



**PEMODELAN *PIPELINE ADVANCED RAG* MENGGUNAKAN
HYBRID SEARCH (BM25-INDOBERT) DAN *RERANKING*
(*CROSS-ENCODER*) UNTUK LAYANAN INFORMASI
AKADEMIK**

**TUGAS AKHIR
SKRIPSI**
HALAMAN SAMBUT

**GEDE JOVAN TIYO ADITYA
41522010019**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**PEMODELAN *PIPELINE ADVANCED RAG* MENGGUNAKAN
HYBRID SEARCH (BM25-INDOBERT) DAN *RERANKING*
(*CROSS-ENCODER*) UNTUK LAYANAN INFORMASI
AKADEMIK**

**TUGAS AKHIR
SKRIPSI**

HALAMAN JUDUL

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

GEDE JOVAN TIYO ADITYA

41522010019

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Gede Jovan Tiyo Aditya

NIM : 41522010019

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Pemodelan *Pipeline Advanced RAG* Menggunakan *Hybrid Search* (BM25-IndoBERT) dan *Reranking (Cross-Encoder)* untuk Layanan Informasi Akademik” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

Jakarta, 27 Juli 2026



Gede Jovan Tiyo Aditya

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
The undersigned, hereby declare that the scientific work written by

Nama /Name : Gede Jovan Tiyo Aditya
NIM /Student ID Number : 41522010019
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

The title:

“Pemodelan Pipeline Advanced RAG Menggunakan Hybrid Search (BM25-IndoBERT) dan Reranking (Cross-Encoder) untuk Layanan Informasi Akademik”

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

29 Juni 2026

dengan nilai persentase sebesar :

/with a percentage value of:

5%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1-wKYQYGH2z1Slak1wLoASdzrsOV4NZo0/view?usp=drive_link



Jakarta, 29 Juni 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto

Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Gede Jovan Tiyo Aditya
NIM : 41522010019
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : Pemodelan *Pipeline Advanced* RAG Menggunakan *Hybrid Search* (BM25-IndoBERT) dan *Reranking* (*Cross-Encoder*) untuk Layanan Informasi Akademik

Telah berhasil dipertahankan pada sidang tanggal 27 Juli 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing



Wawan Gunawan, S.Kom., M.T., M.Kom.

NIDN/NUPTK: 0424108104

Jakarta, 27 Juli 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002

Ketua Program Studi
Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, atas segala berkat, rahmat, dan karunia-Nya yang senantiasa menyertai setiap langkah penulis, sehingga Tugas Akhir ini dapat diselesaikan dengan baik. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Informatika pada Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari sepenuhnya bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak, sejak masa perkuliahan hingga tahap akhir penyusunan Tugas Akhir ini perjalanan ini tidak akan dapat terlaksana dengan baik. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana, atas kepemimpinan dan dedikasi beliau dalam membangun lingkungan akademik yang kondusif bagi seluruh civitas akademika.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer, atas arahan dan kebijakan yang telah mendukung perkembangan akademik mahasiswa di lingkungan fakultas.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana, atas bimbingan dan perhatiannya terhadap perkembangan program studi serta kemajuan para mahasiswanya.
4. Bapak Wawan Gunawan, S.Kom., M.T., M.Kom. selaku dosen pembimbing Tugas Akhir, yang dengan penuh kesabaran telah meluangkan waktu, tenaga, dan pikiran untuk memberikan pengarahan, masukan, serta motivasi yang sangat berarti bagi penulis selama proses penyusunan Tugas Akhir ini.
5. Kedua orang tua penulis yang tiada henti memberikan doa, kasih sayang, semangat, dan dukungan baik secara moral maupun material selama penulis menjalani masa studi di Universitas Mercu Buana.
6. Mimi Lavenia Sinaga dan Putri Wulan Ramadhan, selaku rekan-rekan yang telah menemani dan memberikan dukungan nyata selama proses pengerjaan Tugas Akhir ini, baik dalam diskusi, kolaborasi, maupun semangat yang selalu diberikan di setiap kesempatan.
7. Ni Kadek Dewi, Sulfia, dan Evi Wahyuni Pulungan, serta seluruh sahabat dan teman-teman penulis yang tidak dapat disebutkan satu per satu, yang telah hadir sebagai tempat berbagi cerita, memberikan semangat di saat-saat sulit,

serta menemani perjalanan penulis selama masa kuliah hingga penyelesaian Tugas Akhir ini.

Akhir kata, penulis berharap kiranya Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu dengan balasan yang setimpal. Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna, namun penulis berharap karya ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang Teknik Informatika, serta menjadi kontribusi yang berarti bagi siapa pun yang membutuhkannya.

Jakarta, 27 Juli 2026

Gede Jovan Tiyo Aditya



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR DI REPOSITORI UMB**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Gede Jovan Tiyo Aditya
NIM : 41522010019
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Pemodelan *Pipeline Advanced RAG*
Menggunakan *Hybrid Search* (BM25-
IndoBERT) dan *Reranking* (*Cross-Encoder*)
untuk Layanan Informasi Akademik

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 27 Juli 2026

Yang menyatakan,



Gede Jovan Tiyo Aditya

**PEMODELAN PIPELINE ADVANCED RAG MENGGUNAKAN HYBRID
SEARCH (BM25-INDOBERT) DAN RERANKING (CROSS-ENCODER)
UNTUK LAYANAN INFORMASI AKADEMIK**

GEDE JOVAN TIYO ADITYA

ABSTRAK

Penelitian ini mengembangkan dan mengevaluasi sistem layanan informasi akademik berbasis *Large Language Model* (LLM) di Fakultas Ilmu Komputer. Permasalahan utama yang diangkat adalah keterbatasan layanan informasi manual oleh staf administrasi serta kelemahan bawaan LLM berupa halusinasi dan batasan pembaruan pengetahuan (*knowledge cutoff*). Untuk mengatasi hal tersebut, penelitian ini mengusulkan pemodelan arsitektur *Advanced Retrieval-Augmented Generation* (RAG) yang mengombinasikan pencarian leksikal (BM25) dan pencarian semantik (IndoBERT) melalui mekanisme *Hybrid Search*, serta kalibrasi peringkat dokumen (*Reranking*) menggunakan model *Cross-Encoder*. Metodologi penelitian mencakup perancangan sistem, implementasi menggunakan LLM lokal Qwen2.5 7B, dan pengujian secara komprehensif menggunakan metrik *retrieval* (MRR, *Precision@K*, *Recall@K*) serta metrik kualitas jawaban *end-to-end* dengan *framework* RAGAS (*Faithfulness*, *Answer Relevancy*, *Context Precision*). Hasil pengujian statistik empiris menggunakan *Wilcoxon Signed-Rank Test* menunjukkan bahwa *pipeline Advanced RAG* memberikan peningkatan kinerja penelusuran dokumen yang signifikan dibandingkan RAG standar (*baseline*), yang secara langsung meningkatkan relevansi jawaban dari LLM. Kesimpulan penelitian ini menegaskan bahwa implementasi *Hybrid Search* dan *Reranking* terbukti efektif dalam memitigasi halusinasi, tangguh dalam mengenali istilah dan akronim akademik, serta menjamin privasi data institusi melalui pemrosesan inferensi lokal.

Kata kunci: *Retrieval-Augmented Generation* (RAG), *Large Language Model*, *Hybrid Search*, *Cross-Encoder Reranking*, Layanan Akademik

PEMODELAN PIPELINE ADVANCED RAG MENGGUNAKAN HYBRID SEARCH (BM25-INDOBERT) DAN RERANKING (CROSS-ENCODER) UNTUK LAYANAN INFORMASI AKADEMIK

GEDE JOVAN TIYO ADITYA

ABSTRACT

This research develops and evaluates an academic information service system based on a Large Language Model (LLM) at the Faculty of Computer Science. The main problems addressed are the limitations of manual information services by administrative staff and the inherent weaknesses of LLMs, namely hallucination and knowledge cutoff. To overcome these issues, this research proposes the modeling of an Advanced Retrieval-Augmented Generation (RAG) architecture that combines lexical search (BM25) and semantic search (IndoBERT) through a Hybrid Search mechanism, as well as document rank calibration (Reranking) using a Cross-Encoder model. The research methodology includes system design, implementation using the local Qwen2.5 7B LLM, and comprehensive testing using retrieval metrics (MRR, Precision@K, Recall@K) and end-to-end answer quality metrics using the RAGAS framework (Faithfulness, Answer Relevancy, Context Precision). Empirical statistical test results using the Wilcoxon Signed-Rank Test show that the Advanced RAG pipeline provides a significant improvement in document retrieval performance compared to standard RAG (baseline), which directly increases the relevance of the answers generated by the LLM. The conclusion of this study confirms that the implementation of Hybrid Search and Reranking is proven to be effective in mitigating hallucinations, robust in recognizing academic terms and acronyms, and guarantees institutional data privacy through local inference processing.

Keywords: Retrieval-Augmented Generation (RAG), Large Language Model, Hybrid Search, Cross-Encoder Reranking, Academic Services

DAFTAR ISI

HALAMAN SAMPUL	0
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Batasan Penelitian.....	4
1.5 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	7
2.1 Teori Utama	7
2.2 Teori Pendukung.....	12
2.3 Penelitian Terdahulu	22
2.4 Gap Penelitian	49
BAB III METODE PENELITIAN	51
3.1 Pendekatan Penelitian	51
3.2 Desain Penelitian	51
3.3 Subjek Penelitian	52

3.4 Instrumen Penelitian	53
3.5 Teknik Pengumpulan Data.....	55
3.6 Analisis Data.....	55
3.7 Prosedur Penelitian	56
3.8 Evaluasi Hasil Penelitian	59
BAB IV ANALISIS SISTEM DAN EVALUASI.....	60
4.1 Analisis Pemodelan <i>Pipeline Advanced</i> RAG	60
4.2 Evaluasi Arsitektur dan Hasil Pengujian	76
BAB V Kesimpulan dan saran.....	89
5.1 Kesimpulan	89
5.2 Saran	90
DAFTAR PUSTAKA.....	91
LAMPIRAN.....	95



DAFTAR TABEL

Tabel 2.1 Rangkuman Metrik Evaluasi	19
Tabel 2.2 Penelitian Terkait	22
Tabel 3.1 Spesifikasi Instrumen Penelitian.....	54
Tabel 3.2 Spesifikasi Perangkat Lunak.....	54
Tabel 4.1 Perbandingan Metrik <i>Retrieval</i> antara <i>Pipeline Baseline</i> dan <i>Advanced</i>	81
Tabel 4.2 Perbandingan Metrik RAGAS	83
Tabel 4.3 Hasil Uji Normalitas (Shapiro-Wilk) Metrik <i>Retrieval</i>	85
Tabel 4.4 Hasil Uji Hipotesis (<i>Wilcoxon Signed-Rank Test</i>) Metrik <i>Retrieval</i>	86
Tabel 4.5 Hasil Uji Normalitas (Shapiro-Wilk) Metrik RAGAS	86
Tabel 4.6 Hasil Uji Hipotesis (<i>Wilcoxon Signed-Rank Test</i>) Metrik RAGAS	87



DAFTAR GAMBAR

Gambar 2.1 Arsitektur RAG Standar dengan Tiga Komponen Utama.....	9
Gambar 2.2 Diagram Alur <i>Hybrid Search</i> dengan <i>Reciprocal Rank Fusion</i>	11
Gambar 2.3 Arsitektur <i>Advanced</i> RAG dengan <i>Hybrid Search</i> dan <i>Reranking</i> ...	12
Gambar 2.4 Perbandingan Arsitektur <i>Bi-Encoder</i> dan <i>Cross-Encoder</i>	16
Gambar 3.1 Desain Penelitian <i>Nonequivalent Control Group</i>	52
Gambar 3.2 Diagram Alir Prosedur Penelitian	56
Gambar 4.1 Ekstraksi teks per halaman (A) Panduan ULT (B) Ekstraksi ULT...	62
Gambar 4.2 Penggabungan teks seluruh halaman (A) Kumpulan teks per halaman (B) Teks gabungan keseluruhan.....	62
Gambar 4.3 Pemotongan (<i>chunking</i>) dokumen berbasis token (A) Teks gabungan utuh (B) <i>Array chunk</i> dengan <i>metadata</i>	64
Gambar 4.4 Demonstrasi <i>overlap</i> antar-chunk	64
Gambar 4.5 Normalisasi teks dokumen (A) Teks Markdown kotor (B) Teks bersih hasil normalisasi.....	66
Gambar 4.6 Representasi struktur indeks BM25 dan ChromaDB	68
Gambar 4.7 <i>Hybrid Search</i> dengan fusi RRF (A) Hasil pencarian independen (B) Hasil fusi RRF.....	69
Gambar 4.8 <i>Reranking</i> kandidat dokumen (A) Kandidat hasil RRF (B) Skor <i>Cross-Encoder</i>	71
Gambar 4.9 Penyesuaian prioritas berbasis aturan domain (A) Kandidat dan aturan (B) Hasil penyesuaian	72
Gambar 4.10 Penyusunan konteks (A) Teks dengan nama <i>file</i> asli (B) Teks dengan referensi generik	74
Gambar 4.11 Konstruksi <i>prompt (message list)</i> untuk LLM.....	74
Gambar 4.12 Jawaban akhir dari LLM beserta waktu eksekusi	75
Gambar 4.13 Perbandingan waktu respons (<i>retrieval</i> vs <i>generation</i>) antara <i>Baseline</i> dan <i>Advanced</i>	82
Gambar 4.14 Tampilan laporan visual evaluasi akhir.....	88

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi	95
Lampiran 2 Curriculum Vitae	97
Lampiran 3 Surat Pernyataan & Pengalihan HAKI	98
Lampiran 4 Tanda Terima BNSP	100
Lampiran 5 Surat Permohonan Observasi Data	101

