

DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PERSEMBAHAN.....	ii
ABSTRAK	iii
ABSTRACT.....	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian	4
1.6 Kerangka Pemikiran.....	5
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 <i>State of the Art</i>	8
2.2 Dasar Teori	13
2.2.1 Haji & Umrah.....	13
2.2.2 Kelompok Bimbingan Ibadah Haji (KBIH).....	15
2.2.3 <i>Artificial Intelligence (AI)</i>	17
2.2.4 <i>Natural Language Processing (NLP)</i>	18
2.2.5 <i>Large Language Model (LLM)</i>	18
2.2.6 Chatbot.....	18
2.2.7 <i>Text-to-Text Transfer Transformer (T5)</i>	19
2.2.8 IndoT5	21
2.2.9 <i>Parameter-Efficient Fine-Tuning (PEFT)</i> dan LoRA	22
2.2.10 <i>Semantic Search</i> dan <i>Sentence Embedding</i>	24
2.2.11 <i>Retrieval-Augmented Generation (RAG)</i>	24

2.2.12	Halusinasi AI.....	26
2.2.13	Python	27
2.2.14	CRISP-DM (<i>Cross-Industry Standard Process for Data Mining</i>).....	27
2.2.15	Matriks Evaluasi.....	29
BAB III METODOLOGI PENELITIAN		33
3.1	<i>Business Understanding</i>	35
3.3.1	Identifikasi Masalah & Analisis Tujuan	35
3.3.2	Identifikasi Kebutuhan	35
3.2	<i>Data Understanding</i>	36
3.2.1	Pengumpulan Data	36
3.2.1	Eksplorasi & Analisis Struktur Data	38
3.3	<i>Data Preparation</i>	42
3.3.1	<i>Data Cleaning & Anotasi</i>	44
3.3.2	Teks Labeling dan <i>Pair Construction</i>	44
3.3.3	<i>Tokenization</i>	45
3.3.4	Augmentasi Data.....	47
3.3.5	Pembuatan Basis <i>Retrieval</i> Semantik.....	50
3.4	<i>Modeling</i>	51
3.4.1	Konfigurasi Model Dasar (<i>Baseline</i>)	53
3.4.2	Konfigurasi LoRA.....	54
3.4.3	Proses Pelatihan	55
3.4.4	Integrasi <i>Retrieval</i> Ringan (<i>Hybrid v1</i>).....	60
3.4.5	Integrasi <i>Retrieval</i> Bertingkat (<i>Hybrid v2</i>)	61
3.5	<i>Evaluation</i>	64
3.5.1	Evaluasi model	64
3.5.2	Evaluasi Kualitas <i>Retrieval</i>	66
3.5.3	Evaluasi Manual oleh Pengurus KBIH	66
3.6	Implementasi <i>Prototype</i>	67
3.6.1	Pengembangan <i>Backend</i>	67
3.6.2	Pengembangan <i>Frontend</i>	68
3.6.3	Implementasi pada <i>Website Prototipe</i>	68

BAB IV HASIL DAN PEMBAHASAN.....	69
4.1 Hasil <i>Business Understanding</i>	69
4.2 Hasil <i>Data Understanding</i>	71
4.2.1 Hasil Pengumpulan Data.....	71
4.2.2 Hasil Eksplorasi dan Analisis Data	74
4.3 Hasil <i>Data Preparation</i>	76
4.3.1 Hasil <i>Data Cleaning & Anotasi</i>	76
4.3.2 Hasil Teks Labeling dan <i>Pair Construction</i>	77
4.3.3 Hasil Tokenization.....	78
4.3.4 Hasil Augmentasi Data.....	81
4.3.5 Hasil Pembangunan Basis <i>Retrieval</i> Semantik	84
4.4 Hasil <i>Modeling</i>	84
4.4.1 Hasil Konfigurasi Model Dasar (<i>Baseline</i>) dan LoRA	85
4.4.2 Hasil Proses Pelatihan.....	86
4.4.3 Hasil Integrasi <i>Retrieval</i> Ringan (<i>Hybrid v1</i>).....	94
4.4.4 Hasil Integrasi <i>Retrieval</i> Bertingkat (<i>Hybrid v2</i>).....	96
4.5 Hasil <i>Evaluation</i>	98
4.5.1 Hasil Evaluasi model.....	98
4.5.2 Hasil Evaluasi Kualitas <i>Retrieval</i>	100
4.5.3 Hasil Evaluasi Manual oleh Pengurus KBIH.....	101
4.6 Hasil Implementasi <i>Prototype</i>	109
4.6.1 Hasil pengembangan <i>Backend</i>	109
4.6.2 Hasil Pengembangan <i>Frontend</i>	111
4.6.3 Hasil Implementasi pada <i>Website Prototipe</i>	112
4.7 Pembahasan.....	114
BAB V PENUTUP	121
5.1 Kesimpulan	121
5.2 Saran.....	122
DAFTAR PUSTAKA	123

DAFTAR GAMBAR

Gambar 1. 1 Kerangka pemikiran penelitian	5
Gambar 2. 1 Diagram Prisma	8
Gambar 2. 2 Enam tahapan CRISP-DM yang saling berkesinambungan.....	28
Gambar 3. 1 Diagram alur tahapan metode CRISP-DM.....	33
Gambar 3. 2 <i>Import Library</i>	37
Gambar 3. 3 <i>Load Dataset</i>	38
Gambar 3. 4 <i>Info Dataset</i>	39
Gambar 3. 5 <i>Struktur Dataset</i>	39
Gambar 3. 6 <i>Tahap Exploratory Data Analysis</i>	40
Gambar 3. 7 Distribusi kategori dan pemetaan	41
Gambar 3. 8 Diagram alur tahap <i>data preparation</i>	42
Gambar 3. 9 Konfigurasi awal tahap <i>data preparation</i>	43
Gambar 3. 10 Normalisasi teks dan proses <i>cleaning</i>	44
Gambar 3. 11 Penerapan <i>prefix</i> instruksi.....	45
Gambar 3. 12 Tahap tokenisasi	46
Gambar 3. 13 Pembagian dataset menjadi data latih, data validasi, dan data pengujian	47
Gambar 3. 14 Augmentasi Bahasa gaul.....	48
Gambar 3. 15 Augmentasi parafrase	49
Gambar 3. 16 Filter keamanan <i>fiqh</i>	50
Gambar 3. 17 Kode pembangunan basis retrieval semantik	51
Gambar 3. 18 Diagram alur tahap modeling	52
Gambar 3. 19 Verifikasi struktur <i>layer attention</i> IndoT5-Base.....	53
Gambar 3. 20 Konfigurasi LoRA dan jumlah parameter terlatih	54
Gambar 3. 21 <i>Generation config</i>	56
Gambar 3. 22 Konfigurasi <i>training arguments</i> dan <i>trainer</i>	57
Gambar 3. 23 Fungsi perhitungan metrik rouge dan bleu selama <i>training</i>	60
Gambar 3. 24 diagram alur <i>retrieval</i> bertingkat.....	61
Gambar 3. 25 Fungsi inti <i>hybrid_generate_v2</i> beserta konstanta <i>threshold</i>	63
Gambar 4. 1 Diagram jumlah data per kategori	75
Gambar 4. 2 Output data <i>cleaning</i>	77
Gambar 4. 3 <i>Output</i> kolom final dari dataset	81
Gambar 4. 4 Log output augmentasi parafrase	83
Gambar 4. 5 <i>Output</i> penyimpanan basis <i>retrieval</i> semantik	84
Gambar 4. 6 Verifikasi struktur <i>layer attention</i> IndoT5-Base.....	85
Gambar 4. 7 Konfigurasi LoRA dan jumlah parameter terlatih	86
Gambar 4. 8 Ringkasan hasil akhir proses <i>training</i>	88

Gambar 4. 9 Kurva <i>training loss</i> dan <i>validation loss</i> per <i>epoch</i>	91
Gambar 4. 10 Kurva rouge-1 per <i>epoch</i>	92
Gambar 4. 11 Kurva rouge-1 dan bleu per <i>epoch</i>	93
Gambar 4. 12 Implementasi fungsi <i>hybrid_generate_v1</i> sebagai <i>baseline</i> pembandingan	94
Gambar 4. 13 Implementasi fungsi <i>hybrid_generate_v1</i> sebagai <i>baseline</i> pembandingan	95
Gambar 4. 14 Distribusi Mode Keputusan	100
Gambar 4. 15 Struktur kode layanan inferensi pada modal	110
Gambar 4. 16 Struktur kode <i>gateway</i> API pada Fly.io	111
Gambar 4. 17 Tampilan <i>widget chatbot</i> pada <i>website prototipe</i> KBIH Miftahul Ulum	112
Gambar 4. 18 Hasil interaksi tanya-jawab pada <i>chatbot prototipe</i>	113
Gambar 4. 19 Bar chart perbandingan 3 pendekatan	116
Gambar 4. 20 <i>Donut chart</i> distribusi mode jawaban <i>Hybrid v2</i>	117



DAFTAR TABEL

Tabel 2. 1 <i>State of the art</i>	9
Tabel 2. 2 Rangkuman ritual utama ibadah haji dan umrah	14
Tabel 2. 3 Kategori Layanan dan Konteks Informasi yang Ditangani <i>Chatbot</i> KBIH	15
Tabel 2. 4 Komponen arsitektur <i>encoder-decoder</i> pada model T5	20
Tabel 2. 5 Perbandingan <i>Full Fine-Tuning</i> dan <i>LoRA Fine-Tuning</i>	23
Tabel 3. 1 Konfigurasi <i>hyperparameter fine-tuning</i> model.....	58
Tabel 3. 2 Target metrik evaluasi model.....	64
Tabel 3. 3 Rubrik validasi manual.....	66
Tabel 4. 1 Hasil <i>Business Understanding</i>	69
Tabel 4. 2 Contoh struktur dataset per kategori.....	71
Tabel 4. 3 Distribusi data per kategori layanan	77
Tabel 4. 4 Statistik dataset pada setiap tahapan <i>data preparation</i>	79
Tabel 4. 5 Distribusi Kategori pada Data Latih, Validasi, dan Pengujian	80
Tabel 4. 6 Hasil filtrasi augmentasi parafrase	82
Tabel 4. 7 Progres metrik <i>training baseline</i> per <i>epoch</i>	87
Tabel 4. 8 Progres metrik <i>training Hybrid</i> per <i>epoch</i>	89
Tabel 4. 9 contoh hasil pengujian awal fungsi <i>hybrid_generate_v2</i>	97
Tabel 4. 10 Perbandingan hasil evaluasi model pada tiga pendekatan.....	98
Tabel 4. 11 Contoh Pertanyaan dan Jawaban per Kategori	102
Tabel 4. 12 Rekapitulasi hasil validasi manual per kategori	105