



SKRIPSI

**ANALISIS PERFORMA REASONING MODELS
PADA QUESTION ANSWERING SYSTEM
BERBASIS RAG SEBAGAI LAYANAN
INFORMASI PPMB UPN “VETERAN” JAWA
TIMUR**

MUHAMMAD SURYA ADHI SETIAWAN
NPM 21082010167

DOSEN PEMBIMBING

Arista Pratama, S.Kom., M.Kom.

Prasasti Karunia Farista Ananto, S.Kom., M.Kom., M.IM.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2026**

LEMBAR PENGESAHAN

ANALISIS PERFORMA REASONING MODELS PADA QUESTION ANSWERING SYSTEM BERBASIS RAG SEBAGAI LAYANAN INFORMASI PPMB UPN “VETERAN” JAWA TIMUR

Oleh:

MUHAMMAD SURYA ADHI SETIAWAN

NPM. 21082010167

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur pada tanggal 8 Januari 2026.

Menyetujui

Arista Pratama, S.Kom., M.Kom.
NPT. 17119910320052



(Pembimbing I)

Prasasti Karunia F.A., S.Kom., M.Kom., M.IM.
NIP. 199707042024062001



(Pembimbing II)

Seftin Fitri Ana Wati, S.Kom., M.Kom.
NPT. 21219910320267



(Ketua Penguji)

Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
NIP. 199409292022031008



(Penguji II)

Iqbal Ramadhani Mukhlis, S.Kom., M.Kom.
NIP. 199303052024061002



(Penguji III)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 196811261994032001

LEMBAR PERSETUJUAN

**ANALISIS PERFORMA REASONING MODELS PADA QUESTION ANSWERING
SYSTEM BERBASIS RAG SEBAGAI LAYANAN INFORMASI PPMB UPN
“VETERAN” JAWA TIMUR**

Oleh:

MUHAMMAD SURYA ADHI SETIAWAN

NPM. 21082010167

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,

Koordinator Program Studi Sistem Informasi

Fakultas Ilmu Komputer

Agung Brastama Putra, S.Kom., M.Kom.
NIP. 198511242021211003

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Muhammad Surya Adhi Setiawan
NPM : 21082010167
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya



Surabaya, Januari 2026

Mahasiswa



Muhammad Surya Adhi Setiawan
NPM. 21082010167

ABSTRAK

Nama Mahasiswa : Muhammad Surya Adhi Setiawan
NPM : 21082010167
Judul Skripsi : Analisis Performa Reasoning Models pada Question Answering System Berbasis RAG sebagai Layanan Informasi PPMB UPN “Veteran” Jawa Timur
Dosen Pembimbing : 1.Arista Pratama, S.Kom., M.Kom.
2.Prasasti Karunia Farista Ananto, S.Kom., M.Kom., M.IM.

Akses informasi akurat sangat krusial bagi calon mahasiswa baru, namun layanan konvensional terkendala volume pertanyaan tinggi dan risiko halusinasi model bahasa. Penelitian ini menganalisis performa *reasoning models* pada *Question Answering System* berbasis *Retrieval-Augmented Generation* (RAG) sebagai layanan informasi PPMB UPN “Veteran” Jawa Timur. Metodologi meliputi pembangunan *knowledge base* dari 30 sumber data, penggunaan *embedding* VoyageAI, serta dataset 353 komentar Instagram dalam bahasa Indonesia, Inggris, dan Jawa Suroboyoan. Evaluasi menggunakan *framework* RAGAS pada enam model kategori *reasoning* (o4-mini, gemini-2.5-flash, deepseek-r1) dan *non-reasoning* (gpt-4o-mini, gemini-2.0-flash, deepseek-v3). Hasil menunjukkan *reasoning models* secara konsisten mengungguli *non-reasoning models* dengan rata-rata skor 0,7772 berbanding 0,7289 (naik 6,63%). Keunggulan signifikan terlihat pada metrik *factual correctness* yang meningkat 15,95%, membuktikan efektivitas penalaran internal dalam menyintesis informasi kompleks. gemini-2.5 flash menjadi model paling optimal dengan skor 0,8207 dan ketahanan multibahasa yang stabil. Penelitian menyimpulkan bahwa integrasi *reasoning models* dalam arsitektur RAG mampu meminimalkan kesalahan informasi, meskipun kualitas sistem *retrieval* tetap menjadi faktor pembatas fundamental.

Kata Kunci: PPMB, *Question Answering System*, Ragas, *Reasoning Models*, *Retrieval-Augmented Generation* (RAG).

ABSTRACT

Student Name : Muhammad Surya Adhi Setiawan
Student ID : 21082010167
Thesis Title : *Analisis Performa Reasoning Models pada Question Answering System Berbasis RAG sebagai Layanan Informasi PPMB UPN “Veteran” Jawa Timur*
Advisors : 1. Arista Pratama, S.Kom., M.Kom.
2. Prasasti Karunia Farista Ananto, S.Kom., M.Kom., M.IM.

Accurate information is vital for prospective students, yet conventional services often face high query volumes and model hallucinations. This study analyzes reasoning models in a Retrieval-Augmented Generation (RAG)-based Question Answering System for student admissions at UPN “Veteran” Jawa Timur. The methodology involves building a knowledge base from 30 sources, utilizing VoyageAI embeddings, and a dataset of 353 Instagram comments in Indonesian, English, and Javanese Suroboyoan. Evaluation was conducted using the RAGAS framework across six reasoning (o4-mini, gemini-2.5-flash, deepseek-r1) and non-reasoning models (gpt-4o-mini, gemini-2.0-flash, deepseek-v3). Results show reasoning models consistently outperformed non-reasoning models with average RAGAS scores of 0.7772 versus 0.7289 (+6.63%). The most significant improvement occurred in factual correctness (+15.95%), proving the efficiency of internal reasoning in synthesizing complex data. gemini-2.5-flash emerged as the optimal model with a score of 0.8207 and stable multilingual robustness. The study concludes that integrating reasoning models into RAG architecture effectively minimizes misinformation, although retrieval quality remains a fundamental limiting factor.

Keywords: PPMB, Question Answering System, Ragas, Reasoning Models, Retrieval-Augmented Generation (RAG).

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas rahmat dan karunia-Nya, sehingga skripsi berjudul “**Analisis Performa Reasoning Models pada Question Answering System Berbasis RAG sebagai Layanan Informasi PPMB UPN “Veteran” Jawa Timur**” dapat diselesaikan sebagai salah satu syarat memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Penulis menyadari bahwa selesainya skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. **Kedua Orang Tua dan Keluarga** yang senantiasa memberikan doa, dukungan, serta motivasi yang menjadi kekuatan terbesar bagi penulis.
2. **Prof. Dr. Ir. Novirina Hendrasarie, MT.**, selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
3. **Bapak Arista Pratama, S.Kom., M.Kom.**, selaku Dosen Pembimbing I yang telah memberikan banyak waktu, arahan, serta ilmu yang sangat berharga dalam membimbing penulis menyelesaikan penelitian ini.
4. **Ibu Prasasti Karunia Farista Ananto, S.Kom., M.Kom., M.IM.**, selaku Dosen Pembimbing II yang telah memberikan masukan, saran, dan bimbingan teknis yang mendalam demi kesempurnaan skripsi ini.
5. **Seluruh Dosen dan Staf Program Studi Sistem Informasi** yang telah memberikan bekal ilmu dan bantuan administratif selama masa kuliah.
6. **Rekan-rekan mahasiswa Sistem Informasi angkatan 2021** serta seluruh pihak yang telah membantu secara langsung maupun tidak langsung.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan di bidang *Artificial Intelligence* dan Sistem Informasi.

Surabaya, Januari 2026

Muhammad Surya Adhi Setiawan

DAFTAR ISI

LEMBAR PENGESAHAN.....	iii
LEMBAR PERSETUJUAN.....	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vii
ABSTRAK.....	ix
ABSTRACT.....	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xix
DAFTAR TABEL.....	xxi
DAFTAR LAMPIRAN.....	xxiii
BAB I	
PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	5
1.6. Sistematika Penulisan.....	6
BAB II	
TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu.....	7
2.2. Gap Penelitian.....	10
2.3. Dasar Teori.....	10
2.3.1. Artificial Intelligence.....	11
2.3.2. Natural Language Processing.....	11
2.3.3. Large Language Models.....	11
2.3.4. Reasoning Models.....	12
2.3.5. Question Answering System.....	12
2.3.6. Retrieval-Augmented Generation.....	13
2.3.7. Chunking.....	14
2.3.8. Embedding.....	15
2.3.9. Query Rewriting.....	15
2.3.10. Information Retrieval.....	15
2.3.11. Prompt Engineering.....	16
2.3.12. Natural Language Generation.....	17
2.3.13. Google Colab.....	17
2.3.14. Python.....	17
2.3.15. LangChain.....	18
2.3.16. ChromaDB.....	18
2.3.17. Retrieval Augmented Generation Assessment (Ragas).....	18
2.3.18. Gradio.....	20

BAB III	
METODOLOGI PENELITIAN.....	21
3.1. Studi Literatur.....	21
3.2. Pengumpulan Data.....	21
3.2.1. Analisis Kebutuhan Data.....	21
3.2.2. Identifikasi Sumber Data.....	22
3.2.3. Pra-pemrosesan Data.....	22
3.3. Pembuatan Knowledge Base.....	27
3.3.1. Chunking.....	27
3.3.2. Embedding.....	28
3.3.3. Vector Storing.....	29
3.4. Implementasi RAG.....	30
3.4.1. Query Rewriting.....	30
3.4.2. Query Embedding.....	32
3.4.3. Similarity Search.....	32
3.4.4. Prompt Engineering.....	34
3.4.5. Answer Generation.....	36
3.5. Penyusunan Dataset Pengujian.....	37
3.5.1. Comment Collection.....	37
3.5.2. Comment Classification.....	37
3.5.3. Comment Enrichment.....	37
3.5.4. Ground Truth Generation.....	37
3.6. Pengujian dan Evaluasi.....	38
3.6.1. Batch Retrieval.....	38
3.6.2. Batch Generation.....	39
3.6.3. Batch Evaluation.....	41
3.7. Analisis Hasil.....	42
3.7.1. Analisis Performa Metrik RAGAS.....	42
3.7.2. Analisis Performa Multibahasa.....	42
3.7.3. Analisis Kualitatif.....	43
3.7.4. Analisis Efisiensi Biaya dan Komputasi.....	43
3.7.5. Sintesis Hasil dan Rekomendasi Model Paling Optimal.....	43
3.8. Deployment.....	45
3.8.1. Presentation Layer.....	46
3.8.2. Application Layer.....	46
3.8.3. Data Layer.....	46
BAB IV	
HASIL DAN PEMBAHASAN.....	47
4.1. Pengumpulan Data.....	47
4.1.1. Analisis Kebutuhan Data.....	47
4.1.2. Identifikasi Sumber Data.....	48
4.1.3. Pra-pemrosesan Data.....	51
4.2. Pembuatan Knowledge Base.....	55

4.2.1. Chunking.....	55
4.2.2. Embedding.....	58
4.2.3. Vector Storing.....	58
4.3. Implementasi RAG.....	59
4.3.1. Query Rewriting.....	59
4.3.2. Query Embedding.....	60
4.3.3. Similarity Search.....	61
4.3.4. Prompt Engineering.....	62
4.3.5. Answer Generation.....	63
4.4. Penyusunan Dataset Pengujian.....	64
4.4.1. Comment Collection.....	64
4.4.2. Comment Classification.....	65
4.4.3. Comment Enrichment.....	65
4.4.4. Ground Truth Generation.....	66
4.5. Pengujian dan Evaluasi.....	67
4.5.1. Batch Retrieval.....	67
4.5.2. Batch Generation.....	68
4.5.3. Batch Evaluation.....	69
4.6. Analisis Hasil.....	70
4.6.1. Analisis Performa Metrik RAGAS.....	70
4.6.2. Analisis Performa Multibahasa.....	73
4.6.3. Analisis Kualitatif.....	78
4.6.4. Analisis Efisiensi Biaya.....	82
4.6.5. Sintesis Hasil dan Rekomendasi Model Paling Optimal.....	84
4.7. Deployment.....	85
BAB V	
PENUTUP.....	87
5.1. Kesimpulan.....	87
5.2. Saran.....	89
DAFTAR PUSTAKA.....	91
LAMPIRAN.....	97

DAFTAR GAMBAR

Gambar 2.1. Alur Retrieval-Augmented Generation.....	13
Gambar 3.1. Metodologi Penelitian.....	21
Gambar 3.2. Kode Fungsi identify_drive_url.....	22
Gambar 3.3. Kode Fungsi fetch_content.....	23
Gambar 3.4. Kode Fungsi choose_system_prompt.....	24
Gambar 3.5. Kode Fungsi convert_to_markdown.....	25
Gambar 3.6. Kode Fungsi save_conversion_result.....	26
Gambar 3.7. Alur Pembuatan Knowledge Base.....	27
Gambar 3.8. Kode Fungsi chunk_content.....	27
Gambar 3.9. Kode Fungsi initialize_embedding_function.....	28
Gambar 3.10. Kode Fungsi get_collection dan store_content.....	29
Gambar 3.11. Alur Retrieval-Augmented Generation.....	30
Gambar 3.12. Kode Fungsi rewrite_query.....	31
Gambar 3.13. Kode Fungsi get_collection dan similarity_search.....	33
Gambar 3.14. Kode Variabel system_prompt.....	35
Gambar 3.15. Kode Fungsi prompt_engineering.....	35
Gambar 3.16. Kode Fungsi generate_answer.....	36
Gambar 3.17. Kode Fungsi batch_retrieval.....	38
Gambar 3.18. Chatbot Arena LLM Leaderboard.....	39
Gambar 3.19. Kode Fungsi batch_generation.....	40
Gambar 3.20. Kode Fungsi batch_evaluation.....	41
Gambar 3.21. Arsitektur Prototype QA System berbasis RAG.....	45
Gambar 4.1. Output Fungsi fetch_content dari Dokumen PDF.....	52
Gambar 4.2. Output Fungsi fetch_content dari Endpoint API.....	52
Gambar 4.3. Output Fungsi fetch_content dari Halaman Web.....	52
Gambar 4.4. Hasil Pengumpulan Data dari Dokumen PDF.....	54
Gambar 4.5. Hasil Pengumpulan Data dari Endpoint API.....	54
Gambar 4.6. Hasil Pengumpulan Data dari Halaman Web.....	54
Gambar 4.7. Hasil Pembuatan Knowledge Base.....	58
Gambar 4.8. Output Fungsi prompt_engineering.....	62
Gambar 4.9. Hasil Scrapping Kolom Komentar Akun Instagram @ppmb.upnjatim.....	64
Gambar 4.10. Hasil Klasifikasi Komentar.....	65
Gambar 4.11. Hasil Penerjemahan Komentar ke Berbagai Bahasa.....	65
Gambar 4.12. Kunci Jawaban Bahasa Indonesia.....	66
Gambar 4.13. Kunci Jawaban Bahasa Jawa.....	66
Gambar 4.14. Kunci Jawaban Bahasa Inggris.....	67
Gambar 4.15. Hasil Batch Retrieval.....	68
Gambar 4.16. Hasil Batch Generation.....	68
Gambar 4.17. Hasil Batch Evaluation.....	69
Gambar 4.18. Rata-Rata Skor RAGAS setiap Kategori Model.....	71

Gambar 4.19. Distribusi Skor RAGAS setiap Model.....	71
Gambar 4.20. Rata-Rata Skor Metrik RAGAS setiap Kategori Model.....	72
Gambar 4.21. Distribusi Skor Metrik RAGAS setiap Model.....	72
Gambar 4.22. Rata-Rata Performa Bahasa setiap Kategori Model.....	74
Gambar 4.23. Distribusi Performa Bahasa setiap Model.....	75
Gambar 4.24. Heatmap Performa Bahasa setiap Model.....	76
Gambar 4.25. Robustness Score Multibahasa.....	77
Gambar 4.26. Tab Knowledge Base Manager.....	85
Gambar 4.27. Tab Chat Interface.....	85

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu 1.....	7
Tabel 2.2. Penelitian Terdahulu 2.....	8
Tabel 2.3. Penelitian Terdahulu 3.....	8
Tabel 2.4. Penelitian Terdahulu 4.....	9
Tabel 2.5. Penelitian Terdahulu 5.....	9
Tabel 4.1. Rincian Informasi yang Dibutuhkan Calon Mahasiswa Baru.....	48
Tabel 4.2. Hasil Identifikasi Sumber Data.....	49
Tabel 4.3. Jumlah Token pada 10 Indivisible Information Unit.....	55
Tabel 4.4. Hasil Chunking Informasi Umum SNPMB.....	56
Tabel 4.5. Hasil Embedding Informasi Umum SNPMB.....	58
Tabel 4.6. Hasil Query Rewriting.....	59
Tabel 4.7. Hasil Query Embedding.....	60
Tabel 4.8. Hasil Similarity Search.....	61
Tabel 4.9. Performa Model pada Metrik RAGAS.....	70
Tabel 4.10. Performa Model pada Ketiga Bahasa.....	73
Tabel 4.11. Rata-Rata Biaya per Jawaban.....	82
Tabel 4.12. Cost per Performance Point setiap Model.....	83
Tabel 4.13. Proyeksi Biaya Operasional Bulanan setiap Model.....	83
Tabel 4.14. Matriks Pemilihan Model paling Optimal.....	84

DAFTAR LAMPIRAN

Lampiran 1. Surat Pengantar Skripsi.....	97
Lampiran 2. Kode Deployment Prototype pada File app.py.....	98
Lampiran 3. Requirements yang Dibutuhkan pada requirements.txt.....	118
Lampiran 4. Contoh Environment Variable pada File .env.example.....	118