temuan kami menunjukkan bahwa model rekomendasi yang dikembangkan memiliki tingkat akurasi sebesar 90% dalam mengidentifikasi dosen pembimbing yang relevan berdasarkan kesesuaian topik antara judul tugas akhir mahasiswa dan bidang keahlian dosen yang tergambar dari publikasi mereka. Hasil ini dapat memberikan kontribusi signifikan dalam mendukung mahasiswa menyelesaikan tugas akhir secara lebih efisien, serta meningkatkan kualitas bimbingan akademik melalui pemilihan dosen pembimbing yang lebih tepat.

Kata kunci: *bert*; *cosine similarity*; rekomendasi pembimbing; tugas akhir; *natural language processing* (nlp)

Abstract

Challenges in completing final projects, which often contribute to delays in student graduation, are frequently due to a mismatch between students' research topics and the expertise of their supervisors. Therefore, a method is needed to address this misalignment in the final project process. This study aims to implement a *Bidirectional Encoder Representations from Transformers* (BERT) model and *cosine similarity* to recommend supervisors based on students' final project titles. The research dataset includes 3,723 research titles collected through web scraping from Google Scholar and ResearchGate, representing the expertise of 63 lecturers in the Informatics Engineering Program at Universitas Dian Nuswantoro. Data processing includes preprocessing to generate embedding vectors from lecturers' research titles, which are then compared with students' final project titles. Our findings indicate that the developed recommendation model achieves an accuracy of 90% in identifying relevant supervisors based on topic alignment between students' final project titles and lecturers' areas of expertise, as reflected in their publications. This result can make a significant contribution to supporting students in completing their final projects more efficiently and improving the quality of academic supervision by facilitating more appropriate supervisor selection.

Keywords: *bert*; *cosine similarity*; supervisor recommendation; final project; *natural language processing* (nlp)



PENDAHULUAN

Kompetisi global yang semakin ketat di era modern memaksa pendidikan tinggi untuk terus meningkatkan performa dan daya saingnya. Tantangan tersebut tidak hanya terkait dengan kualitas pengajaran dan penelitian, tetapi juga pada efisiensi proses akademik yang melibatkan mahasiswa dan dosen. Untuk mendapatkan gelar sarjana strata satu, mahasiswa program sarjana harus menyelesaikan tugas akhir (Andriani & Wibawanta, 2020). Pemilihan topik tugas akhir adalah fase penting dalam pendidikan tinggi yang sering menjadi tantangan bagi mahasiswa. Kesulitan dalam menentukan judul tugas akhir secara tepat dapat mengakibatkan penundaan bimbingan, yang berujung pada penurunan performa akademik dan keterlambatan kelulusan (Hanif et al., 2023). Tingkat kecemasan yang tinggi pada mahasiswa dapat mempengaruhi performa akademik serta menunda kelulusan (Sugiharno et al., 2022). Faktor-faktor seperti keterbatasan informasi tentang keahlian dosen, banyaknya jumlah mahasiswa yang membutuhkan bimbingan, serta keterbatasan akses atau komunikasi antara mahasiswa dan dosen turut berkontribusi pada kesulitan menemukan dosen pembimbing yang sesuai. Dalam situasi ini, peran dosen pembimbing sangatlah signifikan, terutama dalam membimbing mahasiswa menuju arah penelitian yang tepat dan sesuai dengan keahlian mereka (Budiono & Eniyati, 2023). Keahlian dosen yang sesuai dengan topik tugas akhir berpengaruh besar pada kualitas bimbingan dan penyelesaian tugas akhir. Selain bimbingan, mahasiswa juga membutuhkan dosen penguji untuk menilai karya ilmiah mereka (Damanik, 2022). Sayangnya, pemilihan dosen pembimbing seringkali dilakukan secara subjektif, sehingga mengakibatkan ketidakcocokan antara topik penelitian mahasiswa dengan keahlian dosen, yang pada akhirnya menghambat efektivitas dan efisiensi penyelesaian tugas akhir (Asfi & Fitrianiingsih, 2020).

Kemajuan teknologi menawarkan berbagai solusi untuk meningkatkan daya saing dan efisiensi pendidikan tinggi, termasuk dalam penentuan dosen pembimbing tugas akhir yang memiliki peran penting dalam meningkatkan efisiensi di berbagai bidang, termasuk pendidikan (Mustaqim et al., 2024). *Natural Language Processing* (NLP) adalah proses analisis teks secara otomatis menggunakan komputer (Khoirunisa, 2020). Sebagai cabang dari kecerdasan buatan (AI), NLP memungkinkan komputer untuk menganalisis dan memahami bahasa manusia (Aprilinda et al., 2022). Salah satu teknik pengumpulan data, yaitu *Web Scraping*, memungkinkan pengumpulan dataset yang relevan secara otomatis dari situs web (Sapanji et al., 2023). Selanjutnya, data tersebut diproses melalui teknik *preprocessing*, yang merupakan langkah penting dalam mempersiapkan data mentah sebelum dilakukan analisis lebih lanjut (Apriani et al., 2021; Aqila & Bijaksana, 2020). Proses *preprocessing* ini sangat penting untuk mengubah teks menjadi bentuk yang dapat dianalisis lebih lanjut, seperti untuk perhitungan kemiripan atau penggunaan algoritma pembelajaran mesin lainnya.

BERT menjadi salah satu model yang paling populer dan efektif dalam tugas pemrosesan bahasa alami, terutama karena kemampuannya untuk memahami konteks kata secara mendalam melalui pendekatan *bidirectional* (Kurniawan et al., 2022; Samosir et al., 2022). BERT sangat efektif untuk analisis teks yang kompleks, seperti judul tugas akhir dan abstrak penelitian, karena mampu memahami konteks yang mendalam dibandingkan *TextRank* yang lebih cocok untuk rangkuman teks. Untuk meningkatkan proses pencocokan antara mahasiswa dan dosen, cosine similarity digunakan untuk mengukur kesamaan antara teks judul tugas akhir dengan bidang penelitian dosen (Azhari & Buulolo, 2022). Dengan pendekatan kombinasi antara BERT dan *cosine similarity*, penelitian ini diharapkan tidak hanya merekomendasikan dosen secara efisien tetapi juga memberikan penilaian yang lebih mendalam tentang kesesuaian antara mahasiswa dan dosen.

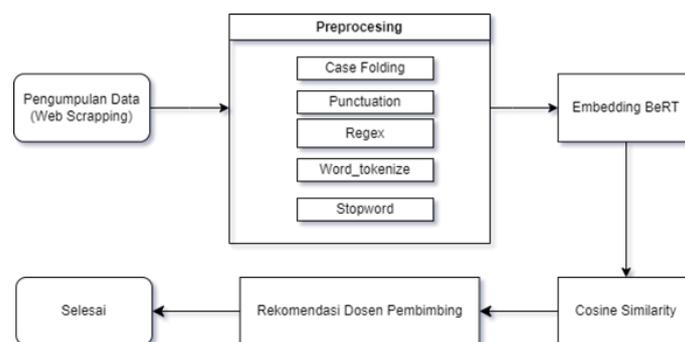
Penelitian sebelumnya telah mengaplikasikan *cosine similarity* dalam berbagai konteks, seperti sistem pengajuan judul tugas akhir (Azhari & Buulolo, 2022) dan pengembangan sistem rekomendasi dosen berbasis *TextRank* (Hairani & Mujahid, 2022). Selain itu, penelitian lain (Braja & Kodar, 2023) menggunakan model *pre-training* IndoBERTBASE dan BERTBASE

Multilingual. Penelitian Widiyanto et al. (2024) dan Daniel & Susanti (2021) yang membandingkan metode *TF-IDF* dan *cosine similarity*. Di sisi lain, penelitian Fitrianto & Editya (2024) mengevaluasi performa beberapa metode deep learning seperti IndoBERT, RoBERTa, BERT, dan BERT *Multilingual* dalam klasifikasi sarkas berbahasa Indonesia di platform X. Meskipun berbagai metode telah diterapkan, kombinasi BERT dan *cosine similarity* menunjukkan keunggulan dalam pemahaman konteks dan akurasi kesamaan semantik, dibandingkan dengan pendekatan konvensional yang sering mengandalkan pencocokan kata kunci secara *literal*. Dengan kemampuan BERT untuk menangkap makna, pendekatan ini berpotensi meningkatkan efektivitas rekomendasi dosen, mengurangi ketidakcocokan antara topik tugas akhir mahasiswa dan keahlian dosen. Penelitian ini berfokus pada pengembangan pemahaman semantik yang lebih baik untuk mengatasi kelemahan metode sebelumnya dan menyediakan rekomendasi yang lebih relevan dan tepat guna bagi mahasiswa.

Tujuan dari penelitian ini adalah untuk menghasilkan rekomendasi dosen pembimbing yang lebih tepat dan efisien berdasarkan kesesuaian antara topik tugas akhir mahasiswa dan keahlian dosen, sekaligus menutupi kekurangan dari penelitian sebelumnya. Implementasi BERT dan *cosine similarity* diharapkan mampu meningkatkan akurasi rekomendasi dengan mengurangi ketidakcocokan antara mahasiswa dan dosen. Hal ini akan berkontribusi pada peningkatan kualitas bimbingan akademik dan efisiensi proses penyelesaian tugas akhir. Dengan rekomendasi yang lebih relevan, mahasiswa dapat menyelesaikan studi tepat waktu, sehingga mendukung peningkatan kualitas pendidikan tinggi dalam menghadapi persaingan di era modern.

METODE

Pada gambar 1 merupakan skema tahapan penelitian. Pada tahap pengumpulan data, penelitian ini memanfaatkan judul-judul penelitian dosen Program Studi Teknik Informatika Universitas Dian Nuswantoro. Terdapat 3.723 judul dari 63 dosen berhasil dikumpulkan melalui *Web Scraping* dari platform *Google Scholar* dan *ResearchGate*. Kedua platform ini dipilih sebagai sumber data karena merupakan repositori publik yang banyak digunakan untuk menyimpan dan mengakses karya ilmiah.



Gambar 1. Tahapan penelitian

Pada tahap *preprocessing*, data yang telah dikumpulkan melalui *web scraping* diolah menggunakan teknik-teknik pemrosesan teks, seperti *case folding* (mengubah semua teks menjadi huruf kecil), penghapusan tanda baca, pembersihan teks dengan *regex*, tokenisasi (memecah teks menjadi kata-kata individu), dan penghapusan *stopword* (kata-kata umum yang tidak relevan seperti "dan", "atau"). Tahapan ini penting untuk menghindari noise dalam data dan memastikan bahwa hanya informasi penting yang digunakan untuk analisis lebih lanjut. *Stemming* dan *lemmatization* tidak digunakan dalam *preprocessing* ini karena dapat mengubah bentuk kata dan mempengaruhi makna teks. Hal ini berpotensi menurunkan akurasi pengukuran kemiripan semantik antara judul tugas akhir mahasiswa dan penelitian dosen. Oleh

karena itu, pendekatan ini dipilih untuk menjaga konteks kata agar hasilnya dapat diproses lebih lanjut oleh model pembelajaran mesin untuk menghasilkan embedding teks. Hasil preprocessing kemudian siap diproses oleh model pembelajaran mesin untuk menghasilkan embedding teks.

Pada tahap *embedding* BERT digunakan untuk mengubah teks judul tugas akhir mahasiswa dan judul penelitian dosen menjadi *vektor embedding*. Model yang digunakan adalah *pretrained* BERTBASE dengan *max_length* 512, memungkinkan pemrosesan teks hingga 512 token. Pemilihan BERTBASE didasarkan pada kemampuannya untuk memahami konteks secara *bidireksional*, yang memungkinkan representasi semantik lebih akurat dari teks. *Embedding* yang dihasilkan akan menangkap makna dari setiap kata dalam konteks keseluruhan kalimat, memungkinkan model untuk menangkap hubungan semantik yang lebih dalam antara judul tugas akhir dan penelitian dosen. Model ini dipilih karena terbukti memberikan performa terbaik dalam tugas NLP, terutama dalam konteks bahasa kompleks dan beragam. Dengan demikian, *embedding* yang dihasilkan memfasilitasi pengukuran kesamaan secara lebih akurat menggunakan *cosine similarity*, meningkatkan relevansi rekomendasi dosen dengan topik tugas akhir mahasiswa.

Setelah vektor *embedding* dari judul-judul penelitian mahasiswa dan dosen dihasilkan, tahap selanjutnya adalah menghitung *cosine similarity* antara vektor-vektor tersebut. *Cosine Similarity* mengukur kemiripan antara dua vektor dengan menghitung nilai kosinus dari sudut antara vektor-vektor tersebut. Semakin kecil sudutnya, semakin besar kemiripannya, yang berarti topik penelitian mahasiswa dan dosen semakin selaras. *cosine similarity* digunakan karena sederhana namun efektif dalam mengukur kemiripan antar teks berdasarkan representasi vektor mereka.

Pada tahap pengujian, kami melakukan uji coba untuk mengidentifikasi kesalahan atau error yang mungkin terlewatkan. Setelah proses pengembangan selesai, pengujian dilakukan untuk menilai performa dan akurasi rekomendasi. Beberapa judul tugas akhir digunakan sebagai masukan untuk mengevaluasi kemampuan model dalam merekomendasikan dosen yang sesuai. Pengujian ini memungkinkan penilaian sejauh mana model mampu menghasilkan rekomendasi yang akurat dan relevan dengan topik tugas akhir mahasiswa. Akurasi metode *cosine similarity* menggunakan pengujian akurasi dengan persamaan 1 untuk mengevaluasi seberapa tepat model dalam merekomendasikan dosen yang sesuai. Akurasi dihitung dengan membandingkan jumlah rekomendasi benar terhadap total data uji, memberikan gambaran tentang performa dan konsistensi model.

$$Akurasi = \frac{Jumlah\ Klasifikasi\ Benar}{Jumlah\ Data\ Testing} \quad (1)$$

HASIL DAN PEMBAHASAN

Hasil

Pengumpulan data dilakukan melalui *web scraping* judul penelitian dosen dari platform *Google Scholar* dan *ResearchGate* pada bulan Juni 2024. Dataset yang berhasil dikumpulkan mencakup 3.723 judul penelitian dari 63 dosen Program Studi Teknik Informatika, Universitas Dian Nuswantoro. Dataset yang dikumpulkan memuat judul penelitian dosen beserta informasi tambahan seperti nama dosen terkait. Pada tabel 1, dapat dilihat hasil *web scraping* yang berisi nama dosen serta judul penelitian dosen. Data ini diolah lebih lanjut melalui proses *preprocessing* untuk keperluan pemodelan dan analisis.

Tabel 1. Dataset hasil *web scrapping*

Nama Dosen	Judul Penelitian
Abdus Salam	Push Notification Using Firebase Cloud Messaging (FCM) on Employee Attendance Application, Web-Based Public Street Lighting Complaint Application with Realtime Whatsapp...
Abu Salam	Analisis Sentimen Data Komentar Sosial Media Facebook Dengan k-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia), Implementasi Algoritma...
Achmad Wahid	Pengenalan motif batik menggunakan deteksi tepi canny dan k-nearest neighbor, Perbandingan penggunaan deteksi tepi dengan metode laplace, sobel dan prewit dan canny pada...
Adhitya Nugraha	A systematic review of intelligence video surveillance: Trends, techniques, frameworks, and datasets, Dinus Intelligent Assistance (DINA) chatbot for university admission services...
Adi Prihandono	Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes, Klasifikasi Tingkat Kerentanan Malaria Pada Suatu Wilayah...

Pada Tabel 2 merupakan tahap *preprocessing*. Setelah *preprocessing*, judul penelitian dosen diubah menjadi representasi vektor *embedding* menggunakan *BERT*. Model *BERT* dipilih karena kemampuannya dalam menangkap konteks dari teks secara mendalam, memungkinkan model untuk memahami makna semantik di dalam teks. Representasi ini memungkinkan analisis kesamaan topik secara *presisi*, sehingga rekomendasi didasarkan pada *relevansi* semantik, bukan hanya kemiripan kata secara *literal*.

Tabel 2. *Preprocessing*

Tahap	Hasil
Awal	Perbandingan Efektivitas Naive Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube tahun 2022
Case Folding	perbandingan efektivitas naive - bayes dan svm dalam menganalisis sentimen kebencanaan di youtube tahun 2022
Punctuation	perbandingan efektivitas naive bayes dan svm dalam menganalisis sentimen kebencanaan di youtube tahun 2022
Regex	perbandingan efektivitas naive bayes dan svm dalam menganalisis sentimen kebencanaan di youtube tahun
Word_tokenize	['perbandingan', 'efektivitas', 'naive', 'bayes', 'dan', 'svm', 'dalam', 'menganalisis', 'sentimen', 'kebencanaan', 'di', 'youtube', 'tahun']
Stopword	['perbandingan', 'efektivitas', 'naive', 'bayes', 'svm', 'menganalisis', 'sentimen', 'kebencanaan', 'youtube']

Gambar 2 menampilkan hasil *word cloud* yang menggambarkan kata-kata dominan dalam judul penelitian dosen setelah *preprocessing*. Kata-kata yang sering muncul, seperti "*knearest neighbor*", "*naive bayes*", atau "*neural netwprk*", mencerminkan topik-topik yang sering diteliti. *Word cloud* ini membantu dalam identifikasi visual tema-tema populer, memudahkan mahasiswa untuk menemukan dosen yang relevan dengan minat mereka. Gambar 3 menampilkan *visualisasi PCA 2D* dari *embedding BERT*, yang digunakan untuk memahami pola dan pengelompokan judul penelitian berdasarkan kesamaan semantik. Titik-titik yang berdekatan menunjukkan judul-judul dengan tema atau konteks yang mirip, sedangkan titik yang berjauhan menandakan perbedaan signifikan dalam topik. Visualisasi ini

Tabel 3. Hasil pengujian data

No	Judul	Rekomendasi	Cosine Similarity	Aktual	Ket
1	A systematic review of intelligence video surveillance: Trends, techniques, frameworks, and datasets	Adhitya Nugraha	0.578	Adhitya Nugraha	valid
		Ifan Rizqa	0.563		
		Nova Rijati	0.559		
2	Teknik Super Enkripsi Menggunakan Transposisi Kolom Berbasis Vigenere Cipher Pada Citra Digital	Vincent	0.888	Daurat Sinaga	valid
		Suhartono	0.886		
		Daurat Sinaga			
3	A Comparative Study of Javanese Script Classification with GoogleNet, DenseNet, ResNet, VGG16 and VGG19	Eko Hari	0.844	Eko Hari Rachmawanto	valid
		Rachmawanto	0.843		
		Fahri Firdausillah	0.840		
4	A Survey and Taxonomy of Distributed Data Mining Research Studies: A Systematic Literature Review	Rama Aria M.	0.619	Fauzi Adi Rafrastara	valid
		Muljono	0.615		
		Adhitya Nugraha	0.603		
5	Document preprocessing with TF-IDF to improve the polarity classification performance of unstructured sentiment analysis	Fauzi Adi R.	0.769	Rama Aria Megantara	valid
		Rama Aria M.	0.768		
		Muljono	0.759		
6	Integration of Fuzzy Multi-Attribute Decision Making and Clustering Methods for Student Apprenticeship Recommendations	De Rosal Ignatius	0.421	Aris Marjuni	invalid
		Moses Setiadi	0.4		
		Ardiawan Bagus Harisa	0.399		
7	Hybrid method using 3-DES, DWT and LSB for secure image steganography algorithm	Feri Agustina	0.866	De Rosal Ignatius Moses Setiadi	valid
		De Rosal Ignatius	0.865		
		Moses Setiadi	0.865		
8	Performance Analysis of Support Vector Classification and Random Forest in Phishing Email Classification	Eko Hari	0.782	L. Budi Handoko	valid
		Rachmawanto	0.778		
		Fahri Firdausillah	0.777		
9	Algoritma Naive Bayes Dengan Fitur Seleksi Untuk Mengetahui Hubungan Variabel Nilai Dan Latar Belakang Pendidikan	De Rosal Ignatius	0.833	Yani Parti Astuti	valid
		Moses Setiadi	0.831		
		L. Budi Handoko	0.830		
10	Analisis Sentimen Data Komentar Sosial Media Facebook Dengan k-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)	Suprayogi	0.936	Abu Salam	valid
		Abu Salam	0.932		
		Vincent Suhartono	0.929		

Dataset penelitian ini mencakup 3.723 judul penelitian dari 63 dosen, namun beberapa topik mungkin kurang terwakili yang mempengaruhi kualitas rekomendasi. Evaluasi menunjukkan model ini efektif dalam mencocokkan judul mahasiswa dengan penelitian dosen, meskipun judul yang luas atau multidisipliner cenderung menurunkan *relevansi* (*cosine* rendah). Sebaliknya, judul dengan fokus teknis menghasilkan rekomendasi lebih akurat (*cosine* tinggi). Analisis kesalahan mengungkap keterbatasan dalam memahami konteks spesifik, sehingga dibutuhkan penambahan faktor seperti bidang minat dosen dan preferensi mahasiswa. Hasil penelitian ini menunjukkan bahwa rekomendasi dosen berbasis BERT dan *cosine similarity* dapat mempercepat pemilihan dosen pembimbing, mengurangi potensi ketidakcocokan, serta mengurangi kecemasan mahasiswa dalam proses pemilihan.

Pembahasan

Penelitian ini terdiri nama dosen serta judul penelitian dosen Program Studi Teknik Informatika Universitas Dian Nuswantoro. Terdapat beberapa keterbatasan dalam dataset yang dikumpulkan, seperti dosen yang tidak aktif, sehingga ada potensi bias dalam data yang diambil. Pada tahap *preprocessing* dilakukan untuk membersihkan data sehingga lebih mudah diolah oleh model. Penambahan langkah-langkah seperti *stemming* atau *lemmatization* juga dapat memengaruhi hasil akhir dari analisis. Proses ini diikuti oleh penggunaan model *BERT* untuk mengubah judul penelitian menjadi vektor *embedding*.

Pemilihan BERT dalam penelitian ini karena kemampuannya yang kuat dalam menangkap makna semantik dan konteks dalam teks. *BERT*, yang berbasis pada arsitektur *transformer*, memiliki keunggulan dalam memahami hubungan antar kata dalam kalimat, dibandingkan model lain seperti *Word2Vec* atau *GloVe*. Hal ini sangat penting dalam penelitian ini karena judul penelitian sering kali memiliki konteks spesifik yang perlu dipahami secara mendalam. Setelah vektor *embedding*, metode *cosine similarity* digunakan untuk mengukur kesamaan antara kedua vektor tersebut. *cosine similarity* dinilai efektif dalam mengukur kesamaan semantik.

Pengujian dilakukan dengan metode *random testing* menggunakan 10 judul tugas akhir. Hasil pengujian menunjukkan bahwa pendekatan ini memberikan rekomendasi dosen pembimbing dengan akurasi 90%. Pengujian untuk melihat sejauh mana merekomendasikan dosen yang relevan (Nurwanda et al., 2024) serta diperoleh pemahaman mengenai tingkat ketepatan prediksi yang dihasilkan oleh model (Pamungkas & Cahyono, 2024). Penelitian ini melakukan pengujian dengan input judul tugas akhir mahasiswa untuk mengukur akurasi dan kesesuaian antara topik penelitian dengan keahlian dosen yang dihasilkan oleh model.

Penelitian Azhari & Buulolo (2022) menggunakan *TextRank* dan *cosine similarity*. Sementara itu Hairani & Mujahid (2022) yang menunjukkan bahwa algoritma berbasis NLP mampu meningkatkan kualitas rekomendasi dalam konteks akademik. Pendekatan ini sebelumnya terbatas pada pencocokan kata kunci literal, kurang mampu menangkap konteks semantik mendalam. Penelitian ini memperkenalkan inovasi dengan menggabungkan BERT, yang memahami konteks serta *cosine similarity* untuk mengukur kesamaan semantik secara efisien. Keunggulan BERT terlihat dalam penelitian (Braja & Kodar, 2023), yang menunjukkan akurasi 94% dalam tugas analisis sentimen, membuktikan efektivitas model ini dalam memahami konteks bahasa yang kompleks. Di sisi lain, hasil penelitian yang dilakukan oleh Widiyanto, et al. (2024) mengidentifikasi kelemahan pendekatan TF-IDF dalam menangkap makna yang lebih dalam. Oleh karena itu, dengan implementasi BERT dengan *cosine similarity* dapat meningkatkan pemahaman semantik dan memungkinkan pemilihan dosen yang lebih tepat sesuai topik tugas akhir mahasiswa. Penelitian ini membuka peluang pengembangan dengan mempertimbangkan preferensi mahasiswa dan minat dosen, meningkatkan relevansi rekomendasi.

SIMPULAN

Penelitian ini menunjukkan bahwa kombinasi model BERT dan *cosine similarity* secara signifikan meningkatkan akurasi rekomendasi dosen pembimbing hingga 90%, menciptakan kecocokan yang lebih baik antara judul penelitian mahasiswa dan keahlian dosen. Pendekatan ini berpotensi diintegrasikan dengan sistem manajemen akademik yang ada, memungkinkan pemilihan dosen secara otomatis dan mengurangi subjektivitas dalam proses ini. Dengan meningkatkan kolaborasi antara mahasiswa dan dosen, serta menyediakan rekomendasi yang lebih responsif terhadap kebutuhan akademik, penelitian ini berkontribusi pada pengembangan lingkungan pembelajaran yang lebih inovatif, efisien, dan berkualitas.

REFERENSI

- Andriani, N., & Wibawanta, B. (2020). Peran Dosen Pembimbing Sebagai Pemimpin Yang Melayani Dalam Pembimbingan Tugas Akhir Mahasiswa Program Sarjana. *Polyglot: Jurnal Ilmiah*, 16(2), 230–251. <https://doi.org/10.19166/pji.v16i2.1927>
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITe)*, 3(1), 19–27. <https://doi.org/10.30812/bite.v3i1.1110>
- Aprilinda, Y., Martavia, T., Erlangga, E., Afandi, F. N., & Rizal, U. (2022). Chatbot Menggunakan Natural Language Processing untuk Pembelajaran Bahasa Inggris Berbasis Android. *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, 12(1), 23–27. <https://doi.org/10.36448/expert.v12i1.2629>
- Aqila, N., & Bijaksana, M. A. (2020). Developing Set of Word Senses of Vocabulary in Al-Qur'an. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 83–90. <https://doi.org/10.29408/edumatic.v4i1.2119>
- Asfi, M., & Fitrianiingsih, N. (2020). Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi. *Jurnal Nasional Informatika Dan Teknologi Jaringan*, 5(1), 45–50.
- Azhari, A., & Buulolo, E. (2022). Sistem Rekomendasi Dosen Pendamping Skripsi Berbasis Text Rank menggunakan Metode Cosine Similarity. *Pelita Informatika*, 10(3), 119–122.
- Braja, A. S. P., & Kodar, A. (2023). Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store. *J I M P - Jurnal Informatika Merdeka Pasuruan*, 7(3), 120–128. <https://doi.org/10.51213/jimp.v7i3.779>
- Budiono, A., & Eniyati, S. (2023). Sistem Rekomendasi Dosen Pembimbing Tugas Akhir Menggunakan Content Based Filtering. *Jurnal Elektronika Dan Komputer*, 16(1), 64–71.
- Damanik, B. E. (2022). Pengaruh Minat Baca Dan Peran Dosen Pembimbing Terhadap Keberhasilan Penulisan Tugas Akhir. *Jurnal Manajemen Pendidikan (JMP)*, 11(1), 28–39. <https://doi.org/10.26877/jmp.v11i1.2994>
- Daniel, & Susanti, W. (2021). Penerapan Algoritma Cosine Similarity Pada Sistem Pengajuan. *Prosiding SENATIKA 2021*, 1–10.
- Fitrianto, R. A., & Editya, A. S. (2024). Klasifikasi Tweet Sarkasme Pada Platform X Menggunakan Bidirectional Encoder Representations from Transformers. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 6(3), 366–371. <https://doi.org/10.47233/jteksis.v6i3.1344>
- Hairani, H., & Mujahid, M. (2022). Recommendations of Thesis Supervisor using the Cosine Similarity Method. *Sistemasi*, 11(3), 646–654. <https://doi.org/10.32520/stmsi.v11i3.2003>
- Hanif, J., Farid, A. N., Hasanah, M., & Barokatun, B. (2023). Penerapan Natural Language Processing untuk Klasifikasi Bidang Minat berdasarkan Judul Tugas Akhir. *Jurnal Sistim*

- Informasi Dan Teknologi*, 5(1), 41–49. <https://doi.org/10.37034/jsisfotek.v5i1.196>
- Khoirunisa, R. (2020). Penggunaan Natural Language Processing Pada Chatbot Untuk Media Informasi Pertanian. *Indonesian Journal of Applied Informatics*, 4(2), 55–63. <https://doi.org/10.20961/ijai.v4i2.38688>
- Kurniawan, B., Ari Aldino, A., & Rahman Isnain, A. (2022). Sentimen Analisis terhadap Kebijakan Penyelenggara Sistem Elektronik (PSE) Menggunakan Algoritma Bidirectional Encoder Representations from Transformers (Bert). *Jurnal Teknologi Dan Sistem Informasi*, 3(4), 98–106.
- Mustaqim, K., Amaresti, F. A., & Dewi, I. N. (2024). Analisis Sentimen Ulasan Aplikasi PosPay untuk Meningkatkan Kepuasan Pengguna dengan Metode K-Nearest Neighbor (KNN). *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 11–20. <https://doi.org/10.29408/edumatic.v8i1.24779>
- Nurwanda, N., Suarna, N., & Prihartono, W. (2024). Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1841–1846. <https://doi.org/10.36040/jati.v8i2.8469>
- Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 1–10. <https://doi.org/10.29408/edumatic.v8i1.24114>
- Samosir, F. V. P., Toba, H., & Ayub, M. (2022). BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(1), 202–217. <https://doi.org/10.28932/jutisi.v8i1.4474>
- Sapanji, R. A. E. V. T., Hamdani, D., & Harahap, P. (2023). Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP). *Journal of Applied Informatics and Computing*, 7(2), 202–211. <https://doi.org/10.30871/jaic.v7i2.6517>
- Sugiharno, R. T., Ari Susanto, W. H., & Wospakrik, F. (2022). Faktor-Faktor yang Mempengaruhi Kecemasan Mahasiswa dalam Menghadapi Tugas Akhir. *Jurnal Keperawatan Silampari*, 5(2), 1189–1197. <https://doi.org/10.31539/jks.v5i2.3760>
- Widianto, A., Pebriyanto, E., Fitriyanti, F., & Marna, M. (2024). Document Similarity using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity. *Journal of Data Science, Information Technology, and Data Analytics (Dinda)*, 4(2), 149–153. <https://doi.org/10.20895/dinda.v4i2.1589>