

Sistem Rekomendasi Referensi Berbasis *mBERT* dan *Analytical Hierarchy Process* Terintegrasi *Semantic Scholar*

Muhamad Raasikh Fil'ilmi¹, Paramitha Nerisafitra²

^{1,2}Teknik Informatika, Universitas Negeri Surabaya

¹muhamadfililmi.22076@mhs.unesa.ac.id

²paramithanerisafitra@unesa.ac.id

Abstrak — Perkembangan teknologi informasi dan meningkatnya jumlah mahasiswa menyebabkan kompleksitas dalam penentuan referensi tugas akhir yang relevan, berkualitas, dan memiliki unsur kebaruan, sehingga proses pencarian referensi seringkali memakan waktu lama dan bersifat subjektif. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi referensi tugas akhir berbasis website yang mampu memberikan rekomendasi jurnal internasional secara otomatis berdasarkan kriteria kualitas *novelty*, *clarity*, *specificity*, dan *relevance*, serta menampilkan referensi yang sesuai dengan tren penelitian global melalui integrasi *Semantic Scholar* API. Metode yang digunakan meliputi analisis semantik berbasis *Multilingual BERT (mBERT)* untuk menghasilkan *embedding* teks dan menghitung kemiripan menggunakan *cosine similarity*, serta metode *Analytical Hierarchy Process (AHP)* untuk pembobotan multi-kriteria dalam proses pemeringkatan referensi. Dataset yang digunakan berasal dari metadata publikasi ilmiah dari *Semantic Scholar*. Hasil evaluasi menunjukkan nilai *accuracy* sebesar 82%, *precision* sebesar 85,2%, *recall* sebesar 92%, dan *F1-score* sebesar 0,885 yang mengindikasikan performa model yang baik dalam klasifikasi kemiripan semantik. Berdasarkan hasil implementasi, waktu rata-rata preprocessing dan pembentukan *embedding* berkisar antara 0,2–0,5 detik per data, sedangkan waktu rekomendasi berada pada kisaran 40–100 detik. Pengujian terhadap 50 responden menggunakan metode *System Usability Scale (SUS)* menghasilkan skor rata-rata 70,1 yang termasuk dalam kategori baik. Dengan demikian, sistem ini efektif dalam membantu mahasiswa dan dosen dalam memperoleh referensi penelitian yang relevan, terstruktur, dan sesuai dengan perkembangan riset terkini.

Kata Kunci — Sistem Rekomendasi, *BERT*, *AHP*, *Semantic Scholar*, Analisis Semantik.

I. PENDAHULUAN

Tugas akhir menjadi tantangan serius bagi mahasiswa. Perkembangan teknologi informasi pada era digital telah membawa perubahan besar dalam pengelolaan data akademik di perguruan tinggi, khususnya pada kegiatan penelitian mahasiswa yang semakin kompleks [1]. Dalam lingkungan akademik, tugas akhir merupakan salah satu syarat wajib kelulusan yang berfungsi sebagai bentuk penelitian komprehensif mahasiswa pada akhir masa studi [2]. Pemilihan referensi penelitian yang relevan, mutakhir, dan berkualitas menjadi faktor penting dalam menentukan kualitas karya ilmiah yang dihasilkan [3]. Oleh karena itu, proses pencarian referensi tugas akhir menjadi aspek yang sangat penting untuk memastikan penelitian memiliki kontribusi ilmiah yang jelas dan sesuai dengan perkembangan ilmu pengetahuan [4].

Universitas Negeri Surabaya (UNESA) merupakan salah satu perguruan tinggi yang telah menerapkan teknologi informasi secara luas untuk mendukung kegiatan akademik [5].

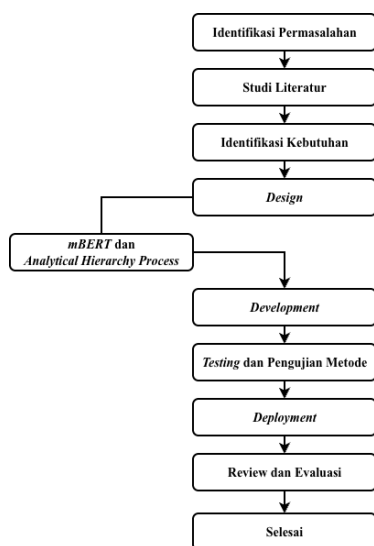
Setelah memperoleh status Perguruan Tinggi Negeri Berbadan Hukum (PTNBH), jumlah mahasiswa aktif UNESA mengalami peningkatan signifikan hingga mencapai 73.347 mahasiswa pada tahun 2025 [6]. Peningkatan jumlah mahasiswa tersebut berdampak pada bertambahnya jumlah penelitian dan judul tugas akhir yang diajukan setiap tahunnya [7]. Kondisi tersebut menyebabkan proses penentuan referensi penelitian menjadi semakin kompleks baik bagi mahasiswa maupun dosen pembimbing.

Meskipun berbagai sistem akademik telah diterapkan, banyak mahasiswa masih mengalami kesulitan dalam menentukan referensi penelitian yang memenuhi unsur *novelty*, *clarity*, *specificity*, dan *relevance* [8]. Proses pencarian referensi masih banyak dilakukan secara manual dan bergantung pada subjektivitas pengguna sehingga membutuhkan waktu yang lama serta kurang terukur secara kuantitatif. Permasalahan tersebut menyebabkan mahasiswa sering mengalami hambatan dalam menentukan arah penelitian dan memperoleh referensi yang sesuai dengan perkembangan riset terkini.

Beberapa penelitian sebelumnya telah membahas pemanfaatan model *BERT* untuk analisis semantik serta metode *Analytical Hierarchy Process (AHP)* untuk pengambilan keputusan multikriteria. Penelitian oleh Pradani dan Suaada menunjukkan bahwa model *BERT* mampu menghasilkan analisis *semantic similarity* yang lebih akurat dibandingkan metode tradisional seperti *TF-IDF* [9]. Selain itu, penelitian Katarina et al. menunjukkan bahwa metode *AHP* mampu menghasilkan proses penilaian yang lebih objektif melalui pembobotan kriteria secara terstruktur [10]. Penelitian lain juga menunjukkan bahwa *novelty* merupakan faktor penting dalam menentukan kualitas dan kontribusi ilmiah suatu penelitian [11]. Namun, belum banyak penelitian yang menggabungkan variabel *novelty*, *clarity*, *specificity*, dan *relevance* dalam satu sistem rekomendasi referensi penelitian berbasis analisis semantik secara otomatis.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pengembangan sistem rekomendasi referensi tugas akhir berbasis *website* menggunakan analisis semantik *Multilingual BERT (mBERT)* dan metode *Analytical Hierarchy Process (AHP)* dengan integrasi *Semantic Scholar* API. Sistem dirancang untuk memberikan rekomendasi referensi jurnal internasional berdasarkan penilaian kualitas *novelty*, *clarity*, *specificity*, dan *relevance* secara otomatis. Dengan adanya sistem ini, diharapkan mahasiswa dan dosen pembimbing dapat memperoleh referensi penelitian yang lebih relevan, objektif, dan sesuai dengan perkembangan tren riset global.

II. METODE PENELITIAN



Gbr. 1 Agile Method

Metode penelitian yang digunakan adalah metode *agile*. Sesuai Gbr. 1, metode *agile* terdiri dari beberapa tahapan, antara lain:

A. Identifikasi Kebutuhan

Tahap ini meliputi proses identifikasi kebutuhan dalam pengembangan sistem rekomendasi referensi tugas akhir berbasis analisis semantik *Multilingual BERT (mBERT)* dan metode *Analytical Hierarchy Process (AHP)*. Identifikasi kebutuhan dilakukan untuk memahami permasalahan utama yang dihadapi mahasiswa dan dosen pembimbing dalam proses pencarian referensi penelitian yang relevan, mutakhir, dan memiliki kualitas ilmiah yang baik.

Proses identifikasi dilakukan melalui studi literatur dari berbagai jurnal ilmiah terkait sistem rekomendasi referensi, *semantic similarity*, *Natural Language Processing (NLP)*, serta metode pengambilan keputusan multikriteria menggunakan *AHP*. Selain itu, identifikasi kebutuhan juga dilakukan dengan menganalisis proses pencarian referensi tugas akhir yang selama ini masih dilakukan secara manual dan bergantung pada subjektivitas pengguna.

Berdasarkan hasil identifikasi, sistem dirancang untuk mampu melakukan proses pencarian, evaluasi, dan pemeringkatan referensi jurnal internasional secara otomatis berdasarkan empat variabel kualitas penelitian, yaitu *novelty*, *clarity*, *specificity*, dan *relevance*. Sistem juga diintegrasikan dengan *Semantic Scholar API* untuk memperoleh metadata publikasi ilmiah sesuai dengan perkembangan tren penelitian global.

2.1.1 Pengguna Sistem

Sistem rekomendasi ini dirancang untuk melayani tiga peran pengguna utama:

1. Mahasiswa
Mahasiswa menjadi pengguna utama dalam sistem rekomendasi referensi penelitian. Sistem membantu mahasiswa memperoleh rekomendasi jurnal internasional yang relevan dengan topik penelitian yang dimasukkan pengguna. Proses rekomendasi dilakukan menggunakan analisis semantik berbasis *Multilingual BERT (mBERT)* dan pembobotan multikriteria menggunakan metode *Analytical Hierarchy Process (AHP)*. Melalui sistem ini, mahasiswa dapat memperoleh referensi penelitian yang lebih terstruktur, relevan, dan sesuai dengan perkembangan riset terkini.
2. Dosen Pembimbing
Dosen pembimbing berperan sebagai pihak yang membantu mengarahkan topik penelitian mahasiswa agar tetap sesuai dengan bidang keilmuan dan perkembangan riset terbaru. Sistem ini membantu dosen pembimbing dalam mengevaluasi kualitas referensi penelitian mahasiswa berdasarkan variabel *novelty*, *clarity*, *specificity*, dan *relevance*.
3. Tim PPTI
Tim PPTI berperan sebagai admin sistem yang bertanggung jawab dalam pengelolaan dan pemeliharaan sistem rekomendasi referensi penelitian. Admin melakukan proses pengambilan data publikasi ilmiah dari *Semantic Scholar API*, *preprocessing* data, pembentukan *embedding* menggunakan *mBERT*, serta pengelolaan dataset pada basis data sistem.

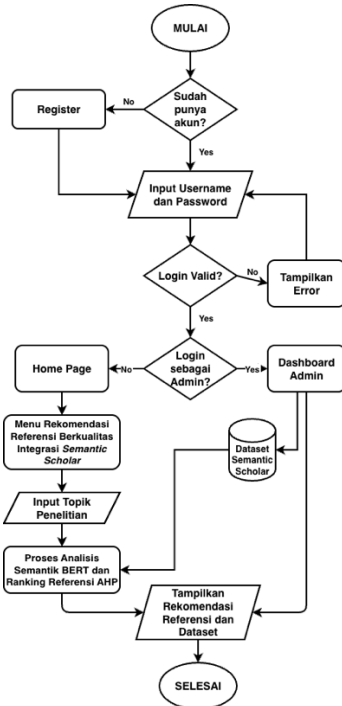
2.1.2 Metode Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan menggunakan dua metode utama, yaitu studi literatur dan pengumpulan data publikasi ilmiah melalui *Application Programming Interface (API)*. Studi literatur dilakukan dengan mengkaji berbagai jurnal ilmiah, artikel penelitian, dan publikasi akademik yang membahas *Natural Language Processing (NLP)*, *semantic similarity*, *Multilingual BERT (mBERT)*, *Analytical Hierarchy Process (AHP)*, serta sistem rekomendasi referensi penelitian. Literatur diperoleh melalui platform seperti *Google Scholar*, *ScienceDirect*, *IEEE Xplore*, dan *Semantic Scholar* dengan fokus pada publikasi beberapa tahun terakhir untuk memastikan relevansi penelitian. Selain itu, pengumpulan data juga dilakukan melalui integrasi *Semantic Scholar API* untuk memperoleh metadata publikasi ilmiah internasional yang meliputi judul penelitian, abstrak, tahun publikasi, *citation count*, dan informasi pendukung lainnya. Setelah proses pengambilan dan filtering data, diperoleh sebanyak 32.644 metadata publikasi ilmiah yang digunakan sebagai dataset utama dalam proses analisis semantik dan rekomendasi referensi penelitian. Selain dataset publikasi ilmiah, penelitian ini juga menggunakan *technical seed list* sebanyak 43.114 istilah teknis yang digunakan untuk mendukung perhitungan variabel *specificity*. Data yang diperoleh selanjutnya digunakan sebagai dasar dalam proses *preprocessing*, pembentukan *embedding* semantik, perhitungan

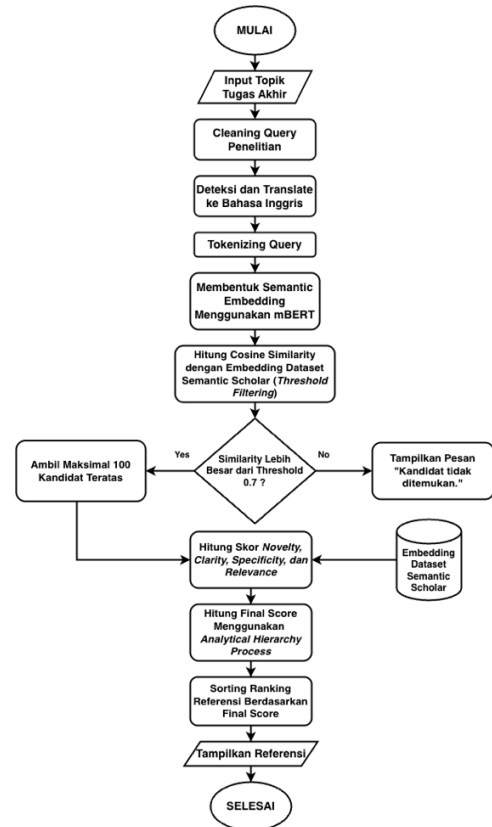
variabel kualitas referensi, serta proses rekomendasi penelitian pada sistem yang dikembangkan.

B. Design

2.2.1 Flowchart



Gbr. 2 General Flowchart

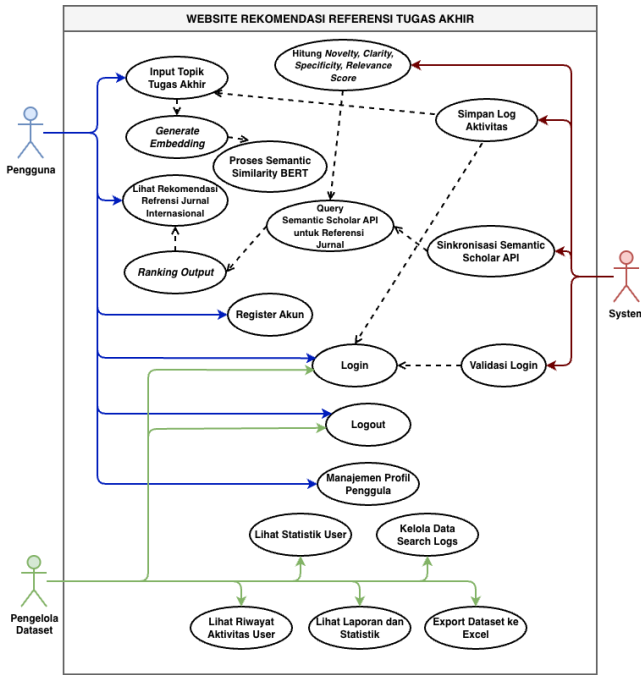


Gbr. 3 System Flowchart

Flowchart merupakan representasi grafis yang menggambarkan urutan dan alur proses dalam sistem dengan menggunakan simbol-simbol standar. Diagram ini memperlihatkan tahapan-tahapan yang dilakukan oleh pengguna (mahasiswa & dosen pembimbing) dan Admin dalam berinteraksi dengan sistem. Alur secara umum dapat dilihat pada Gbr. 2. Sementara *flowchart* sistem rekomendasi referensi digunakan untuk menggambarkan tahapan cara kerja sistem dalam menentukan rekomendasi referensi berdasarkan skor akhir terbaik dapat dilihat pada Gbr. 3.

2.2.2 Usecase Diagram

Sistem rekomendasi referensi penelitian melibatkan tiga aktor utama, yaitu pengguna (mahasiswa dan dosen pembimbing), admin, dan sistem rekomendasi. Pengguna dapat memasukkan topik penelitian untuk memperoleh rekomendasi referensi ilmiah berdasarkan analisis *semantic similarity* menggunakan *Multilingual BERT (mBERT)*. Sistem kemudian melakukan proses *embedding*, perhitungan *similarity*, evaluasi variabel kualitas referensi (*novelty*, *clarity*, *specificity*, dan *relevance*), serta pemeringkatan menggunakan metode *Analytical Hierarchy Process (AHP)*. Admin bertugas mengelola dataset publikasi ilmiah, melakukan *preprocessing* data, memperbarui *technical seed list*, serta memonitor aktivitas sistem. Seluruh interaksi antar aktor dan sistem direpresentasikan melalui usecase diagram pada Gbr. 4.

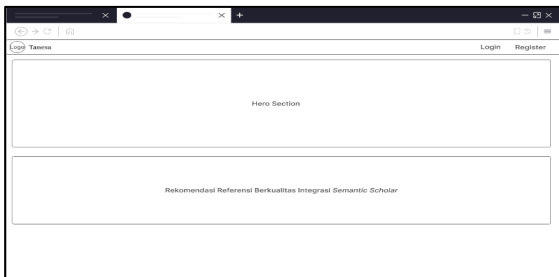


Gbr. 4 Usecase Diagram

2.2.3 Wireframe Lo-Fi

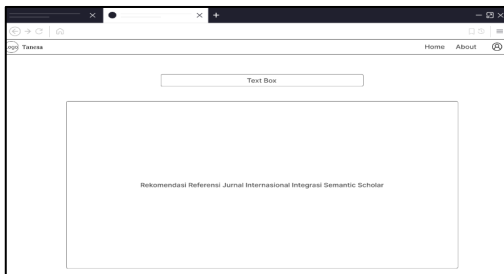
Wireframe low-fidelity adalah gambaran sederhana dari tampilan sistem yang menggambarkan fungsionalitas dasar tanpa elemen desain estetika. Berikut adalah komponen dasar dari Wireframe Lo-Fi:

a. Halaman Labuh



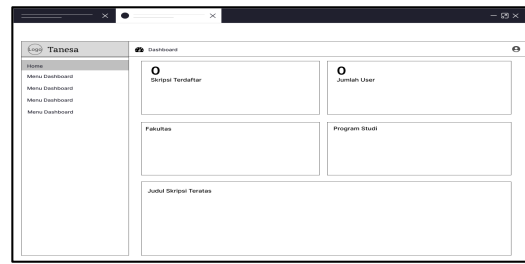
Gbr. 5 Halaman Labuh

b. Halaman Rekomendasi Referensi Tugas Akhir



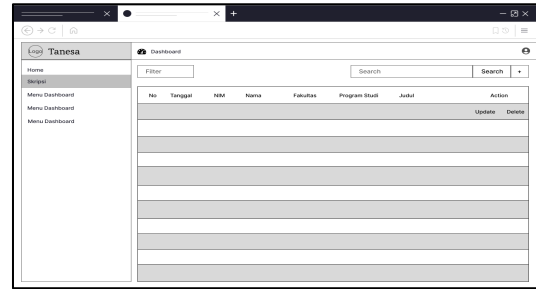
Gbr. 6 Halaman Rekomendasi Referensi Tugas Akhir

c. Dashboard Admin



Gbr. 7 Dashboard Admin

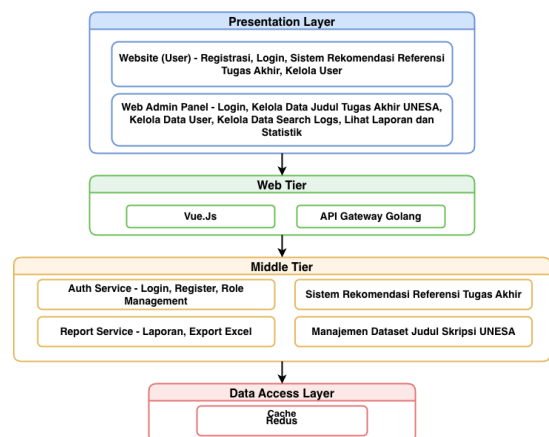
d. Halaman Manajemen Dataset



Gbr. 8 Halaman Manajemen Dataset

2.2.4 Arsitektur Sistem

Arsitektur sistem dalam penelitian ini dirancang menggunakan pendekatan *multi-tier architecture* seperti pada Gbr. 9 yang terdiri dari *presentation layer*, *web tier*, *middle tier*, *data access layer*, dan *data layer* untuk memastikan pemisahan fungsi sistem berjalan secara modular dan terstruktur. Pada *presentation layer*, *framework Vue.js* digunakan sebagai antarmuka pengguna (*user interface*) yang menyediakan layanan *input* judul, serta rekomendasi referensi ilmiah berkualitas urut dari skor akhir tertinggi sampai terendah yang diambil melalui integrasi *Semantic Scholar*. Antarmuka ini berkomunikasi dengan *backend* melalui protokol *RESTful API* yang mendukung format pertukaran data *JSON*, sehingga proses *request* dan *response* data dapat berlangsung secara asinkron serta efisien. Pada *web tier*, sistem dibangun menggunakan *framework Vue.js* pada *frontend* dan bahasa pemrograman *Golang* yang berfungsi sebagai pusat logika bisnis untuk menjalankan proses analisis semantik menggunakan *BERT embedding*, perhitungan *similarity* menggunakan *cosine similarity*, pengambilan data publikasi melalui *Semantic Scholar API*, ekstraksi kata teknis, dan perhitungan skor berdasarkan variabel *novelty*, *clarity*, *specificity*, dan *relevance*. *Layer* ini juga bertanggung jawab menjalankan mekanisme pembobotan menggunakan *Analytical Hierarchy Process (AHP)* untuk menghasilkan skor akhir kualitas referensi berbasis multi-kriteria yang objektif dan terukur. Sementara itu, *data layer* menggunakan *PostgreSQL* sebagai sistem manajemen basis data yang menyimpan *dataset Semantic Scholar*.



sesuai untuk menangani input penelitian berbahasa Indonesia dan referensi ilmiah berbahasa Inggris. Pada penelitian ini, *mBERT* digunakan tanpa proses *fine-tuning* karena belum tersedia dataset berlabel khusus pada domain rekomendasi referensi akademik.

Sementara itu, metode *Analytical Hierarchy Process (AHP)* digunakan pada tahap pembobotan karena mampu melakukan pengambilan keputusan multi-kriteria secara terstruktur melalui *pairwise comparison*. Metode ini dipilih karena variabel *novelty*, *clarity*, *relevance*, dan *specificity* memiliki tingkat kepentingan yang berbeda sehingga diperlukan mekanisme pembobotan yang objektif dan konsisten. Selain menghasilkan bobot prioritas, *AHP* juga menyediakan pengujian konsistensi melalui *Consistency Ratio (CR)* untuk memastikan validitas penilaian antar variabel. Kombinasi *mBERT* dan *AHP* diharapkan mampu menghasilkan sistem rekomendasi referensi penelitian yang lebih relevan, terukur, dan sesuai dengan kebutuhan akademik.

Gbr. 9 Architecture Diagram

2.2.5 Preprocessing dan Translasi Teks

Pada penelitian ini, sistem menerapkan tahap *preprocessing* untuk memastikan konsistensi representasi teks sebelum dilakukan analisis semantik menggunakan *Multilingual BERT (mBERT)*. Mengingat input pengguna dapat berasal dari Bahasa Indonesia maupun Bahasa Inggris, sistem menerapkan mekanisme *language detection* dan translasi otomatis menggunakan *LibreTranslate API*.

Tahap *preprocessing* dimulai dengan proses *cleaning* untuk menghapus karakter khusus, spasi berlebih, dan elemen non-alfabetis. Selanjutnya, sistem melakukan deteksi bahasa terhadap teks input. Jika input terdeteksi sebagai Bahasa Indonesia, sistem akan menerjemahkan teks ke Bahasa Inggris menggunakan layanan translasi otomatis. Sebaliknya, apabila input telah berbahasa Inggris, proses translasi dilewati dan teks langsung diproses ke tahap berikutnya.

Setelah proses normalisasi bahasa selesai, teks dilakukan tokenisasi dan dikonversi menjadi *embedding vector* menggunakan model *mBERT* untuk mendukung proses *semantic similarity* pada sistem rekomendasi referensi penelitian.

2.2.6 Justifikasi Pemilihan *Multilingual BERT* dan *Analytical Hierarchy Process*

Pemilihan metode pada penelitian ini didasarkan pada kebutuhan sistem untuk memahami makna semantik teks penelitian secara kontekstual serta melakukan pemeringkatan referensi secara objektif. Pada tahap analisis semantik, penelitian ini menggunakan *Multilingual BERT (mBERT)* karena mampu menghasilkan *contextual embedding* melalui mekanisme *bidirectional self-attention* sehingga hubungan antar kata dapat dipahami berdasarkan konteks kalimat. Dibandingkan metode representasi teks tradisional seperti *TF-IDF* dan *Word2Vec*, *mBERT* memiliki kemampuan lebih baik dalam menangkap hubungan *semantic similarity* pada teks akademik multibahasa.

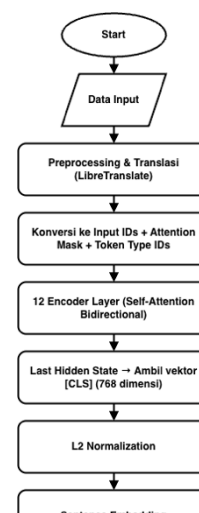
Selain itu, *mBERT* dipilih karena telah dilatih menggunakan korpus multibahasa dari 104 bahasa sehingga

2.2.7 Representasi Semantik Teks Menggunakan *Multilingual BERT*

Pada tahap ini, sistem membentuk representasi semantik teks menggunakan model *Multilingual BERT (mBERT)* dengan mengubah topik pengguna, *title*, dan *abstract* dari *Semantic Scholar* menjadi *embedding vector* untuk merepresentasikan makna teks secara kontekstual. Model *mBERT* yang digunakan terdiri atas 12 *encoder layer*, 12 *attention head*, dan 768 *hidden size*, di mana setiap lapisan menerapkan mekanisme *self-attention* secara *bidirectional* sehingga mampu memahami hubungan kontekstual antar kata dengan lebih baik dan mendukung perbandingan teks berbasis semantik secara lebih akurat. Dalam penelitian ini, *mBERT* diterapkan tanpa proses *fine-tuning (zero-shot/frozen weights)* karena tidak tersedia *dataset* berlabel yang spesifik untuk domain referensi akademik multibahasa. Pemilihan pendekatan ini didasarkan pada kemampuan *mBERT* yang telah dilatih pada korpus Wikipedia dari 104 bahasa sehingga memiliki representasi lintas bahasa yang memadai untuk menangani teks berbahasa Indonesia dan Inggris. Meskipun model seperti *paraphrase-multilingual-mpnet-base-v2* dari *Sentence-BERT* secara teoritis lebih optimal untuk tugas *sentence similarity* karena telah melalui proses *fine-tuning* khusus, *mBERT* dipilih karena cakupan bahasanya yang lebih luas dan lebih sesuai dengan kebutuhan sistem yang dikembangkan. Nilai kemiripan semantik antara query pengguna dan kandidat referensi dihitung menggunakan metode cosine similarity seperti pada Persamaan (1).

$$MaxSim = \max(\text{similarity}(ref_i, ref_j)) \quad (1)$$

Alur kerja *mBERT* pada penelitian ini dijelaskan pada Gbr. 10.



$$SemanticNovelty = \left(\frac{1 - MaxSim}{1 - Treshold} \right) \times 100 \quad (2)$$

Selain aspek kebaruan semantik, penelitian ini juga mempertimbangkan tingkat kebermaknaan ilmiah (*scientific importance*) menggunakan pendekatan *citation rate*.

$$CitationRate = \left(\frac{Citation}{CurrentYear - Year + 1} \right) \quad (3)$$

Skor akhir *novelty* pada penelitian ini diperoleh melalui penggabungan antara *semantic novelty* dan *scientific importance*.

$$NoveltyScore = SemanticNovelty \times ScientificImportance \quad (4)$$

Alur proses perhitungan *novelty* ditunjukkan pada Gbr. 11.

Gbr. 10 Alur Kerja *mBERT*

2.2.8 Perhitungan Skor Variabel Kualitas Referensi

Tahap ini mengubah hasil analisis semantik menjadi skor kuantitatif untuk empat variabel penilaian, yaitu *novelty*, *clarity*, *specificity*, dan *relevance*, menggunakan nilai *cosine similarity* yang diperoleh dari referensi hasil *retrieval* Semantic Scholar. Variabel *novelty* dihitung melalui kombinasi *semantic novelty* berdasarkan kemiripan maksimum antar referensi dan *scientific importance* yang diukur menggunakan *citation rate* yang telah dinormalisasi. Variabel *clarity* mengukur konsistensi makna internal berdasarkan variasi kedekatan semantik antar token pada judul, sedangkan *specificity* dihitung dari proporsi istilah teknis melalui pemeriksaan *embedding* terhadap *technical seed list*. Sementara itu, *relevance* diperoleh dari rata-rata nilai *similarity* antara referensi dan *semantic query* gabungan yang merepresentasikan kebutuhan pengguna dan domain akademik. Hasil perhitungan seluruh variabel kemudian dinormalisasi menjadi skor terstandarisasi yang selanjutnya digunakan pada proses pembobotan menggunakan *Analytical Hierarchy Process (AHP)* untuk menghasilkan penilaian kualitas referensi yang objektif dan konsisten.

1. Novelty Score

Novelty dihitung melalui pendekatan gabungan dua komponen, yaitu keunikan semantik antar kandidat referensi (*text novelty*) dan pengakuan komunitas ilmiah melalui jumlah sitasi (*citation novelty*). Pendekatan ini dikembangkan berdasarkan kritik terhadap metode berbasis kemiripan teks semata, di mana rendahnya kemiripan antar referensi tidak selalu mencerminkan kebaruan yang bermakna, sehingga tidak banyak dibahas oleh peneliti lain [15]. Oleh karena itu, komponen sitasi diintegrasikan sebagai faktor koreksi untuk membedakan referensi yang benar-benar baru dan bernilai dari referensi yang sekadar jarang dibahas karena lemah secara ilmiah. Formulasi perhitungan *semantic novelty* ditunjukkan pada Persamaan (2).



Gbr. 11 Flowchart Proses Penilaian *Novelty Score*

2. Clarity Score

Perhitungan *clarity* bertujuan untuk mengukur tingkat kejelasan dan konsistensi makna suatu judul referensi berdasarkan representasi semantik yang dihasilkan oleh model *BERT* [21]. Proses dimulai dengan *preprocessing* teks judul, kemudian dilakukan pembentukan *embedding vector* untuk memperoleh representasi makna secara kontekstual. Selanjutnya, nilai *semantic similarity* antar token dalam judul dihitung dan digunakan untuk memperoleh nilai *standard deviation* sebagai indikator konsistensi linguistik. Semakin kecil nilai *standard deviation*, semakin konsisten hubungan semantik antar token sehingga judul dianggap lebih jelas dan mudah dipahami. Pendekatan ini memungkinkan evaluasi *clarity*

dilakukan secara objektif berdasarkan hubungan makna antar kata, tidak hanya bergantung pada struktur gramatikal atau panjang judul. Penentuan nilai *standard deviation* dihitung menggunakan Persamaan (5).

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (5)$$

Nilai *standard deviation* yang dihasilkan kemudian dikonversi menjadi skor *clarity* melalui Persamaan (6).

$$Clarity = (1 - s) \times 100 \quad (6)$$

Alur proses perhitungan *clarity* ditunjukkan pada Gbr. 12.



Gbr. 12 Flowchart Proses Penilaian Clarity Score

3. Specificity Score

Perhitungan *specificity* bertujuan untuk mengukur tingkat kekhususan suatu referensi dalam merepresentasikan topik penelitian tertentu. Secara eksplisit, formula perhitungan skor *specificity* ditunjukkan pada Persamaan (7).

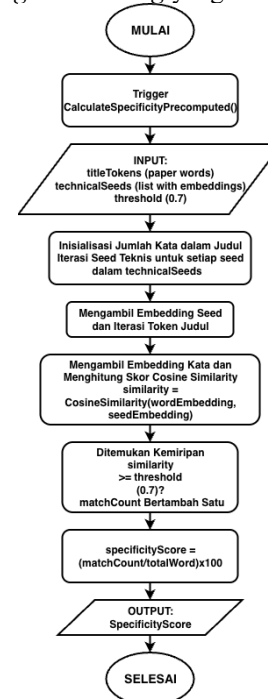
$$Specificity = \frac{\text{Jumlah Kata Teknis}}{\text{Total Kata}} \times 100 \quad (7)$$

Pada Persamaan (7), identifikasi kata teknis dilakukan menggunakan model *BERT* melalui perhitungan kesesuaian semantik terhadap *seed list* istilah teknis yang telah ditetapkan. Berdasarkan referensi teoretis, suatu kata dikategorikan sebagai kata teknis apabila nilai *similarity*-nya berada di atas ambang batas (*threshold*) sebesar 0,7.

public
technical_seed_list
id
words
created_at
created_by
updated_at
updated_by
is_deleted
word_embedding

Gbr. 13 Database Table Technical Seed List

Variabel *Specificity* dihitung melalui *embedding similarity BERT* dengan *technical seed list*. *Technical seed list* merupakan kumpulan istilah teknis yang digunakan sebagai acuan dalam mengidentifikasi kata-kata spesifik domain keilmuan pada judul referensi ilmiah untuk perhitungan variabel *specificity*. Daftar ini disusun melalui proses kurasi dari berbagai sumber terbuka, meliputi *Academic Word List (AWL)*, *Glossary of Computer Science* yang dipublikasikan oleh Wikipedia dan ACM, serta repositori kosakata teknis berbasis *Natural Language Processing* yang tersedia di GitHub.



Gbr. 14 Flowchart Proses Penilaian Specificity Score

Seluruh istilah kemudian disaring dan disesuaikan berdasarkan relevansinya terhadap domain Teknik Informatika dan berbagai bidang keilmuan yang umum dijumpai pada judul tugas akhir di Universitas Negeri Surabaya, sehingga menghasilkan daftar istilah yang mencakup *machine learning*, *software engineering*, *database*, jaringan komputer, *artificial intelligence*, serta bidang ilmu umum lainnya. Proses kurasi

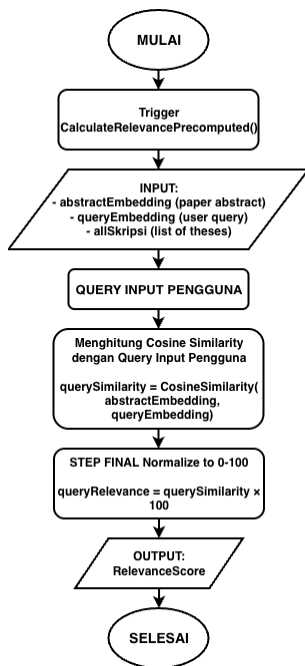
mempertimbangkan tingkat spesifisitas istilah terhadap domain teknologi dan frekuensi kemunculannya dalam publikasi ilmiah, dengan mengecualikan istilah yang terlalu umum serta memasukkan terminologi teknis seperti nama algoritma, metode, dan *framework*. Hasil akhirnya berupa *technical seed list* statis yang disimpan pada tabel *technical seed list* di basis data Gbr. 13 dan dapat diperbarui secara berkala oleh administrator agar tetap selaras dengan perkembangan topik penelitian dan terminologi pada jurnal internasional. Alur proses perhitungan *relevance* ditunjukkan pada Gbr. 14.

4. *Relevance Score*

Perhitungan *relevance* dilakukan untuk mengukur tingkat kesesuaian antara topik penelitian yang dimasukkan oleh pengguna dengan referensi jurnal *Semantic Scholar* [12]. Referensi yang memiliki kesamaan makna yang tinggi dengan input pengguna akan memiliki nilai *relevance* yang lebih tinggi. Tingkat kesamaan semantik kemudian dihitung menggunakan metode *cosine similarity* untuk memperoleh nilai *relevance* menggunakan Persamaan (8).

$$RelSet = \{sim_i | sim_i \geq 0.7\} \quad (8)$$

Alur proses perhitungan *relevance* ditunjukkan pada Gbr. 15.



Gbr. 15 Flowchart Proses Penilaian *Relevance Score*

2.2.9 Metode *Analytical Hierarchy Process*

Penentuan bobot perbandingan antar variabel dalam matriks *Analytical Hierarchy Process (AHP)* didasarkan pada studi literatur mengenai faktor-faktor yang paling sering digunakan dalam menilai kualitas *research output*. Hasil telaah

menunjukkan bahwa *novelty* merupakan variabel yang paling penting karena kontribusi ilmiah baru menjadi indikator utama kualitas penelitian. Dalam penelitian ini, *novelty* diukur tidak hanya berdasarkan keunikan semantik teks, tetapi juga mempertimbangkan *scientific importance* melalui *citation rate* untuk memastikan kebaruan yang diperoleh memiliki nilai ilmiah yang diakui oleh komunitas akademik. Selanjutnya, *clarity* ditempatkan pada prioritas kedua karena kejelasan formulasi ide dan struktur penulisan berpengaruh langsung terhadap pemahaman pembaca serta kualitas penyampaian penelitian.

Variabel *relevance* menempati urutan ketiga karena kesesuaian topik dengan domain keilmuan dan tren penelitian merupakan faktor penting dalam evaluasi referensi, meskipun tingkat kepentingannya berada di bawah *novelty* dan *clarity*. Sementara itu, *specificity* berada pada urutan keempat karena istilah teknis lebih berfungsi sebagai penguat konteks dibandingkan indikator utama kualitas referensi. Berdasarkan hasil studi literatur, diperoleh urutan prioritas *Novelty > Clarity > Relevance > Specificity*, yang kemudian digunakan sebagai dasar dalam penyusunan matriks perbandingan berpasangan. Nilai skala Saaty (1–9) selanjutnya diterapkan untuk merepresentasikan tingkat kepentingan relatif antar variabel sehingga diperoleh matriks perbandingan yang sistematis, logis, dan komprehensif [13]. Seperti ditampilkan pada Tabel I.

TABEL I
NILAI PERBANDINGAN BERPASANGAN SKALA SAATY

Pasangan Variabel	Nilai Saaty	Justifikasi Berdasarkan Literatur
<i>Novelty > Clarity</i>	3	Dvir et al. (2019) menunjukkan bahwa kejelasan judul meningkatkan aksesibilitas, namun Savov et al. (2021) menegaskan bahwa <i>novelty</i> lebih berpengaruh dalam evaluasi ilmiah.
<i>Novelty > Relevance</i>	5	Dewanda et al. (2023) menunjukkan bahwa relevansi penting, tetapi <i>novelty</i> merupakan parameter utama dalam menilai kontribusi ilmiah.
<i>Novelty > Specificity</i>	7	Wong et al. (2023) menekankan bahwa istilah teknis tidak menentukan nilai penelitian, sementara <i>novelty</i> adalah indikator kunci kontribusi ilmiah.
<i>Clarity > Relevance</i>	3	Studi <i>readability</i> oleh P.Pottier et al. (2024) menunjukkan bahwa kejelasan meningkatkan penerimaan editorial lebih tinggi dibanding relevansi.
<i>Clarity > Specificity</i>	5	<i>Specificity</i> hanya mendukung konteks teknis, sedangkan <i>clarity</i> memengaruhi pemahaman langsung pembaca.

<i>Relevance > Specificity</i>	3	<i>Relevance</i> menentukan keterkaitan topik dengan domain ilmiah, sementara <i>specificity</i> bersifat pelengkap linguistik.
-----------------------------------	---	---

Setelah matriks perbandingan berpasangan disusun, setiap kolom matriks dinormalisasi dan bobot masing-masing kriteria dihitung menggunakan rata-rata nilai pada setiap baris matriks ternormalisasi. Bobot yang diperoleh kemudian diuji tingkat konsistensinya melalui perhitungan nilai λ_{max} , *Consistency Index (CI)*, dan *Consistency Ratio (CR)*, dengan nilai $CR < 0,10$ digunakan sebagai indikator bahwa penilaian yang diberikan telah cukup konsisten. Terakhir, skor akhir tiap judul dihitung sebagai jumlah tertimbang (*weighted sum*) dari skor variabel menggunakan Persamaan (9).

$$Final\ Score = \sum_{i=1}^4 \omega_i \times S_i \quad (9)$$

Berdasarkan hasil tinjauan literatur, diperoleh urutan tingkat kepentingan variabel yaitu *Novelty > Clarity > Relevance > Specificity*. Urutan tersebut menjadi dasar dalam penyusunan matriks perbandingan berpasangan menggunakan skala Saaty 1–9 sebagaimana ditunjukkan pada Tabel II.

TABEL II
Matriks Perbandingan Berpasangan

Kriteria	<i>Novelty</i>	<i>Clarity</i>	<i>Relevance</i>	<i>Specificity</i>
<i>Novelty</i>	1	3	5	7
<i>Clarity</i>	0.3333	1	3	5
<i>Relevance</i>	0.2	0.3333	1	3
<i>Specificity</i>	0.1429	0.2	0.3333	1
Jumlah	1.6762	4.5333	9.333	16

Selanjutnya, bobot hasil AHP digunakan untuk menghitung skor akhir setiap judul melalui metode *weighted sum* berdasarkan nilai masing-masing variabel yang telah diperoleh sebelumnya. Matriks ternormalisasi disajikan pada Tabel III.

TABEL III
Matriks Ternormalisasi

Kriteria	<i>Novelty</i>	<i>Clarity</i>	<i>Relevance</i>	<i>Specificity</i>	Rata-rata
<i>Novelty</i>	0.5966	0.6622	0.5357	0.4375	0.5580
<i>Clarity</i>	0.1990	0.2207	0.3214	0.3125	0.2634
<i>Relevance</i>	0.1193	0.0736	0.1071	0.1875	0.1219

<i>Specificity</i>	0.0852	0.0441	0.0357	0.0625	0.0569
--------------------	--------	--------	--------	--------	--------

Bobot akhir (ω) disajikan sebagai persentase:

- $\omega(Novelty) \approx 0.5580$ (55.80 %)
- $\omega(Clarity) \approx 0.2634$ (26.34 %)
- $\omega(Relevance) \approx 0.1219$ (12.19 %)
- $\omega(Specificity) \approx 0.0569$ (5.69 %)

Setelah bobot valid, nilai akhir tiap judul dihitung sebagai penjumlahan bobot \times skor (*score* tiap variabel dalam skala 0–100). Maka *final score* bisa didapatkan menggunakan Persamaan (9).

C. Development

2.3.1 Integrasi Semantic Scholar API

Pada tahap ini, sistem melakukan integrasi dengan *Semantic Scholar API* untuk memperoleh referensi ilmiah yang digunakan sebagai kandidat evaluasi dan dasar penyusunan rekomendasi. Proses dimulai dengan mengirimkan *request* menggunakan *query* yang dibentuk dari kata kunci domain program studi di Universitas Negeri Surabaya, kemudian sistem menerima data berformat *JSON* yang berisi metadata publikasi seperti *title*, *abstract*, *authors*, tahun publikasi, *citation count*, *venue*, dan *semantic URL*. Data yang diperoleh selanjutnya melalui proses *parsing* dan *cleaning* untuk menghapus entri duplikat, tidak lengkap, atau tidak relevan sehingga hanya publikasi valid yang digunakan. Pemilihan sumber data dilakukan dengan membandingkan beberapa *open scholarly APIs*, yaitu *Semantic Scholar API*, *Crossref API*, *OpenAlex API*, dan *Scopus API*, berdasarkan kriteria ketersediaan *abstract*, dukungan terhadap analisis semantik, akses terbuka, serta kompleksitas integrasi. Dibandingkan alternatif lainnya, *Semantic Scholar API* dinilai lebih sesuai karena menyediakan metadata yang mendukung analisis semantik berbasis teks menggunakan *BERT*, bersifat *open-source*, dan memiliki proses integrasi yang lebih sederhana.

2.3.2 Pemberian Skor Pada Variabel Penilaian Kualitas

Tahap ini merupakan inti dari sistem yang bertugas menghitung kualitas setiap referensi ilmiah berdasarkan empat variabel, yaitu *novelty*, *clarity*, *specificity*, dan *relevance*, menggunakan representasi semantik yang dihasilkan oleh model *BERT*. Variabel *novelty* dihitung berdasarkan tingkat kemiripan semantik antar referensi (*intra-reference similarity*), sedangkan *clarity* diukur menggunakan nilai *standard deviation* dari hubungan semantik antar token dalam judul untuk menilai konsistensi linguistik. Variabel *specificity* diperoleh melalui identifikasi istilah teknis menggunakan *technical seed list*, sementara *relevance* dihitung dari rata-rata nilai *cosine similarity* antara referensi dan *semantic query*, dengan hanya mempertimbangkan nilai di atas *threshold* 0,7 sebagai indikator relevansi yang signifikan. Seluruh skor yang dihasilkan kemudian dinormalisasi menjadi nilai terstandarisasi untuk masing-masing variabel. Hasil tersebut selanjutnya

digunakan dalam proses pembobotan menggunakan *Analytical Hierarchy Process (AHP)* sehingga penilaian kualitas referensi dapat dilakukan secara kuantitatif, objektif, dan konsisten berdasarkan parameter linguistik serta semantik yang terukur.

D. Testing

Tahap *testing* dilakukan untuk mengevaluasi fungsionalitas, keandalan, dan validitas metode yang digunakan dalam sistem rekomendasi referensi tugas akhir berbasis *BERT* dan *AHP*. Pengujian fungsional dilakukan menggunakan metode *Black Box Testing* untuk memastikan seluruh fitur pada sisi pengguna, dosen pembimbing, dan administrator berjalan sesuai kebutuhan fungsional yang telah dirancang, dengan fokus pada kesesuaian input, proses, dan output system [14]. Selain itu, dilakukan uji kinerja model *BERT* melalui pendekatan *binary semantic similarity classification* dengan membandingkan hasil prediksi model terhadap *ground truth* yang diperoleh melalui *manual labeling*, kemudian dievaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. Pengujian juga mencakup validasi metode *Analytical Hierarchy Process (AHP)* melalui perhitungan nilai λ_{max} , *Consistency Index (CI)*, dan *Consistency Ratio (CR)* untuk memastikan konsistensi pembobotan kriteria [19]. Seluruh hasil pengujian didokumentasikan dan dianalisis untuk mengidentifikasi kesesuaian sistem terhadap spesifikasi yang telah ditetapkan. Melalui rangkaian pengujian tersebut, sistem diharapkan mampu menghasilkan rekomendasi referensi yang fungsional, akurat secara semantik, serta memiliki mekanisme pembobotan yang konsisten dan dapat dipertanggungjawabkan.

E. Deployment

Tahap *deployment* dilakukan secara bertahap dari lingkungan pengembangan (*localhost*) hingga server produksi untuk memastikan seluruh fitur sistem berjalan dengan baik sebelum diakses secara publik. Setelah pengujian lokal berhasil, aplikasi dipindahkan ke server produksi melalui proses *secure file transfer* yang disertai konfigurasi lingkungan, seperti variabel lingkungan, koneksi basis data, dan layanan server. Implementasi *deployment* menggunakan layanan *cloud hosting* Hostinger VPS KVM1 yang memberikan kontrol penuh terhadap konfigurasi server berbasis Linux. Server dikonfigurasi menggunakan Nginx sebagai *reverse proxy* yang menghubungkan *client* dengan layanan *backend* berbasis Golang, sementara aplikasi *frontend* Vue.js dikompilasi menjadi *static files* dan disajikan melalui Nginx. Selain itu, basis data PostgreSQL ditempatkan pada lingkungan yang terpisah secara logis untuk meningkatkan keamanan, performa, dan stabilitas sistem pada lingkungan produksi.

F. Review

Review website rekomendasi referensi tugas akhir dilakukan menggunakan metode *System Usability Scale (SUS)*. *Review* ini dilakukan kepada pengguna akhir yaitu mahasiswa, dosen pembimbing dan admin. Metode *SUS* dipilih karena

sederhana dan efektif dalam mengevaluasi tingkat kegunaan (*usability*) *website* berdasarkan pengalaman pengguna.

2.6.1 System Usability Scale

SUS atau *System Usability Scale* adalah skala evaluasi kegunaan yang terdiri dari 10 pernyataan [38]. Responden diminta untuk memberikan penilaian pada skala 1-5, dengan rincian:

- 1: Sangat Tidak Setuju
- 2: Tidak Setuju
- 3: Netral
- 4: Setuju
- 5: Sangat Setuju

2.6.2 Pernyataan System Usability Scale

Berikut merupakan tabel yang berisi 10 pernyataan dalam metode *SUS* yang digunakan untuk mengukur kegunaan *website* :

TABEL IV
 FORM PERNYATAAN SUS

Pernyataan	Skala 1-5
Fitur rekomendasi referensi tugas akhir pada aplikasi ini sangat membantu saya mendapatkan referensi jurnal internasional berkualitas mencakup multikriteria (<i>novelty, clarity, specificity, relevance</i>)	
Saya kesulitan memahami istilah atau hasil analisis yang ditampilkan pada fitur rekomendasi referensi tugas akhir.	
Proses penggunaan fitur rekomendasi referensi tugas akhir terasa mudah dan tidak membingungkan.	
Proses rekomendasi referensi tugas akhir pada aplikasi memakan waktu terlalu lama dan terasa melelahkan.	
Informasi rekomendasi tren topik penelitian terkait yang muncul otomatis sangat membantu dan mudah dipahami.	
Bahasa yang digunakan dalam aplikasi sulit dipahami.	
Desain aplikasi responsif dan mudah diakses melalui berbagai perangkat (<i>smartphone, laptop</i>).	
Proses registrasi dan <i>login</i> di aplikasi ini sangat sederhana dan cepat.	
Saya memerlukan bantuan teknis eksternal untuk dapat menggunakan seluruh fitur aplikasi dengan baik.	

Aplikasi ini memberikan informasi yang bermanfaat untuk membantu saya dalam penyusunan judul dan proses akademik.

2.6.3 Interpretasi Hasil *System Usability Scale*

Setelah responden memberikan penilaian pada setiap pernyataan, skor *SUS* dihitung dengan langkah sebagai berikut:

1. Untuk pernyataan bernomor ganjil, nilai skor dikurangi 1 dari jawaban responden. Dapat dilihat pada Persamaan (10).

$$S^i = p^i - 1, \text{ jika } i \text{ ganjil} \quad (10)$$

2. Untuk pernyataan bernomor genap, nilai skor dihitung dengan cara 5 dikurangi jawaban responden. Dapat dilihat pada Persamaan (11).

$$S^i = 5 - p^i, \text{ jika } i \text{ genap} \quad (11)$$

3. Jumlahkan semua skor dari langkah 1 dan 2 menggunakan Persamaan (12).

$$\sum_{i=1}^n S^i \quad (12)$$

4. Hasil penjumlahan kemudian dikalikan dengan 2,5 untuk mendapatkan skor akhir *SUS*. Dapat dilihat pada Persamaan (13).

$$SUS = (\sum_{i=1}^n S^i) \times 2.5 \quad (13)$$

5. Skor akhir tiap responden dihitung, lalu diambil rata-rata untuk mendapatkan nilai *SUS* keseluruhan. Dapat dilihat pada Persamaan (14).

$$\bar{x} = \frac{\sum x}{n} \quad (14)$$

Keterangan :

S^i = skor tiap pernyataan setelah dikonversi

P^i = nilai yang diberikan responden pada pernyataan ke- i

N = jumlah total pernyataan

SUS = skor *SUS* per responden

\bar{x} = skor rata-rata *SUS*

$\sum x$ = jumlah skor *SUS* keseluruhan

n = jumlah responden

Nilai akhir *SUS* berada pada rentang 0 – 100. Secara umum, skor di atas 68 dianggap menunjukkan bahwa sistem memiliki tingkat kegunaan yang baik. Berikut merupakan Skala pada metode *System Usability Scale*:

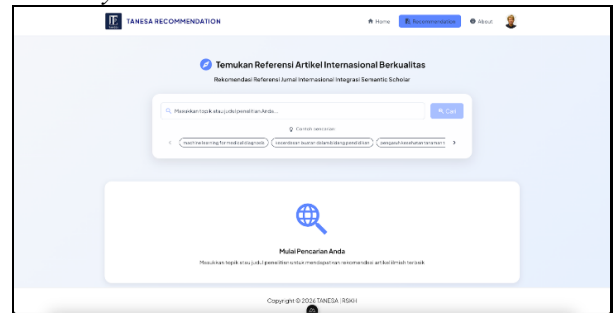
- $\geq 80.3 \rightarrow$ Excellent (Sangat Baik) [A]
- $68 - 80.3 \rightarrow$ Good (Baik) [B]
- $51 - 68 \rightarrow$ Fair (Cukup) [C]
- $\leq 51 \rightarrow$ Poor (Buruk) [D/E]

III. HASIL DAN PEMBAHASAN

A. Hasil

1. Halaman Rekomendasi Referensi

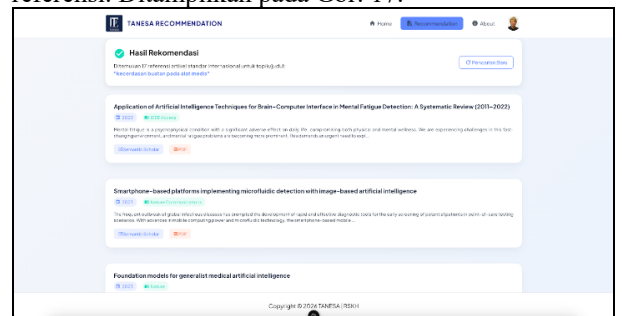
Halaman rekomendasi referensi merupakan halaman utama dalam proses pencarian referensi ilmiah. Pada Gbr. 16 diperlihatkan bahwa halaman ini menyediakan sebuah *form input* berupa *text box* yang berfungsi untuk memasukkan topik atau judul penelitian yang diminati pengguna. Melalui form tersebut, pengguna dapat menentukan kata kunci yang akan digunakan dalam proses pencarian artikel ilmiah. Setelah pengguna menekan tombol “Cari”, sistem akan menampilkan indikator *loading* dan memproses data secara otomatis. Sistem kemudian menampilkan daftar rekomendasi artikel yang diperoleh dari basis data *Semantic Scholar* berdasarkan hasil proses pencarian yang telah dilakukan. Urutan rekomendasi artikel ditentukan berdasarkan nilai akhir yang diperoleh dari hasil pembobotan *Analytical Hierarchy Process*.



Gbr. 16 Antarmuka Halaman Rekomendasi Referensi

2. Hasil *Ranking*

Sistem kemudian menampilkan daftar rekomendasi artikel jurnal internasional yang diperoleh dari basis data *Semantic Scholar* berdasarkan hasil proses pencarian yang telah dilakukan sebelumnya. Urutan rekomendasi artikel jurnal ditentukan berdasarkan nilai akhir yang diperoleh dari hasil pembobotan *Analytical Hierarchy Process* terhadap beberapa kriteria *novelty*, *clarity*, *specificity*, dan *relevance*. Setiap artikel yang direkomendasikan ditampilkan beserta informasi penting yang mendukung pengguna dalam melakukan peninjauan referensi. Ditampilkan pada Gbr. 17.



Gbr. 17 Antarmuka Perankingan Referensi

B. Pembahasan

1. *Activation Multilingual BERT*

Pada tahap implementasi, model *Multilingual BERT* (*mBERT*) digunakan sebagai layanan analisis semantik untuk menghasilkan *embedding vector* dari teks penelitian berbahasa Indonesia dan Inggris. Model yang digunakan adalah *bert-base-multilingual-cased* dari *HuggingFace* karena mampu menghasilkan representasi semantik multibahasa secara kontekstual.

Sistem *backend* berbasis *Golang* berkomunikasi dengan layanan *mBERT* melalui *REST API* menggunakan format *JSON*. Input berupa topik penelitian pengguna dikirimkan ke layanan *embedding* untuk diproses menjadi representasi vektor semantik yang selanjutnya digunakan dalam perhitungan *cosine similarity* pada proses *semantic similarity analysis*.

Layanan *embedding* juga mendukung pemrosesan *batch* sehingga sistem mampu melakukan pencocokan terhadap banyak kandidat referensi dari *Semantic Scholar* secara lebih efisien. Nilai *similarity* yang dihasilkan kemudian digunakan dalam perhitungan variabel *novelty*, *clarity*, *specificity*, dan *relevance* sebagai dasar pemeringkatan referensi penelitian.

2. Perhitungan *Novelty Score*

Implementasi perhitungan *novelty* dilakukan dengan menganalisis tingkat kebaruan relatif suatu referensi terhadap referensi lain yang diperoleh dari *Semantic Scholar*. Sistem menghitung nilai *cosine similarity* antar referensi untuk memperoleh nilai kemiripan tertinggi (*maximum similarity*), kemudian mengonversinya menjadi *semantic novelty* sehingga referensi yang lebih berbeda memperoleh skor lebih tinggi. Untuk menghindari *novelty* semu, sistem juga menghitung *scientific importance* berdasarkan *citation rate* yang dinormalisasi dari jumlah sitasi dan tahun publikasi [16]. Kedua komponen tersebut digabungkan untuk menghasilkan *novelty score* akhir. Hasil implementasi menunjukkan bahwa referensi dengan kebaruan semantik tinggi tetapi sitasi rendah mengalami penurunan skor, sedangkan referensi yang memiliki kebaruan dan pengakuan ilmiah yang baik memperoleh skor *novelty* tertinggi.

3. Perhitungan *Clarity Score*

Sistem melakukan *preprocessing* dan pembentukan *embedding* menggunakan *BERT*, kemudian menghitung nilai *cosine similarity* antar pasangan token dalam judul. Nilai-nilai *similarity* tersebut digunakan untuk menghitung *standard deviation* sebagai indikator konsistensi linguistik. Semakin kecil nilai *standard deviation*, semakin konsisten hubungan semantik antar kata sehingga judul dianggap lebih jelas dan mudah dipahami. Pendekatan ini memungkinkan sistem mengevaluasi kualitas struktur semantik judul secara objektif berdasarkan representasi makna yang dihasilkan model *BERT*.

4. Perhitungan *Specificity Score*

Setelah proses *preprocessing* dan pembentukan *embedding*, setiap token dibandingkan dengan *technical seed list* menggunakan *cosine similarity*. Token yang memiliki nilai *similarity* di atas ambang batas 0,7 dikategorikan sebagai istilah teknis yang relevan dengan domain keilmuan. Selanjutnya, sistem menghitung proporsi istilah teknis terhadap jumlah

keseluruhan token dalam judul untuk menghasilkan skor *specificity*. Semakin banyak istilah teknis yang terdeteksi, semakin tinggi skor *specificity* yang diperoleh referensi tersebut.

5. Perhitungan *Relevance Score*

Sistem membentuk *semantic query* dari input pengguna menggunakan *BERT*, kemudian menghitung *cosine similarity* terhadap *embedding* referensi yang tersimpan pada basis data. Hanya referensi dengan nilai *similarity* di atas *threshold* 0,7 yang dipertimbangkan sebagai kandidat relevan. Nilai *similarity* yang diperoleh selanjutnya dikonversi menjadi *relevance score* untuk setiap publikasi. Dengan pendekatan ini, sistem dapat mengidentifikasi referensi yang memiliki keterkaitan semantik paling tinggi dengan kebutuhan penelitian pengguna secara lebih akurat dibandingkan pencocokan berbasis kata kunci.

6. Perhitungan *Final Score* dengan Pembobotan *AHP*

Pada tahap implementasi *final score*, sistem menggabungkan empat variabel kualitas referensi, yaitu *novelty*, *clarity*, *relevance*, dan *specificity*, menggunakan bobot prioritas yang diperoleh dari metode *Analytical Hierarchy Process (AHP)* untuk menghasilkan skor akhir sebagai dasar pemeringkatan referensi secara otomatis. Proses ini diterapkan pada 100 kandidat referensi hasil *semantic retrieval* yang sebelumnya telah melalui perhitungan masing-masing variabel kualitas. Skor akhir dihitung menggunakan metode *weighted sum*, yaitu dengan mengalikan setiap skor variabel dengan bobot *AHP* yang sesuai, kemudian menjumlahkan seluruh kontribusi variabel. Pendekatan ini memungkinkan sistem tidak hanya mempertimbangkan kesesuaian topik penelitian, tetapi juga aspek kebaruan, kejelasan, dan kekhususan teknis dari setiap publikasi ilmiah. Semakin tinggi nilai *final score* yang dihasilkan, semakin tinggi pula prioritas referensi yang direkomendasikan kepada pengguna.

Sebagai contoh implementasi, publikasi "*Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses*". Berdasarkan hasil implementasi sebelumnya, publikasi tersebut memiliki: *novelty score* = 74,98; *clarity score* = 90,85; *relevance score* = 70,70; dan *specificity score* = 42,85. Selanjutnya, sistem mengambil bobot prioritas hasil *AHP* yang telah diperoleh pada tahap sebelumnya, yaitu:

- $\omega(\text{Novelty}) \approx 0.5580$ (55.80 %)
- $\omega(\text{Clarity}) \approx 0.2634$ (26.34 %)
- $\omega(\text{Relevance}) \approx 0.1219$ (12.19 %)
- $\omega(\text{Specificity}) \approx 0.0569$ (5.69 %)

Dihitung menggunakan Persamaan (9), sistem menghasilkan perhitungan, sebagai berikut:

$$\text{Final Score} = (74.98 \times 0.5580) + (90.85 \times 0.2634) + (70.70 \times 0.1219) + (42.85 \times 0.0569)$$

Kemudian seluruh kontribusi dijumlahkan sehingga menghasilkan:

$$\text{Final Score} = 41.84 + 23.93 + 8.62 + 2.44 = 76.83$$

Hasil tersebut menunjukkan bahwa publikasi memiliki kualitas referensi yang cukup tinggi karena memperoleh *novelty* dan

clarity yang dominan terhadap query penelitian pengguna. Nilai tersebut lebih tinggi dibandingkan publikasi "Customer Sentiment Analysis in Hotel Reviews Through NLP Techniques" yang memperoleh final score sebesar 70,33. Perbedaan hasil ini menunjukkan bahwa meskipun suatu referensi memiliki nilai *relevance* dan *specificity* yang baik, skor akhir tetap sangat dipengaruhi oleh nilai *novelty* karena variabel tersebut memiliki bobot prioritas tertinggi dalam AHP. Hasil implementasi ini menunjukkan bahwa mekanisme pemeringkatan yang dikembangkan mampu menghasilkan rekomendasi referensi yang lebih komprehensif dengan mempertimbangkan kontribusi relatif dari setiap variabel kualitas secara terukur dan konsisten.

7. Evaluasi Kinerja *Semantic Similarity* Menggunakan *mBERT*

Validasi model dilakukan dengan pendekatan *binary semantic similarity classification* untuk mengevaluasi apakah model *BERT* dapat lebih baik dalam membedakan pasangan judul yang memiliki kedekatan makna atau tidak. Proses penilaian dilakukan dengan membuat *ground truth* berdasarkan *manual labeling*, di mana peneliti memberikan label *similar* atau *not similar* pada pasangan judul sampel berdasarkan pemahaman makna konteks topik.

Hasil pelabelan kemudian dibandingkan dengan keluaran model menggunakan *cosine similarity*, di mana nilai ≥ 0.7 dikategorikan sebagai *similar* berdasarkan studi literatur. Performa model dianalisis menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kemampuan model dalam mengklasifikasikan kesesuaian semantik secara konsisten [18]. Hasil evaluasi ini menjadi dasar untuk memastikan bahwa *BERT* memenuhi syarat reliabilitas semantik sebelum digunakan sebagai komponen inti dalam sistem penilaian kualitas referensi.

Untuk mengukur kualitas prediksi digunakan *confusion matrix* yang terdiri dari empat komponen utama, yaitu :

- *True Positive (TP)*: pasangan judul yang secara manual dilabeli *similar* dan diprediksi *similar* oleh model.
- *True Negative (TN)*: pasangan judul yang dilabeli *not similar* dan diprediksi *not similar*.
- *False Positive (FP)*: pasangan judul yang sebenarnya *not similar* namun diprediksi *similar*.
- *False Negative (FN)*: pasangan judul yang sebenarnya *similar* namun diprediksi *not similar*.

Pengujian dilakukan dengan menggunakan 10 *sample query* masukan pengguna, yaitu:

1. "Artificial Intelligence in the Field of Health",
2. "Plant Health Influences on Air Quality",
3. "Criminal Accountability Analysis in Cyberbullying Cases",
4. "Clinical Psychological Roles in Juvenile Behavioral Studies",
5. "Machine Learning in the Auto Parking System",
6. "Artificial Intelligence for the Development of Smart Village in Tanggul",

7. "Full Immunization Relationship with the Impairment on Toddlers",
8. "Analysis of the Role of the Law in the Countermeasures of Female Violence",
9. "Genetic Diversity Analysis of Local Plants",
10. "Digital Marketing Strategies to Improve UMKM Sales".

Setiap query kemudian dibandingkan dengan 10 judul jurnal yang diambil dari basis data *Semantic Scholar*, sehingga total terdapat 100 pasangan data pengujian. Proses pelabelan manual (*ground truth*) dilakukan oleh tiga penilai, yaitu penulis dan dua narasumber ahli dari *Senior Software Engineer* di PT Greatsoft Solusi Indonesia. Untuk meningkatkan reliabilitas pelabelan, proses anotasi dilakukan dengan mempertimbangkan kesepakatan antar penilai (*inter-rater agreement*) terhadap tingkat kesesuaian konteks dan topik antara query serta judul jurnal, sehingga dihasilkan label *mirip* atau *tidak mirip* yang lebih objektif.

Selanjutnya, sistem menghitung nilai *cosine similarity* dari *embedding* kedua teks menggunakan model *BERT*. Berdasarkan *threshold* yang telah ditetapkan, yaitu *cosine similarity* $\geq 0,7$, sistem mengklasifikasikan pasangan teks sebagai mirip, sedangkan nilai di bawah *threshold* dikategorikan sebagai tidak mirip. Hasil klasifikasi sistem kemudian dibandingkan dengan *label ground truth* untuk mengidentifikasi *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Nilai-nilai tersebut selanjutnya digunakan untuk menghitung metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

Implementasi perhitungan metrik evaluasi dilakukan dengan menggunakan nilai *True Positive (TP)* sebesar 69, *True Negative (TN)* sebesar 13, *False Positive (FP)* sebesar 12, dan *False Negative (FN)* sebesar 6. Berdasarkan nilai tersebut, nilai *accuracy* dihitung menggunakan Persamaan (15), nilai *precision* menggunakan Persamaan (16), dan nilai *recall* menggunakan Persamaan (17).

- *Accuracy*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$Accuracy = \frac{69 + 13}{69 + 13 + 12 + 6}$$

$$Accuracy = \frac{82}{100} = 0.82 / 82\%$$

- *Precision*

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Precision = \frac{69}{69 + 12}$$

$$Precision = \frac{69}{81} = 0.852 / 85.2\%$$

- *Recall*

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$Recall = \frac{69}{69 + 6}$$

$$\text{Recall} = \frac{69}{75} = 0.92 / 92\%$$

Setelah didapatkan ketiga nilai *accuracy*, *precision*, dan *recall*, kita perlu mencari nilai *F1-Score* menggunakan Persamaan (18).

- *F1-Score*

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

$$F1 \text{ Score} = 2 \times \frac{0.852 \times 0.92}{0.852 + 0.92}$$

$$F1 \text{ Score} = 0.885$$

Untuk mengevaluasi efektivitas pendekatan semantic similarity yang digunakan, penelitian ini juga melakukan perbandingan performa antara metode *Multilingual BERT (mBERT)*, *TF-IDF*, dan *Word2Vec* menggunakan dataset pengujian yang sama. Evaluasi dilakukan menggunakan pendekatan *binary semantic similarity classification* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Pendekatan *TF-IDF* digunakan sebagai baseline tradisional berbasis frekuensi kata, sedangkan *Word2Vec* digunakan sebagai baseline word *embedding* kontekstual sederhana. Sementara itu, *mBERT* digunakan sebagai pendekatan *contextual embedding* berbasis *transformer* yang mampu memahami hubungan semantik secara *bidirectional*.

TABEL V

PERBANDINGAN PEFORMA METODE SEMANTIC SIMILARITY

Metode	Accuracy	Precision	Recall	F1-Score
<i>TF-IDF</i>	71%	73.4%	78%	75.6%
<i>Word2Vec</i>	76%	79.1%	83%	81%
<i>mBERT</i>	82%	85.2%	92%	88.5%

Berdasarkan hasil pengujian pada Tabel V, metode *Multilingual BERT (mBERT)* menghasilkan performa terbaik pada seluruh metrik evaluasi dibandingkan *TF-IDF* dan *Word2Vec*. Peningkatan performa paling signifikan terlihat pada nilai *recall* dan *F1-score*, yang menunjukkan kemampuan *mBERT* dalam memahami hubungan semantik antar topik penelitian secara lebih konsisten dan kontekstual.

C. Pengujian System Usability Scale (SUS)

Evaluasi sistem *website* Tanesa dilakukan menggunakan metode *System Usability Scale (SUS)* yang melibatkan 50 responden yang terdiri atas mahasiswa, dosen pembimbing, dan administrator sistem yang diminta menggunakan fitur utama Tanesa. Setelah penggunaan sistem, responden mengisi kuesioner SUS yang terdiri atas 10 pernyataan dengan skala Likert 1–5. Hasil perhitungan menunjukkan skor rata-rata SUS sebesar **70,1** dari skala 0–100. Berdasarkan interpretasi SUS, nilai tersebut termasuk dalam kategori *Good*, yang menunjukkan bahwa sistem memiliki tingkat efektivitas, efisiensi, dan kepuasan pengguna yang baik.

D. Pengujian Black Box Testing

Pengujian website dilakukan menggunakan metode *Black Box Testing* untuk memastikan seluruh fungsi sistem berjalan sesuai kebutuhan sebelum diterapkan pada lingkungan produksi. Pengujian melibatkan dua peran utama, yaitu pengguna dan administrator, dengan total 22 skenario pengujian yang mencakup fitur-fitur utama pada masing-masing peran. Setiap skenario dievaluasi berdasarkan kesesuaian antara keluaran sistem dan hasil yang diharapkan tanpa meninjau proses internal aplikasi. Hasil pengujian oleh total 14 penguji menunjukkan bahwa seluruh skenario berhasil dijalankan tanpa ditemukan kegagalan fungsi maupun kendala yang menghambat operasional sistem.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka kesimpulan yang diperoleh sesuai dengan rumusan masalah adalah sebagai berikut :

1. Sistem rekomendasi referensi jurnal berbasis website yang mengintegrasikan analisis semantik menggunakan model *BERT* dan metode *Analytical Hierarchy Process (AHP)* berhasil dibangun dan diimplementasikan dengan baik. Sistem menerapkan mekanisme pemrosesan terpisah antara *query* pengguna yang diproses secara *real-time* dan data referensi yang diproses secara *batch* oleh administrator, sehingga mampu meningkatkan efisiensi komputasi. Hasil pengujian menunjukkan bahwa waktu rata-rata *preprocessing* dan pembentukan *embedding* berada pada kisaran 0,2–0,5 detik per data, yang dipengaruhi oleh panjang teks yang ada dan beban sistem saat menjalankan proses. Selain itu, proses pencarian rekomendasi referensi secara keseluruhan membutuhkan waktu komputasi rata-rata sekitar 40–100 detik, yang dipengaruhi oleh jumlah kandidat referensi, panjang teks, serta beban sistem saat proses berlangsung.
2. Sistem mampu menampilkan daftar referensi ilmiah secara terurut berdasarkan tren dan relevansi melalui integrasi dengan *Semantic Scholar API* serta pembobotan menggunakan metode *AHP*. Hasil evaluasi menunjukkan bahwa model memiliki performa tinggi dengan nilai *accuracy* sebesar 82%, *precision* 85,2%, *recall* 92%, dan *F1-score* 0,885, serta metode *AHP* menghasilkan bobot kriteria yang valid dengan nilai *Consistency Ratio (CR)* sebesar 0,044 (< 0,1). Distribusi bobot menunjukkan bahwa kriteria *novelty* memiliki pengaruh paling dominan sebesar 55,80%, diikuti *clarity* (26,34%), *relevance* (12,19%), dan *specificity* (5,69%). Selain itu, sistem terbukti mampu memberikan rekomendasi referensi yang objektif, relevan, dan sesuai dengan kebutuhan pengguna.

UCAPAN TERIMA KASIH

Penulis mengucapkan syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan. Terima kasih kepada mama, dosen pembimbing, keluarga, sahabat, dan seluruh pihak yang telah memberikan doa, dukungan, serta bantuan selama proses penelitian.

Apresiasi juga penulis sampaikan kepada diri sendiri atas ketabahan, ketekunan, dan kebesaran hati dalam menyelesaikan penelitian ini hingga tuntas.

REFERENSI

- [1] D. N. Lindang, A. Y. Muniar, A. Halid, dan A. Amiruddin, "Sistem Penentuan Kemiripan Antar Skripsi Menggunakan Metode Cosine Similarity Pada Perpustakaan," *SNTEI*, vol. 8, no. 1, 2022.
- [2] A. H. Nasrullah, "Integrasi Tf-Idf Dan Algoritma Cosine Similarity Untuk Deteksi Tingkat Kemiripan Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo)," *INTEC*, vol. 3, no. 3, 2024.
- [3] Andre, N. Suciati, H. Fabroyir, dan E. Pardede, "Educational Data Mining Clustering Approach: Case Study of Undergraduate Student Thesis Topic," *IEEE Access*, vol. 11, hlm. 130072–130088, 2023, doi: 10.1109/ACCESS.2023.3332818.
- [4] A. Hadhiatma, A. Azhari, dan Y. Suyanto, "A Scientific Paper Recommendation Framework Based on Multi-Topic Communities and Modified PageRank," *IEEE Access*, vol. 11, hlm. 25303–25317, 2023, doi: 10.1109/ACCESS.2023.3251189.
- [5] O. Karimah dan R. Bisma, "Analisis Faktor-Faktor Mempengaruhi Terhadap Penerimaan Sistem Monitoring Tugas Akhir dan Skripsi pada Civitas Akademik di Lingkungan UNESA (Universitas Negeri Surabaya)," *J. Emerg. Inf. Syst. Bus. Intell. JEISBI*, vol. 3, no. 2, hlm. 11–17, Apr 2022, doi: 10.26740/jeisbi.v3i2.45725.
- [6] humas-unesa, "Universitas Negeri Surabaya," Universitas Negeri Surabaya. Diakses: 22 Agustus 2025. [Daring]. Tersedia pada: //unesa.ac.id/
- [7] B. H. Irawan, M. S. H. Simarankir, dan E. Erlinna, "DETEKSI KEMIRIPAN JUDUL SKRIPSI MENGGUNAKAN ALGORITMA LEVENSHTAIN DISTANCE PADA KAMPUS STMIK MIC CIKARANG," *Eduatic - Sci. J. Inform. Educ.*, vol. 7, no. 2, Mei 2021, doi: 10.21107/edutic.v7i2.10051.
- [8] Savov, P., 2021. Measuring the Novelty of Scientific Papers. Ph.D. Thesis. Polish-Japanese Academy of Information Technology
- [9] K. A. Pradani dan L. H. Suadaa, "Automated Essay Scoring Menggunakan Semantic Textual Similarity Berbasis Transformer Untuk Penilaian Ujian Esai," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, no. 6, hlm. 1177–1184, Des 2023, doi: 10.25126/jtiik.2023107338.
- [10] D. Katarina, A. Nurrohman, dan A. S. Putra, "Decision Support System For The Best Student Selection Recommendation Using Ahp (Analytic Hierarchy Process) Method," *Int. J. Educ. Res.*, 2025.
- [11] "Systematic Improvement of User Engagement with Academic Titles Using Computational Linguistics," dipresentasikan pada InSITE 2020: Informing Science + IT Education Conferences: Jerusalem, 2020, hlm. 501–512. doi: 10.28945/4338.
- [12] E. L. Services, "FINER: a Research Framework | Elsevier Language Services," Elsevier Author Services - Articles. Diakses: 14 November 2025. [Daring]. Tersedia pada: <https://scientific-publishing.webshop.elsevier.com/research-process/finer-research-framework/>
- [13] R. F. Ramadan dan A. U. Firmansyah, "Perancangan Sistem Pendukung Keputusan Menggunakan Metode Analytic Hierarchy Process (AHP) dalam Menentukan Jenis Tanaman Pangan," *J. Teknol. Inform. Dan Komput.*, vol. 9, no. 1, hlm. 148–159, Mar 2023, doi: 10.37012/jtik.v9i1.1349.
- [14] Bawono, S. A. T., Purnomo, F. A., Purnomo, A., & Nuzuliaty, R. (2017). Conferec (Conference CMS). Indonesian Journal of Applied Informatics (IJAI), 2(1), 1–16. <https://jurnal.uns.ac.id/ijai/article/view/13170>
- [15] Wu, W., Zhang, C., & Zhao, Y. (2025). Automated novelty evaluation of academic paper: A collaborative approach integrating human and large language model knowledge. *Journal of the Association for Information Science and Technology*, 76(11), 1452–1469. <https://ideas.repec.org/a/bla/jinfst/v76y2025i11p1452-1469.html>
- [16] Apartis, S., Catalano, G., Consiglio, G., Costas, R., Delugas, E., Dulong de Rosnay, M., Grypari, I., Karasz, I., Klebel, T., Kormann, E., Manola, N., Papageorgiou, H., Seminaroti, E., Stavropoulos, P., Stoy, L., Traag, V. A., van Leeuwen, T., Venturini, T., Vignetti, S., Waltman, L., & Willemse, T. (2024). Open Science Impact Indicator Handbook. Zenodo. <https://doi.org/10.5281/zenodo.14538442>
- [17] Muhammad Ainurofiq Anwar Buhang dan Paramitha Nerisafitra, "Rancang Bangun Sistem Rekomendasi Deteksi Baby Blues Pasca Melahirkan Menggunakan Metode Certainty Factor," *J. Inform. Comput. Sci. JINACS*, vol. 6, no. 04, hlm. 1099–1108, Jun 2025, doi: 10.26740/jinacs.v6n04.p1099-1108.
- [18] R. Merdiansah, S. Siska, dan A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. Dan Sist. Inf. JIKOMSI*, vol. 7, no. 1, hlm. 221–228, Mar 2024, doi: 10.55338/jikomsi.v7i1.2895.
- [19] I. G. I. Sudipa dan K. Hardiatama, "Implementasi Metode Analytical Hierarchy Process Dan Interpolasi Linier Dalam Penentuan Lokasi Wisata Di Kabupaten Karangasem," *J.SAKTI*, vol. 5, no.2, hlm 866-878, 2021.