

LAPORAN TUGAS AKHIR

AI Engineering

IMPLEMENTASI RETRIEVAL-AUGMENTED GENERATION UNTUK KLASIFIKASI KODE ARSIP

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat
Sarjana Komputer**



Alfa Nada Yulaswara

210711378

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2025**

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

IMPLEMENTASI RETRIEVAL-AUGMENTED GENERATION UNTUK KLASIFIKASI KODE ARSIP

yang disusun oleh

Alfa Nada Yulaswara

210711378

dinyatakan telah memenuhi syarat pada tanggal 26 November 2025

		Keterangan
Dosen Pembimbing 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Dosen Pembimbing 2	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Penguji 2	: Prof. Ir. Suyoto, M.Sc., Ph.D.	Telah Menyetujui
Penguji 3	: Dra. Ernawati, M.T., Ph.D.	Telah Menyetujui

Yogyakarta, 26 November 2025

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Prof. Dr. Ir. Parama Kartika Dewa SP., S.T., M.T., IPU, ASEAN Eng.

PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Alfa Nada Yulaswara
NPM : 210711378
Program Studi : Informatika
Fakultas : Teknologi Industri
Judul Penelitian : Implementasi Retrieval-Augmented-Generation
Untuk Klasifikasi Kode Arsip

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 13 November 2025

Yang menyatakan,

Alfa Nada Yulaswara

210711378

PERNYATAAN PERSETUJUAN DARI INSTANSI ASAL PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama Lengkap Pembimbing : Teguh Satira.S.STP
Jabatan : Kepala Sub bagian Administrasi Umum dan
Keuangan
Departemen : Inspektorat Daerah Kabupaten Purworejo

Menyatakan dengan ini:

Nama Lengkap : Alfa Nada Yulaswara
NPM : 210711378
Program Studi : Informatika
Fakultas : Teknologi Industri
Judul Penelitian : Implementasi Retrieval-Augmented-Generation
Untuk Klasifikasi Kode Arsip

1. Penelitian telah selesai dilaksanakan di perusahaan.
2. Perusahaan telah melakukan sidang internal berupa kelayakan penelitian ini dan akan mencantumkan lembar penilaian secara tertutup kepada pihak universitas sebagai bagian dari nilai akhir mahasiswa.
3. Memberikan persetujuan kepada pihak Universitas Atma Jaya Yogyakarta untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian dalam bentuk produk hasil penelitian meliputi dokumen, perangkat lunak, dan data hasil penelitian selama tetap mencantumkan nama penulis.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Purworejo, 25 Oktober 2025

Yang menyatakan,



Teguh Satira.S.STP

Kepala Sub bagian
Administrasi dan Keuangan

HALAMAN PERSEMBAHAN

Semua akan indah pada waktu-Nya

**Halaman persembahan ini ditujukan untuk orangtua, keluarga, rekan
sejawat.**

KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan Tugas Akhir “Implementasi Retrieval-Augmented-Generation Untuk Klasifikasi Kode Arsip” ini dengan baik. Penulisan Tugas Akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat sarjana komputer dari Program Studi Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta. Penulis menyadari bahwa dalam pembuatan tugas akhir ini penulis telah mendapatkan bantuan, bimbingan, dan dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang selalu membimbing dalam iman-Nya, memberikan berkat-Nya, dan menyertai penulis selalu.
2. Bapak Yohanes Sigit Purnomo Wuryo Putro, S.T., M.Kom., Ph.D., selaku dosen pembimbing I yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
3. Bapak Aloysius Gonzaga Pradnya S., S.T., M.Eng., selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
4. Bapak Teguh Satira.S.STP., selaku pembimbing lapangan yang telah menyediakan tempat dan membantu dalam penelitian.

Demikian laporan tugas akhir ini dibuat, dan penulis mengucapkan terima kasih kepada semua pihak. Semoga laporan ini dapat bermanfaat bagi pembaca.

Yogyakarta, 13 November 2025

Alfa Nada Yulaswara

210711378

DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH.....	ii
PERNYATAAN PERSETUJUAN DARI INSTANSI ASAL PENELITIAN	iv
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
ABSTRAK.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Penelitian	3
1.4. Pertanyaan Penelitian	4
1.5. Tujuan Penelitian	4
1.6. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. Tinjauan Pustaka	6
2.1.1. Penelitian Tentang Pencarian Nomor Surat dalam Sistem Arsip Elektronik.....	6
2.1.2. Penelitian yang menggunakan <i>Semantic Search</i>	7
2.1.3. Penelitian tentang Digitalisasi Pemberian Nomor Surat dan Pengarsipan Arsip	7
2.1.4. Penelitian tentang Penerapan AI dalam Klasifikasi Dokumen	8
2.1.5. Pemanfaatan OpenAI API dalam Aplikasi Teks Otomatis	9
2.1.6. Implementasi <i>Retrieval-Augmented Generation</i> (RAG) dalam	

Pembangunan Sistem Kecerdasan Buatan	9
2.2. Landasan Teori.....	15
2.2.1. Klasifikasi Dokumen.....	15
2.2.2. <i>Natural Language Processing (NLP)</i>	16
2.2.3. <i>Retrieval Augmented Generation (RAG)</i>	16
2.2.4. <i>Large Language Model (LLM)</i>	18
2.2.5. <i>Vector Embedding</i>	19
2.2.6. <i>Semantic Search</i>	19
2.2.6.1. <i>Cosine Similarity</i>	20
BAB III METODOLOGI PENELITIAN.....	21
3.1. Pengantar.....	21
3.2. Tahapan Penelitian	22
3.2.1. <i>Model Development</i>	22
3.2.1.1. <i>Knowledge Document Preparation</i>	23
3.2.1.2. <i>Knowledge Embedding</i>	24
3.2.1.3. <i>Knowledge Storage</i>	24
3.2.1.4. <i>Query Retrieval</i>	24
3.2.1.5. <i>Answer Generation</i>	24
3.2.1.6. <i>Fine Tuning</i>	25
3.2.2. Model Implementasi Dalam Bentuk Website	25
3.2.2.1. Analisis Kebutuhan	26
3.2.2.2. Implementasi Klasifikasi dengan RAG.....	27
3.2.2.3. Implementasi Digitalisasi Pengelolaan Arsip	28
3.3. Usulan Solusi	28
3.4. Pengumpulan Data	29

3.5.	<i>Setting Penelitian</i>	29
3.6.	Alat Bantu atau Tools yang Digunakan	29
3.7.	Evaluasi	30
BAB IV IMPLEMENTASI		32
4.1.	Pengantar.....	32
4.2.	<i>Model Development</i>	32
4.2.1.	<i>Knowledge Document Preparation</i>	32
4.2.2.	<i>Knowledge Embedding</i>	40
4.2.3.	<i>Knowledge Storage</i>	42
4.2.4.	<i>Query Retrieval</i>	45
4.2.5.	<i>Answer Generation</i>	47
4.2.6.	<i>Fine Tuning</i>	49
4.3.	Website Development	54
4.3.1.	Implementasi Model dan Sistem.....	54
4.3.2.	Implementasi Digitalisasi Sistem Pengelolaan Arsip	60
4.3.2.1.	Pengembangan Database.....	60
4.3.2.2.	Pengembangan Antarmuka	63
BAB V HASIL DAN PEMBAHASAN.....		66
5.1.	Pengantar.....	66
5.2.	Hasil	66
5.2.1.	Hasil Klasifikasi	66
5.2.2.	Hasil Implementasi Sistem.....	68
5.3.	Pembahasan.....	71
5.3.1.	Evaluasi	71
5.3.2.	<i>Analisis Kelebihan dan Kekurangan klasifikasi dengan RAG</i>	80
5.3.3.	<i>Analisis Kelebihan dan kekurangan implementasi sistem</i>	81

BAB VI KESIMPULAN DAN SARAN	83
6.1. Kesimpulan	83
6.2. Saran.....	84
DAFTAR PUSTAKA	85

DAFTAR GAMBAR

Gambar 2. 1 Workflow RAG	17
Gambar 3. 1 Arsitektur sistem aplikasi web	26
Gambar 4. 1 Hasil ekstraksi data dokumen	34
Gambar 4. 2 Struktur hierarki pada dokumen	35
Gambar 4. 3 Membuat database vektor di pinecone	43
Gambar 4. 4 Dashboard pembuatan model fine tuning.....	54
Gambar 4. 5 Antarmuka input query.....	55
Gambar 4. 6 Antarmuka hasil klasifikasi	57
Gambar 4. 7 ERD database	61
Gambar 4. 8 Antarmuka menambahkan data	64
Gambar 4. 9 Antarmuka sistem kelola arsip	65
Gambar 5. 1 UI Halaman Utama	68
Gambar 5. 2 UI Setelah Klasifikasi.....	69
Gambar 5. 3 UI Upload File.....	70
Gambar 5. 4 UI Setelah Klasifikasi dengan File.....	71
Gambar 5. 5 Grafik Akurasi.....	72

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan Penelitian Terkait	10
Tabel 2. 2 Format penamaan kode klasifikasi arsip	16
Tabel 4. 1 Contoh kesalahan penulisan pada dokumen	34
Tabel 4. 2 Hasil konversi hierarkis	39
Tabel 5. 1 Contoh Hasil Klasifikasi berdasar query	66
Tabel 5. 2 Evaluasi dengan Kueri	72
Tabel 5. 3 Evaluasi dengan Input File.....	77

ABSTRAK

IMPLEMENTASI RETRIEVAL-AUGMENTED GENERATION UNTUK KLASIFIKASI KODE ARSIP

Alfa Nada Yulaswara

210711378

Penelitian ini bertujuan untuk membangun sistem klasifikasi kode arsip berbasis kecerdasan buatan yang dapat membantu pegawai pemerintah daerah dalam menentukan kode klasifikasi arsip secara otomatis sesuai dengan Peraturan Menteri Dalam Negeri Nomor 83 Tahun 2022. Sistem ini dikembangkan untuk mengatasi permasalahan dalam proses penentuan kode klasifikasi yang selama ini masih dilakukan secara manual, sehingga memakan waktu, rawan kesalahan, dan menghambat efisiensi pengelolaan arsip di instansi pemerintahan.

Metode penelitian yang digunakan adalah pendekatan eksperimen rekayasa perangkat lunak dengan penerapan *Retrieval-Augmented Generation (RAG)* sebagai dasar pengembangan model. Tahapan penelitian meliputi *knowledge document preparation*, *knowledge embedding*, *knowledge storage*, *query retrieval*, dan *answer generation*, serta penyempurnaan melalui *fine-tuning* menggunakan OpenAI API. Implementasi sistem dilakukan dalam bentuk aplikasi web berbasis CodeIgniter yang terhubung dengan Flask, serta didukung oleh basis data MySQL sebagai penyimpanan arsip digital.

Hasil penelitian menunjukkan bahwa sistem mampu mengklasifikasikan kode arsip secara otomatis berdasarkan isi dokumen dengan tingkat relevansi yang baik. RAG yang dikembangkan dapat memahami konteks isi teks dan memberikan kode klasifikasi yang sesuai. Implementasi sistem berbasis web juga berhasil mengintegrasikan proses klasifikasi dan penyimpanan arsip digital secara efisien, sehingga mempermudah pencarian, pencatatan, dan pelacakan dokumen.

Penelitian ini menunjukkan bahwa penerapan metode RAG efektif dalam mendukung digitalisasi pengelolaan arsip di lingkungan pemerintah daerah. Sistem ini berpotensi diterapkan secara lebih luas untuk meningkatkan efisiensi administrasi dan kepatuhan terhadap regulasi kearsipan nasional. Namun, sistem masih terbatas pada dokumen teks dan belum mencakup pengolahan arsip dalam format non-teks seperti gambar atau hasil pemindaian.

Kata Kunci: *Retrieval-Augmented Generation, Kode Arsip, Cosine Similarity*