

DAFTAR ISI

	Hlm.
HALAMAN JUDUL	
LEMBAR PERSETUJUAN	
LEMBAR PENGESAHAN	
LEMBAR PERNYATAAN KARYA SENDIRI	
HALAMAN PERSEMBAHAN	
ABSTRAK	i
<i>ABSTRACT</i>	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah Penelitian	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah Penelitian	4
1.5 Kerangka Pemikiran Penelitian	5
1.6 Sistematika Penulisan	6
BAB II KAJIAN LITERATUR	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori	14
2.2.1 <i>Frequently Asked Question</i>	14
2.2.2 Instagram	14
2.2.3 <i>Open-domain Question Answering</i>	15
2.2.4 <i>Retriever-Reader</i>	15
2.2.5 BM25	16
2.2.6 BERT	17
2.2.7 IndoBERT	18
2.2.8 CRISP-DM	18
2.2.9 <i>Data Flow Diagram</i>	20

2.2.10 <i>Haystack</i>	20
2.2.11 <i>Elasticsearch</i>	20
2.2.12 <i>Python</i>	20
2.2.13 <i>Flask</i>	21
2.2.14 Metrik Evaluasi.....	21
BAB III METODOLOGI PENELITIAN	23
3.1 <i>Business Understanding</i>	24
3.2 <i>Data Understanding</i>	24
3.3 <i>Data Preparation</i>	25
3.3.1 <i>Data Indexing</i>	26
3.3.2 <i>Data Extraction</i>	26
3.3.3 <i>Data Splitting</i>	26
3.3.4 Pembersihan <i>Whitespace</i>	26
3.3.5 <i>Tokenization</i>	27
3.4 <i>Modeling</i>	27
3.4.1 <i>Fine-Tuning Model Reader</i>	33
3.4.2 Evaluasi Hasil <i>Fine-Tuning</i>	34
3.4.3 Integrasi Model <i>Retriever</i> dan Model <i>Reader</i>	34
3.5 <i>Evaluation</i>	34
3.6 <i>Deployment</i>	35
3.6.1 Arsitektur Aplikasi.....	36
3.6.2 <i>Data Flow Diagram</i> Aplikasi.....	38
3.6.3 Rancangan Antarmuka Aplikasi.....	44
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Hasil <i>Business Understanding</i>	45
4.2 Hasil <i>Data Understanding</i>	46
4.2.1 Analisis Pembangkit Pertanyaan.....	46
4.2.2 Pengambilan Data.....	48
4.3 Hasil <i>Data Preparation</i>	51
4.3.1 <i>Data Indexing</i>	52
4.3.2 <i>Data Extraction</i>	53
4.3.3 <i>Data Splitting</i>	53
4.3.4 Pembersihan <i>Whitespace</i>	54
4.3.5 <i>Tokenization</i>	54

4.4 Hasil <i>Modeling</i>	57
4.4.1 <i>Fine-Tuning</i> Model <i>Reader</i>	57
4.4.2 Evaluasi Hasil <i>Fine-Tuning</i>	64
4.4.3 Integrasi Model <i>Retriever</i> dan Model <i>Reader</i>	66
4.5 Hasil <i>Evaluation</i>	67
4.5.1 Pengujian Model <i>Retriever</i>	67
4.5.2 Pengujian Model <i>Reader</i>	71
4.6 Hasil <i>Deployment</i>	77
BAB V SIMPULAN DAN SARAN	79
5.1 Simpulan.....	79
5.2 Saran	79
DAFTAR PUSTAKA	80



DAFTAR GAMBAR

Gambar 1.1 Kerangka Pemikiran.....	5
Gambar 2.1 Ilustrasi <i>OpenQA</i> [6]	15
Gambar 2.2 Arsitektur <i>Retriever-Reader</i> [6]	16
Gambar 2.3 <i>Pre-Training</i> dan <i>Fine-Tuning</i> pada BERT [32].....	17
Gambar 2.4 Tahapan pada CRISP-DM [36]	18
Gambar 3.1 Alur pada Metode Penelitian.....	23
Gambar 3.2 Alur Proses Persiapan Data	25
Gambar 3.3 <i>Mask Language Model</i> BERT [44]	32
Gambar 3.4 Skenario <i>Modeling</i>	33
Gambar 3.5 Skenario Evaluasi Model	34
Gambar 3.6 Arsitektur Pembangkit FAQ.....	36
Gambar 3.7 Contoh Data Komentar.....	37
Gambar 3.8 Diagram Konteks.....	38
Gambar 3.9 DFD Level 1	39
Gambar 3.10 DFD Level 2 Proses 1.1	40
Gambar 3.11 DFD Level 2 Proses 1.2	42
Gambar 3.12 Rancangan Tampilan Awal Halaman Generate FAQ	44
Gambar 3.13 Rancangan Tampilan Setelah Generate FAQ.....	44
Gambar 4.1 Sampel Data yang Ditarik dari Kolom Komentar Instagram.....	46
Gambar 4.2 Hasil <i>Bigram</i>	47
Gambar 4.3 Sampel Data Jawaban Pembangkit Pertanyaan.....	49
Gambar 4.4 Sampel Data SQuAD 2.0	49
Gambar 4.5 Jumlah Data pada <i>Dataset</i> SQuAD 2.0 yang Telah Diekstrak	50
Gambar 4.6 Histogram Data <i>Train</i> SQuAD 2.0 Berbahasa Indonesia	50
Gambar 4.7 Histogram Data <i>Dev</i> SQuAD 2.0 Berbahasa Indonesia.....	51
Gambar 4.8 Proses <i>Data Indexing</i>	52
Gambar 4.9 Hasil Proses <i>Data Indexing</i>	52
Gambar 4.10 Hasil Ekstraksi Kolom <i>Paragraphs</i> SQuAD 2.0	53
Gambar 4.11 Contoh <i>Data Splitting</i>	54
Gambar 4.12 Proses Pembersihan <i>Whitespace</i>	54
Gambar 4.13 Proses Pemisahan Input.....	54
Gambar 4.14 Proses Pemotongan Konteks	55
Gambar 4.15 Proses Konfigurasi <i>Overflowing Tokens</i>	55
Gambar 4.16 Proses Konfigurasi <i>Offset Mapping</i>	56
Gambar 4.17 Proses Tokenisasi	56
Gambar 4.18 Contoh Hasil Tokenisasi	57
Gambar 4.19 Sampel Hasil Tokenisasi	57
Gambar 4.20 Konfigurasi <i>Hyperparameter</i> dengan <i>TrainingArguments</i>	60
Gambar 4.21 Memuat model <i>pre-trained</i> dan <i>tokenizer</i>	60
Gambar 4.22 Implementasi Fungsi <i>Trainer</i>	61
Gambar 4.23 Visualisasi Nilai <i>Loss</i>	63
Gambar 4.24 Sampel Data Evaluasi.....	64

Gambar 4.25 Implementasi Fungsi <i>Pipeline</i>	64
Gambar 4.26 Sampel Hasil Pengujian Performa Model <i>reader</i>	65
Gambar 4.27 Contoh Integrasi Model <i>Retriever</i> dan Model <i>Reader</i>	66
Gambar 4.28 Sampel Data Pengujian	67
Gambar 4.29 Fungsi untuk Mendefinisikan Metrik Evaluasi <i>Retriever</i>	68
Gambar 4.30 Hasil Pengujian Model <i>Retriever</i> dengan <i>top_k</i> = 1	68
Gambar 4.31 Hasil Pengujian Model <i>Retriever</i> dengan <i>top_k</i> = 2	69
Gambar 4.32 Hasil Pengujian Model <i>Retriever</i> dengan <i>top_k</i> = 3	69
Gambar 4.33 Fungsi untuk Mendefinisikan Metrik Evaluasi <i>Reader</i>	71
Gambar 4.34 Hasil Prediksi Model <i>Reader</i> Sebelum Perbaikan	73
Gambar 4.35 Mendefinisikan Fungsi <i>Post-Processing</i>	74
Gambar 4.36 Penggunaan Fungsi <i>Post-Processing</i>	75
Gambar 4.37 Jawaban yang Dihasilkan Model Performa Paling Unggul	76
Gambar 4.38 Integrasi Pembangkit Pertanyaan dan Pembangkit Jawaban	77
Gambar 4.39 Tampilan Awal Halaman Pembangkit FAQ	78
Gambar 4.40 Tampilan Hasil <i>Generate</i> FAQ	78



DAFTAR TABEL

Tabel 2.1 <i>State of The Art</i>	7
Tabel 3.1 Karakteristik model <i>retriever</i> dan model <i>reader</i> yang digunakan.....	27
Tabel 3.2 Contoh Dokumen	28
Tabel 3.3 Contoh Perhitungan IDF	29
Tabel 3.4 Perhitungan BM25 pada Dokumen 1	30
Tabel 3.5 Perhitungan BM25 pada Dokumen 2	30
Tabel 3.6 Perhitungan BM25 pada Dokumen 3	31
Tabel 3.7 Hasil perhitungan BM25 pada ketiga dokumen.....	31
Tabel 3.8 Spesifikasi Perangkat Keras	35
Tabel 3.9 Spesifikasi Perangkat Lunak	36
Tabel 3.10 Struktur Data pada <i>Database</i>	37
Tabel 3.11 P-Spec Pembentukan FAQ	40
Tabel 3.12 P-Spec Mengambil Data Komentar	41
Tabel 3.13 P-Spec Mengambil Data Komentar yang Sering Muncul.....	41
Tabel 3.14 P-Spec Membuat dan Mengevaluasi Pertanyaan	42
Tabel 3.15 P-Spec Mengambil Dokumen Relevan dengan Pertanyaan.....	43
Tabel 3.16 P-Spec Membuat Kesimpulan Jawaban	43
Tabel 4.1 Daftar Pertanyaan yang Dihasilkan Pembangkit Pertanyaan.....	47
Tabel 4.2 Variasi <i>Data Splitting</i>	53
Tabel 4.3 Daftar <i>Hyperparameter</i> Yang Digunakan.....	58
Tabel 4.4 Eksperimen Nilai <i>Hyperparameter</i>	59
Tabel 4.5 Parameter Pada Fungsi <i>Trainer</i>	61
Tabel 4.6 Nilai <i>Loss</i> Setiap Model	62
Tabel 4.7 Pengujian Model Hasil <i>Fine-Tuning</i>	65
Tabel 4.8 Hasil Pengujian Model <i>Retriever</i>	70
Tabel 4.9 Hasil Pengujian Model <i>Reader</i>	72
Tabel 4.10 Pengujian Model <i>Reader</i> yang Telah Diperbaiki.....	75