



Sistem Rekomendasi Judul Skripsi Berdasarkan Kemiripan Topik Menggunakan Sentence-BERT dan Cosine Similarity

Dicki Danuarta^{1*}, Muhammad Haris²

^{1,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Email: ^{1*}dikidanuarta99@gmail.com, ²muhammadharis@umsu.ac.id

Abstrak

Pemilihan judul skripsi merupakan tahapan penting dalam penyusunan tugas akhir mahasiswa, namun mahasiswa sering mengalami kesulitan menentukan judul yang sesuai minat penelitian sekaligus memastikan judul tersebut tidak memiliki kemiripan topik yang tinggi dengan penelitian terdahulu. Proses pengecekan kemiripan judul secara manual, yang umumnya hanya mengandalkan pencocokan kata kunci (keyword matching), membutuhkan waktu lama dan belum mampu memahami kemiripan makna antarjudul yang disusun dengan susunan kata berbeda. Penelitian ini bertujuan mengembangkan sistem rekomendasi judul skripsi berbasis website yang menerapkan Sentence-BERT (Sentence Bidirectional Encoder Representations from Transformers) untuk mengubah judul skripsi menjadi representasi vektor semantik (sentence embedding), serta Cosine Similarity untuk menghitung tingkat kemiripan antara judul yang dimasukkan pengguna dengan 2.764 data judul skripsi yang diperoleh melalui web scraping dari repository Universitas Sumatera Utara (USU), Universitas Negeri Medan (UNIMED), dan Universitas Islam Negeri Sumatera Utara (UINSU). Sistem dikembangkan menggunakan framework Flask (Python) sebagai backend dan basis data MySQL, dengan antarmuka berbasis HTML, CSS, dan Bootstrap. Hasil pengujian fungsional terhadap lima variasi kueri judul menunjukkan bahwa sistem berhasil menampilkan sepuluh rekomendasi judul teratas (Top 10) untuk setiap kueri, dengan nilai kemiripan tertinggi mencapai 76,00% pada kueri uji "Sistem Pakar" dan waktu pemrosesan pencarian rata-rata sekitar 4,357 detik terhadap keseluruhan dataset. Sistem yang dikembangkan terbukti mampu memahami kemiripan makna antarjudul secara semantik, bukan hanya kesamaan kata, sehingga berpotensi membantu mahasiswa memperoleh referensi judul yang lebih relevan dan mengurangi kemungkinan duplikasi topik penelitian.

Kata Kunci: K Sistem Rekomendasi, Judul Skripsi, Sentence-BERT, Cosine Similarity, Website.

PENDAHULUAN

Perkembangan teknologi informasi memberikan dampak signifikan terhadap berbagai aspek kehidupan, termasuk dalam bidang pendidikan tinggi. Dalam proses akademik, pengajuan judul skripsi merupakan tahapan awal yang sangat penting karena menjadi dasar pelaksanaan penelitian mahasiswa. Namun, mahasiswa sering mengalami kesulitan menentukan judul skripsi yang sesuai dengan bidang keilmuannya sekaligus memastikan bahwa judul yang diajukan tidak memiliki tingkat kemiripan yang tinggi dengan penelitian yang telah ada, sehingga usulan judul kerap mengalami revisi atau penolakan yang memperpanjang proses penyelesaian skripsi. Proses pengecekan kemiripan judul pada umumnya masih dilakukan secara manual dengan membandingkan usulan judul terhadap data skripsi yang tersimpan pada repository perguruan tinggi, suatu cara yang membutuhkan waktu lama dan berpotensi menghasilkan kesalahan identifikasi seiring bertambahnya jumlah data skripsi setiap tahun. Selain itu, metode pencarian berbasis kata kunci (keyword matching) belum mampu memahami makna suatu kalimat secara menyeluruh, sehingga judul dengan topik serupa namun susunan kata berbeda sering kali tidak terdeteksi sebagai penelitian yang memiliki kemiripan.

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah Natural Language Processing (NLP) melalui model Sentence-BERT (SBERT), yaitu pengembangan dari model BERT yang dirancang untuk menghasilkan representasi vektor (sentence embedding) dari suatu kalimat melalui arsitektur siamese network, sehingga tingkat kemiripan antarjudul dapat dihitung berdasarkan makna kalimat secara efisien, bukan hanya berdasarkan kesamaan kata (Reimers & Gurevych, 2019). Lubis dkk. (2026) menunjukkan bahwa penerapan Sentence-BERT pada sistem pencarian semantik repository akademik mampu meningkatkan relevansi hasil pencarian karena sistem dapat memahami

hubungan makna antarkalimat secara lebih baik dibandingkan metode pencarian konvensional. Penelitian lain yang dilakukan Fathuddin dkk. (2025) mengenai penerapan Sentence-BERT dan Cosine Similarity pada pencarian semantik dokumen skripsi dalam format PDF menunjukkan bahwa kombinasi kedua metode tersebut mampu menghasilkan pencarian dokumen yang lebih relevan meskipun menggunakan susunan kata yang berbeda, sehingga pendekatan berbasis representasi semantik terbukti memberikan hasil yang lebih baik dibandingkan metode pencarian yang hanya mengandalkan kesamaan kata.

Penerapan Sentence-BERT yang dipadukan dengan Cosine Similarity juga telah dimanfaatkan pada berbagai domain lain, misalnya sistem rekomendasi buku berbasis konten (Hartatik & Syafrianto, 2025) dan chatbot FAQ akademik (Holis dkk., 2025), yang keduanya membuktikan bahwa representasi semantik berbasis SBERT menghasilkan tingkat relevansi yang lebih tinggi dibandingkan pencocokan kata biasa. Meskipun demikian, penelitian mengenai sistem rekomendasi maupun pendeteksian kemiripan judul skripsi umumnya masih berfokus pada proses identifikasi tingkat kemiripan antarjudul. Suhardi Rahman dkk. (2025) menunjukkan bahwa pendekatan semantic similarity mampu mendeteksi tingkat kemiripan judul skripsi dengan baik, namun sistem yang dikembangkan belum menyediakan fitur rekomendasi alternatif bagi mahasiswa ketika tingkat kemiripan yang diperoleh tergolong tinggi.

Berdasarkan kondisi tersebut, penelitian ini mengembangkan Sistem Rekomendasi Judul Skripsi Berdasarkan Kemiripan Topik Menggunakan Sentence-BERT dan Cosine Similarity Berbasis Website. Sistem yang dikembangkan memanfaatkan 2.764 data skripsi dari repository Universitas Sumatera Utara (USU), Universitas Negeri Medan (UNIMED), dan Universitas Islam Negeri Sumatera Utara (UINSU) sebagai data pembandingan untuk menghitung tingkat kemiripan judul menggunakan Sentence-BERT dan Cosine Similarity. Selain menghitung tingkat kemiripan, sistem juga memberikan rekomendasi alternatif judul berdasarkan topik yang relevan, sehingga berbeda dari penelitian deteksi kemiripan yang hanya berhenti pada tahap identifikasi. Dengan adanya sistem ini, diharapkan mahasiswa memperoleh referensi judul yang lebih sesuai, mengurangi kemungkinan terjadinya duplikasi topik penelitian, serta membantu program studi dalam meningkatkan efektivitas proses seleksi usulan judul skripsi.

METODE

Penelitian ini menggunakan pendekatan Research and Development (R&D) untuk mengembangkan sistem rekomendasi judul skripsi berbasis website yang menerapkan algoritma Sentence-BERT (SBERT) dan metode Cosine Similarity. Tahapan penelitian meliputi pengumpulan dataset, perancangan sistem, implementasi algoritma, serta pengujian fungsional terhadap sistem yang dikembangkan.

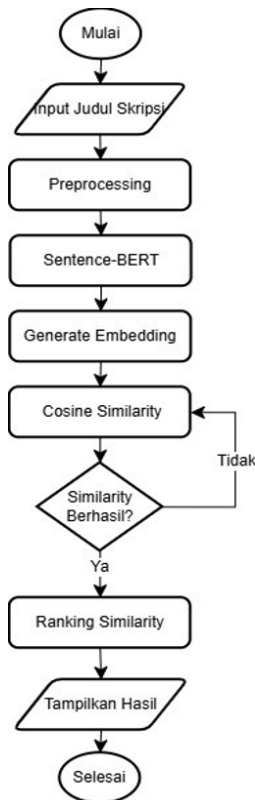
Pengumpulan Dataset

Dataset penelitian dikumpulkan melalui proses web scraping terhadap repository tiga perguruan tinggi, yaitu Universitas Sumatera Utara (USU), Universitas Negeri Medan (UNIMED), dan Universitas Islam Negeri Sumatera Utara (UINSU), sehingga diperoleh 2.764 data skripsi yang kemudian disimpan dalam format CSV dan diimpor ke dalam basis data MySQL. Dataset tersebut memuat lima atribut, yaitu judul skripsi, nama penulis (author), abstrak, nama kampus asal, dan tautan (link) menuju halaman repository asli. Keragaman sumber data dari tiga perguruan tinggi yang berbeda memberikan variasi topik penelitian yang cukup luas sebagai basis pembandingan dalam proses pencarian kemiripan judul.

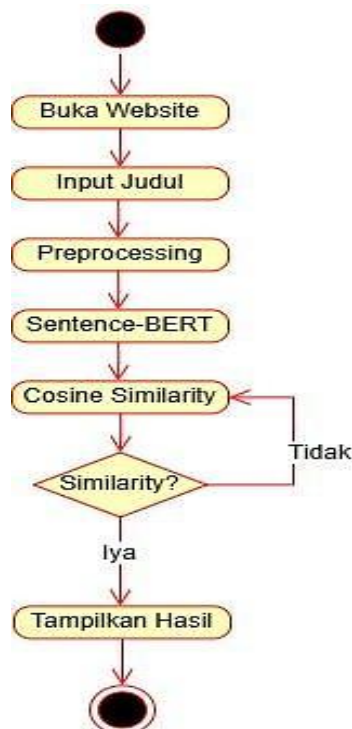
Arsitektur dan Alur Sistem

Sistem dirancang menggunakan Unified Modeling Language (UML) berupa Use Case Diagram dan Activity Diagram untuk menggambarkan interaksi antara pengguna dan sistem. Secara umum, alur kerja sistem terdiri atas tujuh tahapan sebagai berikut:

1. Pengguna memasukkan usulan judul skripsi melalui halaman beranda website tanpa memerlukan proses login terlebih dahulu.
2. Sistem melakukan tahap praproses (preprocessing) terhadap teks judul yang dimasukkan, meliputi case folding (penyeragaman huruf menjadi huruf kecil) dan penghapusan spasi berlebih.
3. Judul hasil praproses diproses menggunakan model Sentence-BERT sehingga menghasilkan representasi vektor (sentence embedding) berdimensi tetap yang merepresentasikan makna semantik judul tersebut.
4. Embedding judul pengguna dibandingkan dengan embedding seluruh judul pada basis data menggunakan metode Cosine Similarity untuk memperoleh nilai kemiripan pada rentang 0 hingga 1.
5. Nilai kemiripan yang diperoleh diurutkan dari nilai tertinggi hingga terendah.
6. Sistem memilih dan menampilkan sepuluh judul dengan nilai kemiripan tertinggi (Top 10) sebagai rekomendasi.
7. Hasil rekomendasi ditampilkan kepada pengguna beserta informasi pendukung berupa judul skripsi, nama penulis, asal universitas, nilai similarity dalam bentuk persentase, dan progress bar tingkat kemiripan.



Gambar 1. Flowchart Sistem



Gambar 2. Activity Diagram Sistem

Model Sentence-BERT

Sentence-BERT merupakan pengembangan dari model BERT yang dirancang khusus untuk menghasilkan representasi vektor (sentence embedding) dari suatu kalimat secara efisien melalui arsitektur siamese network, sehingga proses pencarian maupun perbandingan dokumen dapat dilakukan dengan lebih cepat dibandingkan BERT konvensional tanpa mengorbankan akurasi representasi semantik (Reimers & Gurevych, 2019). Dalam penelitian ini, setiap judul skripsi pada dataset ditransformasikan menjadi vektor embedding menggunakan model Sentence-BERT, sehingga kemiripan antarjudul dapat diukur berdasarkan makna kalimat dan bukan hanya berdasarkan kesamaan kata.

Perhitungan Cosine Similarity

Tingkat kemiripan antara embedding judul yang dimasukkan pengguna dan embedding judul pada basis data dihitung menggunakan metode Cosine Similarity, yang mengukur nilai kosinus sudut yang terbentuk antara dua vektor sesuai persamaan berikut:

$$\text{Similarity}(A, B) = (A \cdot B) / (\|A\| \times \|B\|)$$

dengan A adalah vektor embedding judul yang dimasukkan pengguna, B adalah vektor embedding judul pada basis data, $A \cdot B$ adalah hasil dot product kedua vektor, serta $\|A\|$ dan $\|B\|$ masing-masing adalah panjang (norma) vektor A dan vektor B. Nilai similarity yang dihasilkan berada pada rentang 0 hingga 1, dengan nilai yang semakin mendekati 1 menunjukkan tingkat kemiripan makna yang semakin tinggi antara kedua judul yang dibandingkan.

Implementasi Perangkat Lunak

Sistem rekomendasi judul skripsi diimplementasikan sebagai aplikasi berbasis website menggunakan framework Flask (Python) sebagai backend, yang berfungsi menghubungkan antarmuka pengguna dengan proses pengolahan data menggunakan Sentence-BERT, melakukan komunikasi dengan basis data MySQL, serta menampilkan hasil rekomendasi kepada pengguna (Ma'arif & Kurniasih, 2024). Antarmuka pengguna (front-end) dibangun menggunakan HTML, CSS, dan Bootstrap sehingga tampilan sistem responsif dan dapat diakses melalui komputer maupun perangkat seluler. Rancangan antarmuka sistem terdiri atas halaman Beranda sebagai media input judul, halaman Hasil Rekomendasi untuk menampilkan Top 10 judul dengan nilai kemiripan tertinggi, serta halaman Admin untuk pengelolaan data judul skripsi.

Teknik Pengujian Sistem

Pengujian sistem dilakukan secara fungsional (black-box testing) untuk mengetahui apakah sistem mampu menghasilkan rekomendasi judul skripsi sesuai dengan masukan pengguna. Pengujian dilakukan terhadap lima variasi kueri judul yang mewakili topik penelitian berbeda, yaitu "Sistem Pakar", "Sistem Informasi Akademik", "Data Mining", "Machine Learning", dan "Website E-Commerce". Setiap kueri diuji untuk memastikan sistem berhasil menampilkan sepuluh rekomendasi judul (Top 10) beserta nilai kemiripan dan waktu pemrosesan pencarian (search time).

HASIL DAN PEMBAHASAN

Implementasi Sistem

Sistem rekomendasi judul skripsi berhasil diimplementasikan sebagai aplikasi berbasis website menggunakan framework Flask. Sistem memanfaatkan model Sentence-BERT untuk menghasilkan representasi semantik dari judul skripsi, kemudian menghitung tingkat kemiripan menggunakan metode Cosine Similarity terhadap dataset sebanyak 2.764 judul skripsi yang berasal dari repository USU, UNIMED, dan UINSU.

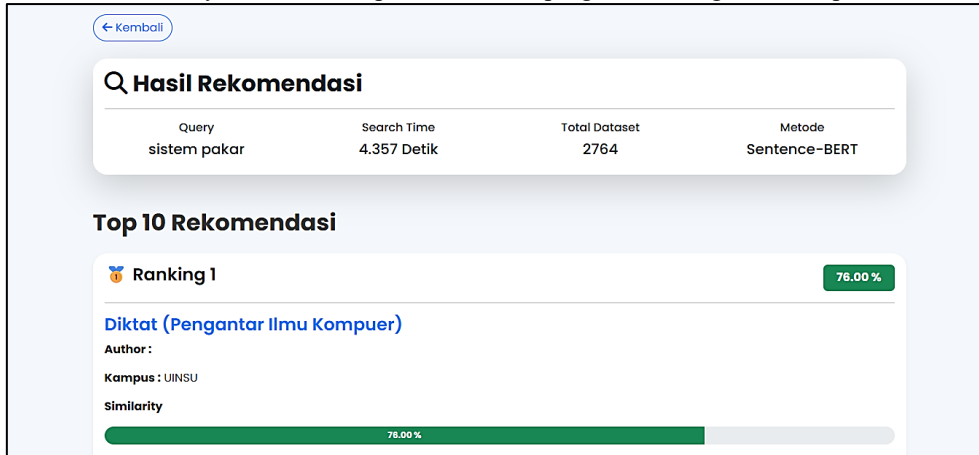
Implementasi Antarmuka

Halaman Beranda merupakan halaman utama yang pertama kali ditampilkan ketika pengguna membuka website. Pada halaman ini tersedia kolom input untuk memasukkan judul skripsi yang akan dianalisis, serta informasi mengenai jumlah dataset yang digunakan (2.764 data skripsi), jumlah universitas sumber data (3 universitas), dan metode yang digunakan (Sentence-BERT dan Cosine Similarity).



Gambar 1. Halaman Beranda Sistem

Halaman Hasil Rekomendasi ditampilkan setelah sistem selesai melakukan proses pencarian kemiripan. Pada halaman ini ditampilkan kueri yang dimasukkan pengguna, lama proses pencarian (search time), jumlah dataset yang digunakan, metode yang digunakan, serta daftar sepuluh judul skripsi (Top 10) dengan nilai similarity tertinggi. Setiap hasil rekomendasi menampilkan judul skripsi, nama penulis, asal universitas, nilai similarity dalam bentuk persentase, dan progress bar tingkat kemiripan.



Gambar 2. Halaman Hasil Rekomendasi

Pengujian Sistem

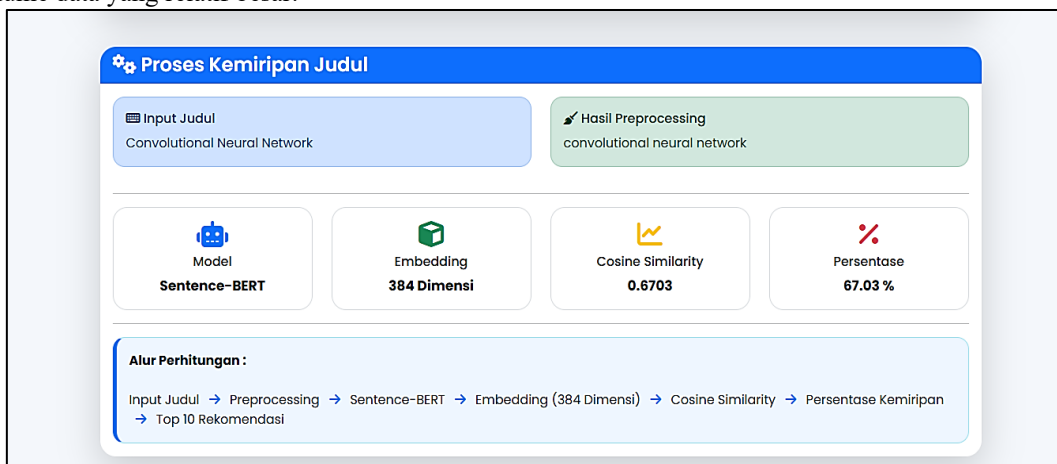
Pengujian dilakukan untuk mengetahui apakah sistem mampu menghasilkan rekomendasi judul skripsi sesuai dengan masukan pengguna. Pengujian dilakukan terhadap lima variasi kueri judul yang mewakili topik penelitian berbeda, sebagaimana dirangkum pada Tabel 1.

Tabel 1. Hasil Pengujian Sistem

No	Input Judul	Hasil
1	Sistem Pakar	Berhasil menampilkan Top 10 rekomendasi
2	Sistem Informasi Akademik	Berhasil menampilkan Top 10 rekomendasi
3	Data Mining	Berhasil menampilkan Top 10 rekomendasi
4	Machine Learning	Berhasil menampilkan Top 10 rekomendasi
5	Website E-Commerce	Berhasil menampilkan Top 10 rekomendasi

Sumber: Hasil pengujian sistem (2026)

Seluruh skenario pengujian menunjukkan bahwa sistem berhasil memproses kueri judul yang dimasukkan dan menampilkan sepuluh rekomendasi judul teratas secara konsisten tanpa kegagalan fungsi. Sebagai ilustrasi, pengujian menggunakan kueri "Sistem Pakar" menghasilkan sepuluh judul skripsi dengan tingkat kemiripan tertinggi, dengan salah satu hasil rekomendasi memperoleh nilai similarity sebesar 76,00%, yang menunjukkan keterkaitan topik yang cukup tinggi terhadap kueri yang dimasukkan pengguna. Selain menghasilkan rekomendasi, sistem juga mampu menyelesaikan proses pencarian dalam waktu sekitar 4,357 detik pada dataset sebanyak 2.764 data skripsi, menunjukkan bahwa kombinasi Sentence-BERT dan Cosine Similarity dapat diterapkan secara responsif meskipun dijalankan terhadap volume data yang relatif besar.



Gambar 3. Proses Alur Kerja Sistem

Pembahasan

Hasil pengujian menunjukkan bahwa sistem mampu memberikan rekomendasi judul skripsi berdasarkan kemiripan topik menggunakan metode Sentence-BERT dan Cosine Similarity. Sentence-BERT digunakan untuk menghasilkan representasi semantik dari judul skripsi sehingga sistem tidak hanya membandingkan kesamaan kata, tetapi juga memahami makna dari kalimat yang dimasukkan pengguna. Hal ini sejalan dengan temuan Reimers dan Gurevych (2019) yang memperkenalkan arsitektur Sentence-BERT berbasis siamese network sebagai metode yang efisien untuk menghasilkan sentence embedding dengan performa yang lebih baik dibandingkan BERT konvensional dalam tugas pengukuran kemiripan semantik.

Pendekatan yang diterapkan pada penelitian ini sejalan dengan temuan Fathuddin dkk. (2025) yang menunjukkan bahwa kombinasi Sentence-BERT dan Cosine Similarity mampu menghasilkan pencarian dokumen skripsi yang lebih relevan meskipun menggunakan susunan kata yang berbeda, serta Lubis dkk. (2026) yang membuktikan bahwa penerapan Sentence-BERT pada sistem pencarian semantik repository akademik mampu meningkatkan relevansi hasil pencarian dibandingkan metode pencarian berbasis kata kunci konvensional. Dibandingkan dengan penelitian Suhardi Rahman dkk. (2025) yang berfokus pada deteksi tingkat kemiripan judul skripsi tanpa menyediakan fitur rekomendasi alternatif, sistem yang dikembangkan pada penelitian ini memiliki nilai tambah berupa fitur Top 10 rekomendasi judul alternatif yang dapat langsung dipertimbangkan oleh mahasiswa, sehingga tidak hanya mendeteksi kemiripan tetapi juga membantu proses eksplorasi ide penelitian baru.

Waktu pemrosesan pencarian yang relatif singkat (sekitar 4,357 detik) terhadap 2.764 data skripsi menunjukkan bahwa pendekatan sentence embedding berbasis Sentence-BERT dapat diterapkan secara praktis pada skala data repository perguruan tinggi tanpa mengorbankan responsivitas sistem. Hasil ini juga memperkuat temuan Hartatik dan Syafrianto (2025) yang melaporkan bahwa penerapan Sentence-BERT pada sistem rekomendasi berbasis konten mampu meningkatkan ketepatan hasil pencarian karena sistem dapat menganalisis karakteristik data secara lebih mendalam dibandingkan pendekatan pencarian konvensional.

Meskipun demikian, pengujian yang dilakukan pada penelitian ini masih bersifat fungsional (black-box testing) dan belum mengukur akurasi sistem secara kuantitatif menggunakan metrik evaluasi seperti precision, recall, atau Mean Average Precision (MAP) terhadap ground truth kemiripan judul yang divalidasi secara manual oleh pakar. Keterbatasan ini perlu menjadi perhatian pada penelitian lanjutan, mengingat evaluasi kuantitatif yang lebih ketat dapat memberikan gambaran yang lebih objektif mengenai tingkat akurasi rekomendasi yang dihasilkan sistem, sebagaimana juga dilakukan pada penerapan NLP untuk deteksi kemiripan dokumen ilmiah lain yang mengombinasikan algoritma pencocokan string dengan metrik evaluasi akurasi kuantitatif (Rofiq & Khair, 2026). Secara keseluruhan, hasil penelitian ini menegaskan bahwa kombinasi Sentence-BERT dan Cosine Similarity memiliki kemampuan yang baik dalam memahami kemiripan makna antarjudul skripsi, sehingga sistem yang dikembangkan berpotensi membantu mahasiswa memperoleh referensi judul yang lebih relevan, mengurangi kemungkinan duplikasi topik penelitian, serta membantu program studi dalam melakukan evaluasi awal terhadap usulan judul skripsi secara lebih cepat dan efisien.

KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi judul skripsi berbasis website menggunakan framework Flask yang mampu membantu pengguna memperoleh rekomendasi judul skripsi berdasarkan kemiripan topik. Sistem dibangun dengan memanfaatkan dataset sebanyak 2.764 judul skripsi yang berasal dari repository Universitas Sumatera Utara (USU), Universitas Negeri Medan (UNIMED), dan Universitas Islam Negeri Sumatera Utara (UINSU). Metode Sentence-BERT berhasil diterapkan untuk mengubah judul skripsi menjadi representasi vektor (sentence embedding), sedangkan metode Cosine Similarity berhasil digunakan untuk menghitung tingkat kemiripan antara judul yang dimasukkan pengguna dengan seluruh judul pada dataset. Hasil perhitungan tersebut digunakan sebagai dasar dalam menentukan rekomendasi judul yang memiliki tingkat kemiripan tertinggi. Pengujian fungsional terhadap lima variasi kueri judul menunjukkan bahwa sistem konsisten berhasil menampilkan sepuluh rekomendasi judul skripsi (Top 10) beserta informasi pendukung berupa persentase kemiripan, nama penulis, dan asal universitas, dengan nilai similarity tertinggi mencapai 76,00% dan waktu pemrosesan pencarian sekitar 4,357 detik terhadap keseluruhan dataset. Dengan demikian, sistem dapat membantu mahasiswa memperoleh referensi judul skripsi yang lebih relevan serta mengurangi kemungkinan pengajuan judul yang memiliki tingkat kemiripan tinggi dengan penelitian yang telah ada. Penelitian selanjutnya disarankan menambah cakupan dataset dari lebih banyak repository perguruan tinggi, mengembangkan sistem agar turut menganalisis abstrak atau isi dokumen secara lebih mendalam, menambahkan fitur penyaringan berdasarkan program studi atau bidang penelitian, serta melakukan evaluasi kuantitatif menggunakan metrik seperti precision, recall, atau Mean Average Precision (MAP) terhadap ground truth yang divalidasi secara manual untuk mengukur akurasi rekomendasi secara lebih objektif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, atas dukungan fasilitas dan bimbingan akademik selama penelitian ini dilaksanakan.

DAFTAR PUSTAKA

- Al-Hassan, M., Abu-Salih, B., Alshdaifat, E., Aloqaily, A., & Rodan, A. (2024). An Improved Fusion-Based Semantic Similarity Measure for Effective Collaborative Filtering Recommendations. *International Journal of Computational Intelligence Systems*, 17(1), Article 45. <https://doi.org/10.1007/s44196-024-00429-4>
- Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., Sung, Y. H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder for English. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 169–174. <https://doi.org/10.18653/v1/D18-2029>
- Choi, H., Kim, J., Joe, S., & Gwon, Y. (2021). Evaluation of BERT and ALBERT Sentence Embedding Performance on Downstream NLP Tasks. *2020 25th International Conference on Pattern Recognition (ICPR)*, 5482–5487. <https://doi.org/10.1109/ICPR48806.2021.9412102>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*, 1, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Fathuddin, M. A. H., Mandyartha, E. P., & Nurlaili, A. L. (2025). Penerapan Sentence-BERT dan Cosine Similarity untuk Pencarian Semantik Dokumen Skripsi dalam Format PDF. *Ranah Research: Journal of Multidisciplinary Research and Development*, 8(1), 322–337. <https://doi.org/10.38035/rj.v8i1.1865>
- Febianto Agiharta, K., Renaldy Suteja, B., & Ayub, M. (2024). Application of Sentence BERT for Job Competency and Course Similarity. *JUTISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 10(3), 449–467.
- Gao, T., Yao, X., & Chen, D. (2021). SimCSE: Simple Contrastive Learning of Sentence Embeddings. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6894–6910. <https://doi.org/10.18653/v1/2021.emnlp-main.552>
- Gusdevi, H., Setyanto, A., Kusri, K., & Utami, E. (2025). Cosine Similarity-Based Evidence Selection for Fact Verification Using SBERT on the FEVER Dataset. *Cogito Smart Journal*, 11(1), 52–66. <https://doi.org/10.31154/cogito.v11i1.917.52-66>
- Hartatik, H., & Syafianto, A. (2025). Penerapan Model Sentence-BERT untuk Sistem Rekomendasi Buku Berbasis Konten di Perpustakaan Digital. *Jurnal Dialektika Informatika (Detika)*, 6(1), 12–19. <https://doi.org/10.24176/detika.v6i1.15916>
- Henderson, M., Al-Rfou, R., Strope, B., Sung, Y. H., Lukács, L., Guo, R., Kumar, S., Miklós, B., & Kurzweil, R. (2017). Efficient Natural Language Response Suggestion for Smart Reply. arXiv. <http://arxiv.org/abs/1705.00652>
- Holis, R. M., Utomo, P. E. P., & Hutabarat, B. F. (2025). Semantic FAQ Chatbot Using SBERT (Sentence-BERT) and Cosine Similarity for Academic Services. *Brilliance: Research of Artificial Intelligence*, 5(2), 915–922. <https://doi.org/10.47709/brilliance.v5i2.7027>
- Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. T. (2020). Dense Passage Retrieval for Open-Domain Question Answering. *Proceedings of EMNLP 2020*, 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
- Khattab, O., & Zaharia, M. (2020). ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 39–48. <https://doi.org/10.1145/3397271.3401075>
- Lubis, I., Lubis, H., & Nur Wahidah, I. (2026). Implementation of Semantic Search in an Academic Repository Using Sentence-BERT and FAISS. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 10(2), 1060–1069. <https://doi.org/10.33395/sinkron.v10i2.15940>

- Ma'arif, O. M., & Kurniasih, T. (2024). Perancangan Sistem Inventory Berbasis Web Menggunakan Framework Flask: PT. Gagas Mitra Jaya (Area Salatiga). *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(2), 1947–1959. <https://doi.org/10.35870/jimik.v5i2.822>
- Mahadi, M. R. S., & Utama, N. P. (2023). Indonesian Text-to-Image Synthesis with Sentence-BERT and FastGAN. arXiv. <http://arxiv.org/abs/2303.14517>
- Malik, A., Srinivasan, S., & Prakash, P. (2024). Enhancing Recommendations of Items by Making Some Changes in Layers of BERT Model. *International Journal of Computer Applications*, 186(3), 28–36. <https://doi.org/10.5120/ijca2024923366>
- Noorian, A., Harounabadi, A., & Hazratifard, M. (2024). A Sequential Neural Recommendation System Exploiting BERT and LSTM on Social Media Posts. *Complex & Intelligent Systems*, 10(1), 721–744. <https://doi.org/10.1007/s40747-023-01191-4>
- Pangestu, F. A., Athallah, F., Sofwan, A., & Christyono, Y. (2024). Pengembangan Aplikasi Sistem Informasi Reservasi Ruangan Berbasis Website. *Transient: Jurnal Ilmiah Teknik Elektro*, 13(3), 115–124.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. *Proceedings of EMNLP-IJCNLP 2019*, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>
- Rofiq, M. A., & Khair, R. (2026). Aplikasi Pendeteksi Plagiat Karya Tulis Ilmiah Berbasis Web Menggunakan NLP dan Algoritma KMP. *Indonesian Journal of Multidisciplinary Scientific Studies (IJOMSS)*, 4(5), 318–327. <https://doi.org/10.33151/ijomss.v4i5.1114>
- Sabilillah, F. T., Winarno, S., & Abiyyi, R. B. (2024). Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing Berdasarkan Judul Tugas Akhir. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 585–594. <https://doi.org/10.29408/edumatic.v8i2.27791>
- Suhardi Rahman, E., Iswal Burhan, M., & T Mangesa, R. (2025). Development of Sentence Similarity Detection Application with Semantic Similarity and Machine Learning Approaches (Case Study: Student Thesis Title). *Jurnal Media Elektrik*, 23(1), 19–36. <https://doi.org/10.59562/metrik.v23i1.10503>
- Syah, M. P., Wardani, A. P., Idhom, M., & Trimono. (2025). Perbandingan Representasi Teks TF-IDF dan BERT terhadap Akurasi Cosine Similarity dalam Penilaian Otomatis Jawaban Berbasis Teks. *Data Sciences Indonesia (DSI)*, 5(1), 47–59. <https://doi.org/10.47709/dsi.v5i1.6021>