

DAFTAR ISI

SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR.....	iv
PERNYATAAN BEBAS PLAGIASI.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
DAFTAR PERSAMAAN.....	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian.....	3
1.6 Tahapan Penelitian.....	3
1.7 Sistematika Penulisan	4
BAB II TINJAUAN LITERATUR.....	5
2.1 Kesehatan Ternak.....	5
2.2 Chatbot	5
2.3 Large Language Model (LLM).....	6
2.3.1 Generative Pre-Trained Transformer (GPT).....	8
2.3.2 GPT-4.....	9
2.4 Retrieval Augmented Generation (RAG)	9
2.4.1 Konsep Dasar dan Arsitektur RAG	9
2.4.2 Cara Kerja RAG	10
2.4.3 Pemotongan Dokumen (<i>Document Chunking</i>).....	11
2.4.4 Dense Retrieval.....	11
2.4.5 Embedding Model	12
2.4.6 FAISS.....	13
2.5 RAG-Fusion	14
2.6 <i>Semantic Deduplication</i> (SemDedup)	15
2.7 Metode Evaluasi	16
2.7.1 BLEU (Bilingual Evaluation Understudy)	17
2.7.2 BERTScore	17
2.7.3 LLM as a Judge	18
2.7.4 Ragas	18

2.8	Penelitian Terdahulu	19
BAB III METODOLOGI PENELITIAN		26
3.1	Pengumpulan Basis Pengetahuan	26
3.1.1	Pengumpulan Dokumen	26
3.1.2	Pengolahan Dokumen.....	27
3.1.3	Token Based Chunking.....	28
3.1.4	Penyimpanan ke <i>Vector Database</i> FAISS	28
3.2	Pengembangan Sistem RAG	29
3.2.1	RAG-Fusion	29
3.2.2	Deduplikasi Dokumen dengan SemDedup.....	31
3.2.3	Pembuatan Jawaban Oleh GPT-4	35
3.3	Pengujian Model.....	36
3.3.1	Pembuatan Data Pengujian.....	36
3.3.2	Pembuatan Rangkaian Pengujian	37
3.3.3	Pengujian Menggunakan Ragas, BLEU, dan BertScore	38
3.3.4	Pembuatan Antarmuka Aplikasi	43
3.4	Pengembangan Antarmuka.....	44
3.4.1	Perancangan.....	44
3.4.2	Pengembangan dengan Streamlit.....	44
3.4.3	Integrasi dengan Sistem RAG	45
BAB IV HASIL PENGUJIAN DAN PEMBAHASAN		46
4.1	Implementasi	46
4.1.1	Persiapan Basis Pengetahuan.....	46
4.1.2	Pengembangan Sistem RAG	48
4.1.3	Pengujian Model.....	53
4.1.4	Pengembangan Antarmuka.....	58
4.2	Hasil.....	59
4.2.1	Metrik Evaluasi Ragas	59
4.2.2	Metrik Evaluasi BLEU	60
4.2.3	Metrik Evaluasi BertScore.....	60
4.2.4	Kecepatan Respons dan Penggunaan Token.....	60
4.2.5	Antarmuka <i>Chatbot</i>	61
4.2.6	Perbandingan dengan ChatGPT.....	62
4.3	Pembahasan	65
4.3.1	Analisis Reduksi Redundansi oleh SemDedup	65
4.3.2	Evaluasi Pengaruh SemDedup terhadap Kualitas Respons.....	66
4.3.3	Evaluasi Efisiensi Waktu dan Kualitas Respons	67

4.3.4 Perbandingan Respons terhadap ChatGPT	67
BAB V PENUTUP	68
5.1 Kesimpulan.....	68
5.2 Saran	68
DAFTAR PUSTAKA	70
LAMPIRAN	73

DAFTAR TABEL

Tabel 2. 1 Perbedaan Embedding Teks Pada Konteks yang Berbeda.....	12
Tabel 2. 2 Penelitian Terdahulu	19
Tabel 3. 1 Contoh Pemotongan Dokumen.....	28
Tabel 3. 2 Perbandingan Teks Asli dengan Teks <i>Embedding</i>	29
Tabel 3. 3 Pengambilan <i>top-k Chunk</i> dengan FAISS.....	31
Tabel 3. 4 Contoh Chunks Hasil RRF	31
Tabel 3. 5 Contoh Hasil Embedding Chunk	32
Tabel 3. 6 Contoh <i>Assignment Step</i> K-Means	33
Tabel 3. 7 Contoh <i>Centroid</i> Baru.....	33
Tabel 3. 8 Contoh Hasil <i>Clustering Chunks</i>	34
Tabel 3. 9 Contoh Hasil <i>Pairwise Cosine Similarity</i>	34
Tabel 3. 10 Contoh Hasil SemDedup	35
Tabel 3. 11 <i>Pairwise Cosine Similarity</i>	43
Tabel 4. 1 Hasil Evaluasi Metrik BLEU	60
Tabel 4. 2 Hasil Evaluasi Metrik BertScore	60
Tabel 4. 3 Rata-rata Kecepatan Respons	61
Tabel 4. 4 Rata-rata Penggunaan Token dan Biaya	61
Tabel 4. 5 Perbandingan Respons dengan ChatGPT	62
Tabel 4. 6 Validasi dan Skoring Respons Tiap Model oleh Pakar	64
Tabel 4. 7 Perbandingan Metrik Evaluasi Baseline dan Semdedup	65
Tabel 4. 8 Perbandingan Evaluasi oleh Pakar.....	65

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur Model Transformer (Vaswani et al., 2017)	7
Gambar 2. 2 Arsitektur GPT-1 (Radford et al., 2018)	8
Gambar 2. 3 Arsitektur Umum RAG (Nguyen, 2023)	10
Gambar 2. 4 Alur RAG Secara Umum	10
Gambar 2. 5 Alur Proses SemDedup	16
Gambar 3. 1 Metodologi Penelitian	26
Gambar 3. 2 Contoh <i>Prompt</i> Pembuatan Basis Pengetahuan	27
Gambar 3. 3 Contoh Respons Deepseek	27
Gambar 3. 4 Contoh JSON Nama File	28
Gambar 3. 5 Contoh <i>Prompt</i> Perluasan Kueri	30
Gambar 3. 6 Contoh Hasil Perluasan Kueri	30
Gambar 3. 7 Contoh Hasil <i>Embedding</i> Kueri	30
Gambar 3. 8 Contoh Inisialisasi <i>Centroid</i>	32
Gambar 3. 9 Contoh Visualisasi Hasil <i>Clustering</i>	34
Gambar 3. 10 Contoh <i>Prompt</i> Menjawab Kueri Pengguna	36
Gambar 3. 11 Contoh Respons GPT-4 dari API	36
Gambar 3. 12 Contoh <i>Prompt</i> untuk Menghasilkan Data Evaluator	36
Gambar 3. 13 Alur Pengujian	37
Gambar 3. 14 Contoh Data Evaluasi	38
Gambar 3. 15 Contoh Pemberian Konteks Awal	38
Gambar 3. 16 Contoh <i>Prompt</i> Untuk Menilai <i>Chunks</i>	39
Gambar 3. 17 Contoh Respons dari <i>Prompt Faithfulness</i>	39
Gambar 3. 18 Contoh <i>Prompt</i> Menghasilkan Pertanyaan	39
Gambar 3. 19 Contoh Pertanyaan Dihasilkan Dari Respons	40
Gambar 3. 20 Contoh Potongan Embedding	40
Gambar 3. 21 Contoh Respon <i>factual correctness</i>	41
Gambar 3. 22 Contoh <i>Prompt Context Precision</i>	41
Gambar 3. 23 Contoh Respons <i>Context Precision</i>	42
Gambar 3. 24 Rancangan Kerangka Antarmuka Chatbot	44
Gambar 3. 25 Alur Proses Interaksi Pengguna, Antarmuka <i>Chatbot</i> , dan RAG	45
Gambar 4. 1 Gagal Membuat Respons GPT-4	59
Gambar 4. 2 Grafik Evaluasi <i>Faithfulness</i>	59
Gambar 4. 3 Grafik Evaluasi <i>Answer Correctness</i>	59
Gambar 4. 4 Grafik Evaluasi <i>Answer Correctness</i>	60
Gambar 4. 5 Grafik Evaluasi <i>Context Precision</i>	60
Gambar 4. 6 Perbandingan Kecepatan Respons	61
Gambar 4. 7 Perbandingan Penggunaan Token dan Biaya	61
Gambar 4. 8 Antarmuka Chatbot	62

DAFTAR PERSAMAAN

Persamaan 2.1 Kemiripan Pertanyaan dengan <i>Passage</i>	11
Persamaan 2.2 Pencarian Tetangga Terdekat.....	14
Persamaan 2.3 <i>K</i> -Tetangga Terdekat (K-NN).....	14
Persamaan 2.4 Pencarian Berdasarkan Radius	14
Persamaan 2.5 Penyimpanan FAISS dengan Kuantisasi	14
Persamaan 2.6 <i>Inner Product</i>	14
Persamaan 2.7 Jarak Euclidean (L2)	14
Persamaan 2.8 Skor RRF.....	15
Persamaan 2.9 Kesamaan Kosinus (<i>Cosine Similarity</i>)	16
Persamaan 2.10 Kesamaan Kosinus (<i>Cosine Similarity</i>).....	17
Persamaan 2.11 Skor <i>Faithfulness</i>	18
Persamaan 2.12 Skor <i>Answer Relevancy</i>	19
Persamaan 2.13 Skor <i>Context Relevancy</i>	19

DAFTAR LAMPIRAN

Lampiran 1 Surat Izin Pengambilan Data.....	73
Lampiran 2 Profil Pakar.....	74
Lampiran 3 Form Validasi dan Skoring Respons	75