JIKTEKS : Jurnal Ilmu Komputer dan Teknologi Informasi
Vol. 04, No. 03, Agustus 2026, hlm. 187 - 195
ISSN : 2986-5417
DOI : doi.org/10.70404/jikteks.v4i03.724
Laman Jurnal: jurnal.faatuatua.com/index.php/JIKTEKS

Implementasi GPT API pada SIMANTAP untuk Deteksi Similarity dan Novelty Teks Akademik Berbasis NLP

Raafi Syarahil Azhar[1*], Subhanjaya Angga Atmaja[2], Yasri[3]
[1*, 2, 3]Teknik Informatika, Univeristas Kebangsaan Republik Indonesia, Indonesia
e-mail: [1*]r.syarahilazhar@gmail.com; [2]subhanjaya.angga.atmaja@fiksi.ukri.ac.id; [3]yasri.ukri@gmail.com

Keywords:

GPT API;
Natural Language Processing;
Similarity Detection;
Novelty Detection;
Large Language Model;
Semantic Similarity:
Academic Integrity.

ABSTRACT

The digitalization of academic documents increases the risk of overlapping ideas and paraphrase-based plagiarism, while existing similarity tools generally report only a numerical score without explaining the position of a study relative to prior work. This study implements the GPT API within the SIMANTAP application to support similarity and novelty detection of academic texts based on Natural Language Processing, and evaluates the capability of the system as well as user perception of its implementation. A quantitative approach was applied using the Waterfall development method. The system was built with PHP, Laravel, and MySQL as a nine-stage analytical pipeline covering text extraction, chunk segmentation, GPT-based keyword extraction, internal and external reference retrieval, chunk-level embedding, cosine similarity matching, score aggregation, and contextual novelty analysis. Functional and non-functional requirements were verified through black-box testing, output testing was conducted on 29 cross-domain proposal documents, and user perception was measured through a five-point Likert questionnaire completed by five lecturers. All requirements were declared valid, with a mean similarity score of 36.09% (SD 24.33) and a mean novelty score of 58.07% (SD 9.92). Threshold calibration revealed that the conventional cosine value of 0.85 produced a 0% similarity index on thematically relevant documents, therefore the threshold was recalibrated to 0.60. User evaluation reached an overall score of 91.2%, categorized as Very Good.

Kata Kunci:

GPT API;
Natural Language Processing;
Similarity Detection;
Novelty Detection;
Large Language Model;
Semantic Similarity;
Integrritas Akademik.

ABSTRAK

Digitalisasi dokumen akademik meningkatkan potensi kemiripan ide dan praktik plagiarisme yang tidak hanya berupa penyalinan langsung, tetapi juga parafrase dan penggantian sinonim. Sistem deteksi kemiripan yang tersedia umumnya hanya menghasilkan skor kemiripan tanpa menjelaskan posisi suatu penelitian terhadap penelitian terdahulu. Penelitian ini bertujuan menganalisis implementasi dan kemampuan GPT API dalam mendeteksi similarity dan novelty teks akademik pada aplikasi SIMANTAP, serta mengidentifikasi kebutuhan sistem, perspektif pengguna, dan implikasinya terhadap integritas akademik. Penelitian menggunakan pendekatan kuantitatif dengan metode pengembangan Waterfall. Sistem dibangun menggunakan PHP, Laravel, dan MySQL dalam bentuk pipeline analisis sembilan tahap yang mencakup ekstraksi teks, segmentasi dokumen menjadi chunk, ekstraksi kata kunci berbasis GPT, pencarian dokumen pembanding internal dan eksternal, pembangkitan embedding, perhitungan cosine similarity antar-chunk, agregasi skor, hingga analisis kebaruan kontekstual. Pengujian kebutuhan dilakukan dengan black-box testing, pengujian keluaran terhadap 29 dokumen proposal lintas domain, dan evaluasi persepsi terhadap lima dosen menggunakan kuesioner skala Likert. Seluruh kebutuhan dinyatakan valid, dengan similarity score rata-rata 36,09% (SD 24,33) dan novelty score rata-rata 58,07% (SD 9,92). Kalibrasi menunjukkan ambang konvensional 0,85 menghasilkan similarity index 0% pada dokumen yang relevan secara tematik sehingga ambang dikalibrasi ke 0,60. Evaluasi pengguna memperoleh nilai 91,2% berkategori Sangat Baik.

Korespondensi Penulis *):

Raafi Syarahil Azhar
Universitas Kebangsaan Republik Indonesia
Jl. Pelajar Pejuang 45 No.8 Kota Bandung, Indonesia 40263.

Diajukan: 06-05-2026 | Direvisi: 11-07-2026 | Diterima: 18-08-2026 | Diterbitkan: Agustus 2026

1. PENDAHULUAN

Perkembangan teknologi informasi mendorong digitalisasi dokumen akademik, termasuk proposal penelitian, skripsi, tesis, dan artikel ilmiah. Repositori *digital* memudahkan mahasiswa memperoleh referensi secara cepat dan luas, namun kemudahan tersebut sekaligus meningkatkan potensi kemiripan topik, duplikasi ide, dan plagiarisme yang menjadi tantangan bagi perguruan tinggi dalam menjaga kualitas serta integritas akademik [1], [2], [3]. Plagiarisme tidak lagi terbatas pada penyalinan langsung, melainkan semakin banyak melibatkan parafrase, penggantian sinonim, dan penyusunan ulang kalimat tanpa mengubah makna dasar dokumen [1], [2].

Berbagai penelitian telah membahas deteksi kemiripan dokumen akademik. Pendekatan semantik yang mempertimbangkan hubungan makna antardokumen terbukti meningkatkan kemampuan deteksi berbagai tipe plagiarisme [4]. Kombinasi *Synonym Recognition* dengan *Winnowing* dan *Cosine Similarity* meningkatkan hasil deteksi kemiripan rata-rata sebesar 3,11% dibandingkan metode sebelumnya [1], sedangkan *text mining* yang dipadukan dengan *Cosine Similarity* digunakan untuk mengidentifikasi kesamaan dokumen skripsi secara otomatis.

Kemajuan *Natural Language Processing* (*NLP*) mengarahkan penelitian pada representasi semantik berbasis *embedding*. *Word Embedding* yang dikombinasikan dengan *Cosine Similarity* mencapai akurasi hingga 85% dalam mendeteksi kemiripan judul skripsi [5], sementara *Sentence Transformer* mampu mengenali plagiarisme tingkat ide yang tidak tertangkap metode leksikal konvensional [6]. Pendekatan *Sentence Embedding* berbasis arsitektur *Transformer* juga terbukti lebih unggul dalam mendeteksi plagiarisme parafrase dibandingkan metode leksikal tradisional [2], dan evaluasi yang lebih luas mengonfirmasi bahwa model bahasa berbasis *Transformer* dapat menghasilkan sekaligus mendeteksi plagiarisme tingkat ide dan parafrase dengan akurasi jauh lebih tinggi daripada *baseline* leksikal [7], [8]. Representasi *embedding* yang dirancang khusus untuk dokumen ilmiah bahkan memungkinkan pengukuran kemiripan berbasis aspek tertentu dari sebuah artikel [9].

Meskipun demikian, sebagian besar penelitian terdahulu berhenti pada pengukuran kemiripan atau klasifikasi plagiat sehingga keluarannya hanya berupa skor atau label [2], [4]. Dalam konteks evaluasi proposal penelitian, skor kemiripan yang tinggi tidak selalu menandakan ketiadaan kontribusi baru, dan skor rendah tidak otomatis menunjukkan kebaruan yang signifikan. Deteksi kebaruan secara inheren merupakan tugas yang lebih kompleks karena menuntut pemahaman hubungan antardokumen serta kemampuan menyimpulkan apakah informasi baru dapat diturunkan dari penelitian sebelumnya [10]. Pemanfaatan *semantic similarity* untuk menilai kebaruan ide skripsi telah dirintis, namun keluarannya masih berupa nilai kemiripan tanpa penjelasan naratif atas kontribusi penelitian [11].

Kehadiran *Large Language Model* (*LLM*) membuka peluang untuk mengatasi keterbatasan tersebut. *GPT* menunjukkan kemampuan pemahaman konteks bahasa, penalaran, dan penyelesaian berbagai tugas *NLP* melalui *in-context learning* tanpa pelatihan ulang khusus [12]. Pendekatan berbasis *prompt* memungkinkan model beradaptasi terhadap tugas analitis baru melalui instruksi yang disisipkan langsung pada masukan [13], sementara integrasi kemampuan pembelajaran mesin ke dalam aplikasi spesifik domain terbukti meningkatkan kinerja dan efisiensi analitis [14], [15]. *LLM* juga berpotensi mendukung aktivitas akademik seperti analisis dokumen, dukungan penyuntingan, dan pemahaman hubungan antarkonsep [16], [17].

Berdasarkan kondisi tersebut, penelitian ini mengimplementasikan *GPT API* pada aplikasi SIMANTAP (Sistem Manajemen Tugas Akhir dan Praktik) untuk mendukung deteksi *similarity* dan *novelty* teks akademik berbasis *NLP*. Kebaruan penelitian ini terletak pada tiga hal. Pertama, analisis dilakukan pada level *chunk* (potongan teks), bukan pada level dokumen utuh, sehingga kemiripan dapat ditelusuri hingga bagian teks tertentu. Kedua, dokumen pembanding diperoleh dari kombinasi repositori internal dan tiga sumber akademik eksternal, sehingga cakupan pembanding tidak bergantung pada satu basis data. Ketiga, nilai ambang *Cosine Similarity* tidak diadopsi begitu saja dari konvensi, melainkan dikalibrasi secara empiris terhadap data nyata, mengingat nilai *cosine* pada ruang *embedding* tidak selalu merepresentasikan kemiripan sebagaimana yang diasumsikan [18]. Dengan demikian, sistem tidak hanya menghasilkan skor kemiripan, tetapi juga penjelasan kebaruan penelitian sehingga mendukung evaluasi karya ilmiah yang lebih objektif, komprehensif, dan efisien.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

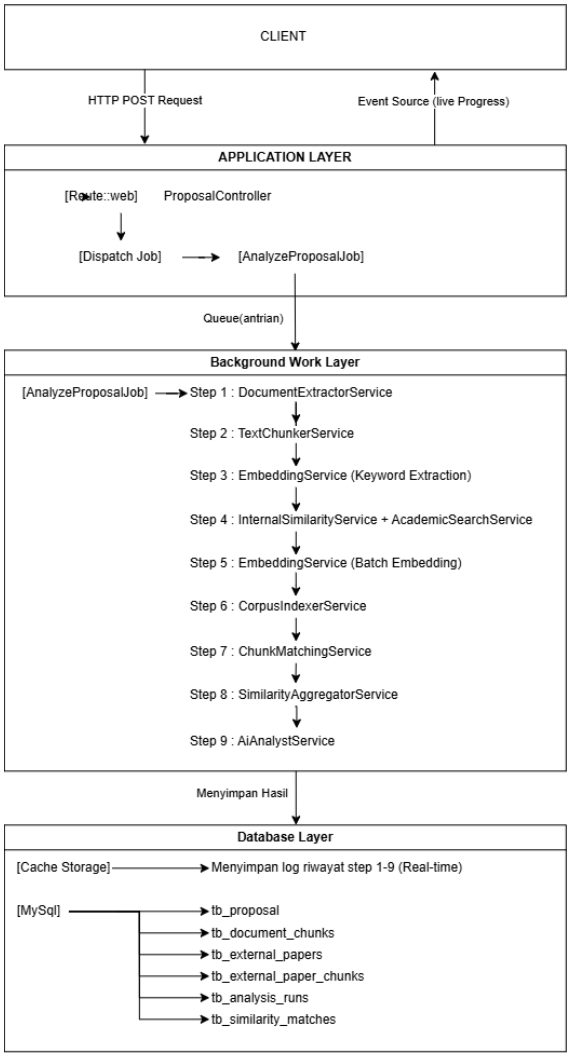
Penelitian ini menggunakan pendekatan kuantitatif dengan metode pengembangan sistem *Waterfall* yang dilaksanakan secara terstruktur dan berurutan, meliputi analisis kebutuhan, perancangan, implementasi, pengujian, dan pemeliharaan sistem. Meskipun bersifat sekuensial, penyesuaian parameter teknis, khususnya nilai ambang

kemiripan, dilakukan secara iteratif pada tahap Pemeliharaan berdasarkan hasil pengujian terhadap data nyata sebelum implementasi final.

Objek penelitian terdiri atas tiga komponen, yaitu dokumen teks akademik sebagai data uji, sistem SIMANTAP berbasis *GPT API* sebagai alat analisis, serta dosen sebagai responden penilai. Pengujian keluaran dilakukan terhadap 29 dokumen proposal skripsi lintas *domain* (*Internet of Things*, *computer vision*, *machine learning*, sistem informasi berbasis web, keamanan siber, dan *NLP*), sedangkan evaluasi persepsi melibatkan lima dosen Program Studi Teknik Informatika Universitas Kebangsaan Republik Indonesia yang dipilih menggunakan teknik *purposive sampling* dengan kriteria aktif membimbing atau menguji skripsi dalam dua tahun terakhir.

2.2. Arsitektur dan Teknologi Sistem

Arsitektur SIMANTAP terdiri atas empat lapisan. *Client Layer* berfungsi sebagai antarmuka pengguna untuk mengirim permintaan analisis dan menerima progres secara *real-time* melalui *Server-Sent Events* (*SSE*). *Application Layer* menerima permintaan melalui *controller* dan mendelegasikan proses ke *background job* melalui mekanisme *queue*. *Background Work Layer* merupakan lapisan pemrosesan inti yang terdiri atas DocumentExtractorService, TextChunkerService, InternalSimilarityService, AcademicSearchService, EmbeddingService, CorpusIndexerService, ChunkMatchingService, SimilarityAggregatorService, dan AiAnalystService. *Database Layer* menyimpan data proposal, *chunk* dokumen, dokumen pembanding beserta *embedding*-nya, riwayat analisis, dan pasangan *chunk* yang tercatat sebagai kecocokan.



Gambar 1. Arsitektur Sistem SIMANTAP

Sistem dibangun menggunakan bahasa pemrograman *PHP* dengan *framework Laravel* berpola *Model-View-Controller*, serta *MySQL* sebagai sistem manajemen basis data. Konfigurasi *GPT API* tidak disimpan secara statis pada berkas *environment*, melainkan pada basis data sehingga *administrator* dapat mengganti *API key* dan model *chat* melalui halaman pengaturan tanpa mengubah kode program maupun melakukan *deployment* ulang. *API key* disimpan dalam bentuk terenkripsi, sedangkan nama model divalidasi terhadap daftar model yang diizinkan dan dialihkan ke model *fallback* apabila nilainya tidak dikenali. Berbeda dengan model *chat*, model *embedding* bersifat tetap dan menghasilkan vektor berdimensi 1536, karena perubahan dimensi akan merusak konsistensi vektor yang telah tersimpan.

2.3. Alur Analitik dan Perhitungan Skor

Proses analisis berjalan sebagai *pipeline* sembilan tahap: (1) ekstraksi teks dari berkas *PDF*, *DOCX*, atau *DOC*; (2) segmentasi teks menjadi *chunk* berukuran sekitar 250 kata dengan *overlap* 40 kata antar-*chunk*; (3) ekstraksi kata kunci melalui *GPT API*; (4) pencarian dokumen pembanding dari proposal internal SIMANTAP dan tiga sumber akademik eksternal, yaitu Semantic Scholar, OpenAlex, dan CrossRef, yang hasilnya digabungkan dan dideduplikasi; (5) pembangkitan *embedding* untuk setiap *chunk*; (6) pengindeksan permanen *embedding* dokumen pembanding agar tidak dihitung ulang; (7) pencocokan antar-*chunk* menggunakan *Cosine Similarity*; (8) agregasi skor menjadi *similarity index*; dan (9) analisis *novelty* berbasis *GPT API*.

Nilai *Cosine Similarity* antara vektor *chunk* A dan B dihitung menggunakan persamaan berikut:

$$(x+a)^n = \sum_{k=0}^{n} ? \left(\frac{n}{k}\right) x^k a^{n-k} \; Similarity(A,B) \; = \; \frac{A \,.\, B}{||A||\,||B||} \quad \text{..... (1)}$$

Nilai *cosine* yang mendekati 1 menunjukkan kemiripan semantik tinggi, sedangkan nilai mendekati 0 menunjukkan tidak adanya hubungan makna yang berarti. Pasangan *chunk* dengan nilai kemiripan melewati ambang batas ditandai sebagai kecocokan, kemudian *similarity index* dihitung berdasarkan proporsi jumlah kata pada *chunk* yang cocok terhadap total kata dokumen:

$$(x+a)^n = \sum_{k=0}^{n} ? \left(\frac{n}{k}\right) x^k a^{n-k} \; Similarity\ Index = \left(\frac{\sum W\ Cocok}{W\ total}\right) \times 100\% \quad \text{.....(2)}$$

Skor tersebut dipecah menjadi *internal index* dan *external index* sesuai kategori sumber dokumen pembanding, kemudian diteruskan ke *GPT API* bersama daftar dokumen pembanding untuk menghasilkan skor *novelty* beserta penjelasan naratifnya.

2.4. Prosedur Pengujian dan Teknik Analisis Data

Pengujian dilakukan dalam empat bagian. Pertama, pengujian kebutuhan fungsional (FR-01 hingga FR-08) dan non-fungsional (NR-01 hingga NR-05) menggunakan metode *black-box testing* melalui *endpoint pipeline* yang bersesuaian, sehingga keterunutan antara kebutuhan dan bukti pengujian dapat ditelusuri. Kedua, pengujian kalibrasi ambang dilakukan dengan menjalankan analisis berulang pada dokumen yang sama menggunakan lima nilai ambang berbeda, lalu mengamati perubahan *similarity index* dan *novelty score*. Ketiga, pengujian keluaran pada 29 dokumen uji dianalisis menggunakan statistik deskriptif berupa rata-rata, nilai minimum, nilai maksimum, dan standar deviasi, serta dikelompokkan ke dalam empat kategori distribusi, yaitu Rendah (0%–<25%), Sedang (25%–<50%), Tinggi (50%–<75%), dan Sangat Tinggi (≥75%). Keempat, evaluasi persepsi pengguna dilakukan melalui kuesioner sepuluh butir pernyataan berskala *Likert* lima poin yang dikelompokkan ke dalam empat kriteria, yaitu kejelasan kebutuhan sistem, kemudahan menginterpretasikan keluaran, persepsi kebermanfaatan, dan persepsi kontribusi terhadap integritas akademik.

Data kuesioner dianalisis menggunakan metode persentase skor ideal. Skor aktual setiap butir dihitung dengan menjumlahkan hasil kali bobot skala dan jumlah responden yang memilih skala tersebut:

$$(x+a)^n = \sum_{k=0}^{n} ? \left(\frac{n}{k}\right) x^k a^{n-k} \mathrm{X} = \Sigma\ (\mathrm{Nj} \times \mathrm{Rj}) \quad \text{.....(3)}$$

Nilai persentase tiap butir diperoleh dengan membandingkan skor aktual terhadap skor ideal, yaitu perkalian skala tertinggi dengan jumlah responden:

$$(x+a)^n = \sum_{k=0}^{n} ? \left(\frac{n}{k}\right) x^k a^{n-k} \mathrm{Y} = \left(\frac{X}{Skor\ Ideal}\right) \times 100\% \; ; \quad \text{Skor Ideal} \; = \; \mathrm{S} \times \mathrm{r} \quad \text{.....(4)}$$

Nilai persentase tiap kriteria kemudian dikonversi kembali ke rata-rata skor *Likert* agar dapat diinterpretasikan pada skala pengukuran asli:

$$(x+a)^n = \sum_{k=0}^{n} ? \left(\frac{n}{k}\right) x^k a^{n-k} \bar{\mathrm{x}}\mathrm{k} = \left(\frac{Yk}{100}\right) \times S \quad \text{.....(5)}$$

Panjang interval kelas ditentukan sebesar 0,8 sehingga diperoleh lima kategori interpretasi sebagaimana disajikan pada Tabel 1

Tabel 1. Interval interpretasi penilaian responden

***Interval* Persentase**	***Interval* Skor *Likert***	**Kategori**
0% – 20%	1,00 – 1,80	Sangat Kurang
21% – 40%	1,81 – 2,60	Kurang
41% – 60%	2,61 – 3,40	Cukup
61% – 80%	3,41 – 4,20	Baik
81% – 100%	4,21 – 5,00	Sangat Baik

3. HASIL DAN ANALISIS

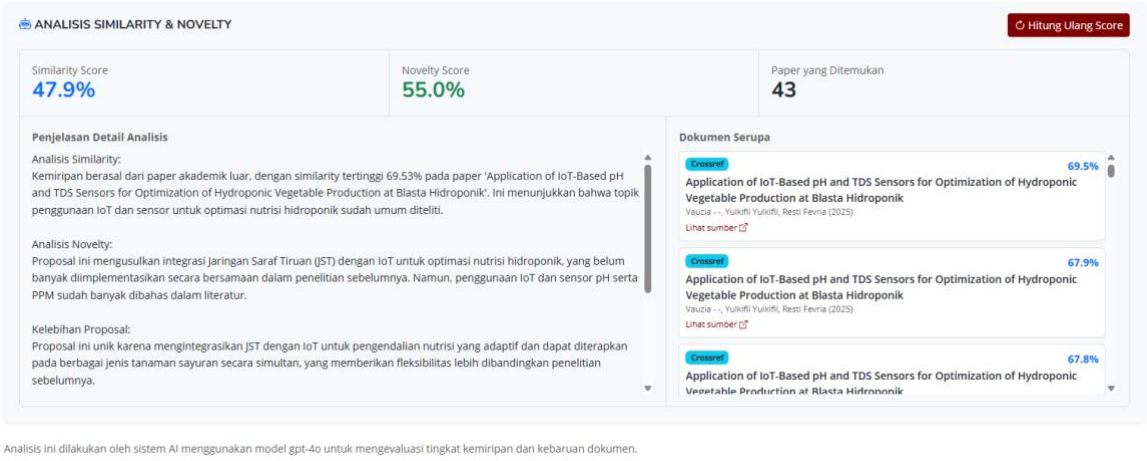
3.1 Hasil Implementasi dan Pengujian Kebutuhan

Seluruh kebutuhan fungsional dan non-fungsional yang ditetapkan pada tahap analisis berhasil diimplementasikan dan dinyatakan valid pada pengujian *black-box*. Ringkasan hasil pengujian disajikan pada Tabel 2.

Tabel 2. Hasil pengujian kebutuhan fungsional dan non-fungsional

Kode	Kebutuhan yang Diuji	Hasil yang Diharapkan	Status
FR-01	Ekstraksi teks dokumen	Teks terekstraksi dari format PDF, DOCX, dan DOC	Valid
FR-02	*Chunking* dokumen	Dokumen terpecah menjadi *chunk* sebagai satuan analisis	Valid
FR-03	Ekstraksi kata kunci	Kata kunci utama dokumen diperoleh melalui *GPT API*	Valid
FR-04	Pencarian dokumen pembanding	Daftar pembanding unik gabungan internal dan eksternal diperoleh	Valid
FR-05	Pembuatan dan penyimpanan *embedding*	*Embedding* tersimpan permanen dan tidak dihitung ulang	Valid
FR-06	Perhitungan *similarity* antar-*chunk*	*Similarity index* internal dan eksternal terhitung terpisah	Valid
FR-07	Analisis *novelty*	Skor dan penjelasan naratif kebaruan dihasilkan	Valid
FR-08	Penyimpanan riwayat analisis	Hasil analisis dan riwayatnya tersimpan	Valid
NR-01	Keamanan (*security*)	*API key* tersimpan terenkripsi dan tidak tampil sebagai teks biasa	Valid
NR-02	Keandalan (*reliability*)	Mekanisme *retry* dan *fallback* menjaga keluaran tetap dihasilkan	Valid
NR-03	Performa (*performance*)	Antarmuka tetap responsif karena analisis berjalan asinkron	Valid
NR-04	Kemudahan penggunaan (*usability*)	Progres tampil *real-time* dan hasil mudah dipahami	Valid
NR-05	Pemeliharaan (*maintainability*)	Konfigurasi model dapat diubah tanpa perubahan kode	Valid

Keluaran akhir sistem ditampilkan dalam bentuk ringkasan skor beserta penjelasan naratif dan daftar dokumen serupa, sebagaimana ditunjukkan pada Gambar 2. Pada contoh tersebut, sistem menghasilkan *similarity score* 47,9% dan *novelty score* 55,0% dari 43 dokumen pembanding yang ditemukan, dengan kecocokan tertinggi 69,5% terhadap artikel eksternal mengenai sensor pH dan TDS berbasis *IoT* untuk hidroponik. Penjelasan naratif yang menyertainya menyatakan bahwa integrasi Jaringan Saraf Tiruan dengan sistem *IoT* untuk optimasi nutrisi merupakan kontribusi yang relatif baru, meskipun pemanfaatan *IoT* dan sensor pada hidroponik telah banyak diteliti



Gambar 2. Antarmuka keluaran analisis *similarity* dan *novelty*

3.2 Kalibrasi Ambang *Cosine Similarity*

Pengujian kalibrasi dilakukan pada satu proposal yang sama dengan lima nilai ambang berbeda. Hasilnya disajikan pada Tabel 3.

Tabel 3. Hasil pengujian kalibrasi ambang *Cosine Similarity*

Ambang	*Similarity Index*	*Internal Index*	*External Index*	*Novelty Score*	Observasi
0,85	0%	0%	0%	85	Tidak ada *chunk* lolos ambang; dua artikel relevan (74,07% dan 70,48%) tidak terdeteksi
0,75	0%	0%	0%	85	Masih belum ada kecocokan meskipun pembanding relevan sudah ditemukan
0,70	8,61%	0%	8,61%	65	Kecocokan mulai terdeteksi pada tingkat wajar
0,65	30,26%	0%	30,26%	55	Cakupan kecocokan meningkat signifikan
0,60	36,94%	0%	36,94%	55	Ambang final; paling representatif terhadap kemiripan riil topik

Pada ambang 0,85 dan 0,75 sistem tidak mendeteksi kemiripan sama sekali meskipun terdapat dua artikel eksternal dengan kemiripan tematik tinggi pada level *chunk*, sehingga sistem justru menghasilkan *novelty score* 85 yang menyesatkan karena mencerminkan kegagalan deteksi, bukan kebaruan penelitian. Penurunan ambang secara bertahap meningkatkan *similarity index* secara konsisten hingga 36,94% pada ambang 0,60, di mana sistem berhasil mengenali artikel *Smart Nutrient Management in Hydroponics* dengan kecocokan 71,2%. Temuan ini menunjukkan bahwa ambang konvensional yang terlalu ketat tidak sesuai dengan karakteristik distribusi nilai *cosine* pada teks akademik berbahasa Indonesia, sekaligus memperkuat argumen bahwa nilai *cosine* pada ruang *embedding* perlu diinterpretasikan secara hati-hati dan dikalibrasi terhadap data [18].

3.3 Hasil Pengujian pada *Dataset* Uji

Pengujian keluaran dilakukan terhadap 29 dokumen proposal skripsi lintas domain. Hasil pengukuran untuk setiap dokumen disajikan pada Tabel 4.

Tabel 4. Hasil pengujian *similarity* dan *novelty* pada 29 dokumen uji

No.	Judul/Topik Proposal	*Similarity*	*Novelty*	Dokumen Serupa
1	Forensik digital pada smart lock berbasis IoT (NIST)	48,1%	58,0%	10
2	Deteksi asap dan gas berbahaya berbasis Arduino	75,0%	55,0%	92
3	Keamanan infrastruktur microcloud LXD dengan HIDS/IPS	43,5%	65,0%	12
4	Monitoring dan controlling suhu kandang ayam berbasis IoT	25,1%	45,0%	45
5	Analisis kesejahteraan masyarakat dengan ANN	0,0%	85,0%	0
6	Prototipe irigasi sprinkler otomatis berbasis IoT	72,8%	55,0%	125
7	Lengan robot pemilah telur berbasis computer vision	9,8%	55,0%	7
8	Brankas pintar dengan password dan alarm berbasis Arduino	48,4%	55,0%	9
9	Deteksi makanan dan estimasi gizi (YOLOv8 dan LLM)	37,8%	55,0%	7
10	Deteksi kelelahan pengemudi berbasis IoT dan fuzzy logic	47,2%	55,0%	5
11	Deteksi burung hama berbasis suara dan ESP32-CAM (YOLOv5)	56,9%	65,0%	27
12	Sistem informasi penjualan rumput pakan ternak berbasis web	4,4%	45,0%	1
13	Analisis kinerja website dengan metode PIECES	58,5%	65,0%	19
14	Pemantauan aktivitas dan kesehatan hewan peliharaan (IoT)	0,0%	85,0%	0
15	Sistem pelayanan desa berbasis web dengan QR code dan chatbot	71,8%	65,0%	7
16	Deteksi jenis makanan dengan YOLOv8 untuk pencegahan obesitas	48,3%	55,0%	10
17	K-Means pada arsitektur Django untuk zonasi performa politik	1,0%	65,0%	1
18	Klasterisasi ketahanan ekonomi desa dengan K-Means	20,5%	55,0%	33
19	Klasifikasi kematangan buah tomat berbasis citra (CNN)	91,3%	65,0%	77
20	Klasifikasi kehadiran siswa dengan LSTM	21,0%	45,0%	5
21	Klasifikasi kualitas telur ayam konsumsi dengan CNN	17,4%	65,0%	5
22	Sistem booking online pasien berbasis web	33,2%	55,0%	14

No.	Judul/Topik Proposal	*Similarity*	*Novelty*	Dokumen Serupa
23	Implementasi GPT API untuk deteksi similarity dan novelty	18,9%	55,0%	12
24	Klasifikasi jenis sampah rumah tangga dengan CNN	32,4%	55,0%	43
25	JST untuk optimasi nutrisi hidroponik adaptif berbasis IoT	55,8%	58,0%	169
26	Monitoring mesin ribbon mixer produksi baglog jamur (IoT)	26,1%	45,0%	5
27	Caesar cipher untuk pengamanan data resep pasien apotek	29,2%	45,0%	6
28	Deteksi jenis makanan dengan Faster R-CNN untuk diabetes	38,4%	58,0%	4
29	Estimasi kebutuhan pupuk nitrogen padi (YOLOv5, BWD)	13,9%	55,0%	30

Berdasarkan hasil tersebut, diperoleh *similarity score* rata-rata 36,09% (minimum 0,0%; maksimum 91,3%; standar deviasi 24,33) dan *novelty score* rata-rata 58,07% (minimum 45,0%; maksimum 85,0%; standar deviasi 9,92). Distribusi kategori skor kemiripan disajikan pada Tabel 5.

Tabel 5. Distribusi kategori *similarity score* pada 29 dokumen uji

Kategori	Rentang	Jumlah Dokumen	Persentase
Rendah	0% – <25%	10	34,5%
Sedang	25% – <50%	12	41,4%
Tinggi	50% – <75%	5	17,2%
Sangat Tinggi	≥75%	2	6,9%

Sebanyak 75,9% dokumen berada pada kategori Rendah hingga Sedang, sesuai dengan keragaman domain penelitian pada *dataset* uji. Skor tertinggi sebesar 91,3% diperoleh proposal klasifikasi tingkat kematangan buah tomat berbasis *CNN* dan konsisten dengan jumlah dokumen serupa yang ditemukan (77 dokumen), mengindikasikan topik tersebut merupakan area yang telah jenuh diteliti. Sebaliknya, dua proposal mencatat *similarity score* 0,0% tanpa dokumen serupa sekaligus *novelty score* tertinggi 85,0%, yang menunjukkan bahwa topik tersebut belum banyak terwakili dalam korpus pembanding yang tersedia.

3.4 Hasil Evaluasi Persepsi Pengguna

Kuesioner sepuluh butir disebarkan kepada lima dosen sehingga skor ideal setiap butir sebesar 25. Nilai persentase tiap butir disajikan pada Tabel 6.

Tabel 6. Nilai persentase tiap butir kuesioner

Kode	Butir Pernyataan	Nilai Persentase (Y)
A1	Sistem dibutuhkan untuk mengatasi keterbatasan evaluasi manual	88,0%
A2	Fitur SIMANTAP sesuai kebutuhan evaluasi proposal	88,0%
A3	Analisis *novelty* merupakan aspek penting dalam evaluasi	96,0%
B1	Skor *similarity* dan *novelty* mudah dipahami maknanya	92,0%
B2	Penjelasan naratif *GPT* membantu memahami alasan skor	88,0%
B3	Daftar dokumen serupa membantu menelusuri sumber kemiripan	88,0%
C1	Sistem mempercepat proses evaluasi dibandingkan cara manual	96,0%
C2	Hasil analisis membantu keputusan yang lebih objektif	92,0%
D1	Sistem berpotensi mencegah plagiarisme dan duplikasi ide	88,0%
D2	Penerapan berkelanjutan mendukung integritas akademik	96,0%
	Rata-rata keseluruhan	91,2%

Nilai rata-rata persentase tiap kriteria beserta konversinya ke rata-rata skor *Likert* disajikan pada Tabel 7.

Tabel 7. Nilai rata-rata persentase dan skor Likert tiap kriteria

Kriteria	Rata-rata Persentase	Rata-rata *Likert*	Kategori
A. Kejelasan kebutuhan sistem	90,7%	4,53	Sangat Baik

Kriteria	Rata-rata Persentase	Rata-rata *Likert*	Kategori
B. Kemudahan menginterpretasikan keluaran	89,3%	4,47	Sangat Baik
C. Persepsi kebermanfaatan	94,0%	4,70	Sangat Baik
D. Kontribusi terhadap integritas akademik	92,0%	4,60	Sangat Baik

Nilai rata-rata keseluruhan sebesar 91,2% menempatkan seluruh kriteria, bahkan seluruh butir pernyataan tanpa terkecuali, pada kategori Sangat Baik. Nilai tertinggi diperoleh kriteria persepsi kebermanfaatan (94,0%), sedangkan nilai terendah diperoleh kriteria kemudahan menginterpretasikan keluaran (89,3%).

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa pendekatan berbasis *GPT API* melampaui kemampuan perangkat deteksi kemiripan konvensional. Jika pendekatan leksikal hanya mampu mengidentifikasi kesamaan kata [4], representasi *embedding* memungkinkan sistem mengenali kesamaan gagasan meskipun diungkapkan dengan struktur kalimat berbeda, konsisten dengan temuan mengenai deteksi plagiarisme tingkat ide berbasis *Transformer* [6], [7], [8]. Analisis pada level *chunk* memberi nilai tambah karena kemiripan dapat ditelusuri hingga bagian teks tertentu, bukan sekadar sebagai satu angka agregat untuk keseluruhan dokumen [9].

Temuan paling menentukan terletak pada kalibrasi ambang. Nilai 0,85 yang lazim dipakai sebagai konvensi justru menghasilkan *similarity index* 0% pada dokumen yang secara substansi memiliki kemiripan, karena nilai kemiripan mentah tertinggi yang teramati hanya berkisar pada rentang 0,59–0,66. Hal ini menegaskan bahwa nilai *cosine* antar-*embedding* tidak memiliki makna absolut yang dapat dipindahkan begitu saja antar-model maupun antar-domain, sehingga penetapan ambang harus diperlakukan sebagai keputusan empiris yang dikalibrasi terhadap data, bukan sebagai konstanta yang diwarisi dari literatur.

Perbedaan mencolok antara standar deviasi *similarity* (24,33) dan *novelty* (9,92) membuktikan secara empiris bahwa kebaruan tidak berperilaku sebagai kebalikan linear dari kemiripan. Sistem tetap mempertimbangkan konteks penerapan dan kombinasi metode ketika menilai kebaruan, sehingga proposal dengan kemiripan tinggi tidak otomatis dinilai tidak baru. Temuan ini memperkuat argumen bahwa evaluasi karya ilmiah tidak cukup bersandar pada skor kemiripan semata [10], [11].

Dari perspektif pengguna, aspek kebermanfaatan memperoleh nilai tertinggi, yang menunjukkan bahwa penjelasan kualitatif hasil *GPT* memberi nilai praktis melebihi skor numerik dan sejalan dengan penelitian terdahulu mengenai peran penjelasan kontekstual dalam meningkatkan kepercayaan pengguna terhadap perangkat evaluasi berbasis *AI* [16], [17]. Sebaliknya, nilai terendah pada aspek kemudahan menginterpretasikan keluaran mengindikasikan bahwa penyajian hasil masih perlu disederhanakan bagi pengguna non-teknis. Kebutuhan *maintainability* dan *reliability* juga terbukti bukan sekadar pelengkap, melainkan penentu kelayakan sistem, karena keduanya memungkinkan penggantian model dan kalibrasi ambang dilakukan secara iteratif tanpa perubahan kode sekaligus menjaga keberlangsungan analisis ketika layanan pihak ketiga gagal merespons.

Implikasi terhadap integritas akademik bersifat dua arah. Di satu sisi, ssistem memperluas jangkauan deteksi ke ranah semantik sehingga parafrase dan kesamaan gagasan dapat teridentifikasi, sekaligus menggeser fokus pembimbingan dari besaran persentase kemiripan menuju substansi kontribusi penelitian. Di sisi lain terdapat risiko ketergantungan berlebih, potensi halusinasi *Large Language Model* pada narasi kebaruan, serta sensitivitas hasil terhadap nilai ambang yang dipilih. Penelitian ini juga memiliki keterbatasan, yaitu belum tersedianya penilaian pakar independen sebagai *ground truth* pembanding, jumlah responden yang terbatas pada satu program studi sehingga hasilnya bersifat deskriptif dan tidak dimaksudkan untuk generalisasi, serta belum diujinya ketahanan sistem terhadap parafrase *adversarial* dan teks akademik multibahasa.

4. KESIMPULAN

Implementasi *GPT API* pada aplikasi SIMANTAP berhasil diwujudkan sebagai *pipeline* analisis sembilan tahap yang mencakup ekstraksi teks, segmentasi *chunk*, ekstraksi kata kunci, pencarian dokumen pembanding internal dan eksternal, pembangkitan *embedding*, pencocokan antar-*chunk*, agregasi skor, hingga analisis kebaruan kontekstual. Seluruh kebutuhan fungsional (FR-01 sampai FR-08) dan non-fungsional (NR-01 sampai NR-05) dinyatakan valid pada pengujian *black-box*. Pengujian terhadap 29 dokumen proposal lintas domain menghasilkan *similarity score* rata-rata 36,09% (standar deviasi 24,33) dan *novelty score* rata-rata 58,07% (standar deviasi 9,92), dengan distribusi kategori Rendah 34,5%, Sedang 41,4%, Tinggi 17,2%, dan Sangat Tinggi 6,9%.

Pengujian kalibrasi membuktikan bahwa ambang *Cosine Similarity* sebesar 0,85 yang lazim digunakan sebagai konvensi justru menghasilkan *similarity index* 0% pada dokumen yang secara tematik jelas relevan, sehingga ambang dikalibrasi ke 0,60 sebagai nilai paling representatif terhadap tingkat kemiripan riil dokumen akademik berbahasa Indonesia. Evaluasi terhadap lima dosen memperoleh nilai rata-rata keseluruhan 91,2% dengan seluruh kriteria berkategori Sangat Baik, yaitu kejelasan kebutuhan sistem 90,7% (4,53), kemudahan menginterpretasikan keluaran 89,3% (4,47), persepsi kebermanfaatan 94,0% (4,70), dan kontribusi terhadap integritas akademik 92,0% (4,60).

Penelitian ini menyimpulkan bahwa *GPT API* mampu menghasilkan analisis kemiripan sekaligus penjelasan kebaruan yang lebih kontekstual dibandingkan pendekatan berbasis skor semata, namun sistem tetap harus diposisikan sebagai alat bantu pengambilan keputusan, bukan penentu putusan akhir atas orisinalitas karya ilmiah. Penelitian selanjutnya disarankan melibatkan penilaian pakar independen sebagai *ground truth*, menguji stabilitas nilai ambang pada korpus yang lebih luas dan lintas bahasa maupun lintas model *embedding*, memperbaiki penyajian hasil bagi pengguna non-teknis, memperluas jumlah dan asal responden, serta mengkaji penggunaan model sumber terbuka yang dijalankan secara mandiri untuk mengurangi ketergantungan pada layanan pihak ketiga

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika, Universitas Kebangsaan Republik Indonesia, serta seluruh pihak yang telah memberikan dukungan sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

[1] S. Purwaningrum, A. Susanto, and A. Kristiningsih, “Pengaruh Synonym Recognition dalam Deteksi Kemiripan Teks Menggunakan Winnowing dan Cosine Similarity”, doi: 10.22146/jnteti.v12i3.6375.

[2] N. Tamim Syuja’i, D. Imanuel Daeli, I. Alfarizi Darnela, E. A. Thio Candra, and J. Putra Tanjung, “Deteksi Plagiarisme Tugas Mahasiswa Menggunakan Sentence Embedding Berbasis Transformer dan Metode Cosine Similarity,” *Jurnal Algoritma*, vol. 23, no. 1, May 2026, doi: 10.33364/algoritma/v.23-1.3523.

[3] S. Pudasaini, L. Miralles-Pechuán, D. Lillis, and M. L. Salvador, “Survey on Plagiarism Detection in Large Language Models: The Impact of ChatGPT and Gemini on Academic Integrity,” Jun. 2024, [Online]. Available: http://arxiv.org/abs/2407.13105

[4] S. Agustian, “PENDEKATAN SEMANTIK DALAM DETEKSI BERBAGAI TIPE PLAGIARISME PADA DOKUMEN TEKS,” *JURNAL TEKNIK INFORMATIKA*, vol. 14, no. 2, pp. 101–114, Oct. 2021, doi: 10.15408/jti.v14i2.22841.

[5] Salman, “SISTEM PENDETEKSI KEMIRIPAN JUDUL SKRIPSI BERBASIS WEB MENGGUNAKAN NLP DAN WORD EMBEDDINGS”, doi: 10.59818/jpi.v5i5.1964.

[6] S. L. Rahma and U. Taufiq, “Analisis Tingkat Akurasi Metode Pendeteksian Plagiarisme Ide dengan Menggunakan Yake dan Sentence Transformer,” *Journal of Internet and Software Engineering*, vol. 5, no. 1, 2024, doi: https://doi.org/10.22146/jise.v5i1.9073.

[7] J. P. Wahle, T. Ruas, F. Kirstein, and B. Gipp, “How Large Language Models are Transforming Machine-Paraphrased Plagiarism,” Nov. 2022, doi: 10.18653/v1/2022.emnlp-main.62.

[8] J. Lee, T. Le, J. Chen, and D. Lee, “Do Language Models Plagiarize?,” in *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, Association for Computing Machinery, Inc, Apr. 2023, pp. 3637–3647. doi: 10.1145/3543507.3583199.

[9] M. Ostendorff, T. Blume, T. Ruas, B. Gipp, and G. Rehm, “Specialized document embeddings for aspect-based similarity of research papers,” in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, Institute of Electrical and Electronics Engineers Inc., Jun. 2022. doi: 10.1145/3529372.3530912.

[10] T. Ghosal, T. Saikh, T. Biswas, A. Ekbal, and P. Bhattacharyya, “Novelty Detection: A Perspective from Natural Language Processing under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license,” *Computational Linguistics*, vol. 48, no. 1, 2022, doi: 10.1162/COLI.

[11] D. Wardani and C. Achmad, “SSTI: Semantic Similarity to detect Novelty of Thesis Ideas,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2022, pp. 376–381. doi: 10.1145/3575882.3575955.

[12] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” Jul. 2020, doi: 10.48550/arXiv.2005.14165.

[13] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing,” *ACM Comput. Surv.*, vol. 55, no. 9, Sep. 2023, doi: 10.1145/3560815.

[14] S. A. Atmaja, A. W. R. Emanuel, and Suyoto, “A Systematic Literature Review on Intelligent Optimization for Virtual Machine Allocation in Cloud Data Centers,” in *ICCSIT 2025 - Proceedings of the 2025 18th International Conference on Computer Science and Information Technology*, Association for Computing Machinery, Inc, Mar. 2026, pp. 58–68. doi: 10.1145/3783862.3783871.

[15] J. Budiasto, “Advancements in Machine Learning Algorithms for Big Data Analytics,” *Journal of Hunan University Natural Sciences*, Mar. 2026, doi: 10.55463/issn.1674-2974.53.2.5.

[16] E. Kasneci *et al.*, “ChatGPT for good? On opportunities and challenges of large language models for education,” *Learn. Individ. Differ.*, vol. 103, p. 102274, Apr. 2023, doi: 10.1016/J.LINDIF.2023.102274.

[17] W. Castillo-González, C. O. Lepez, and M. C. Bonardi, “Chat GPT: a promising tool for academic editing,” *Data and Metadata*, vol. 1, Oct. 2022, doi: 10.56294/dm202223.

[18] H. Steck, C. Ekanadham, and N. Kallus, “Is Cosine-Similarity of Embeddings Really About Similarity?,” in *WWW 2024 Companion - Companion Proceedings of the ACM Web Conference*, Association for Computing Machinery, Inc, May 2024, pp. 887–890. doi: 10.1145/3589335.3651526.