

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI	v
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan dan Manfaat.....	6
1.4 Ruang Lingkup	7
1.5 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA	10
2.1 <i>State of the Art</i>	10
2.2 <i>Named Entity Recognition (NER)</i>	14
2.3 <i>Web Scraping</i>	15
2.4 Prapemrosesan Teks	16
2.5 Anotasi Data	16
2.5.1 Skema Pelabelan <i>Beginning, Inside, Outside (BIO)</i>	17
2.5.2 <i>Intra-annotator Agreement</i>	17
2.6 Augmentasi Data	19
2.6.1 <i>Mention Replacement (MR)</i>	20
2.6.2 <i>Contextual Word Replacement (CWR)</i>	20
2.7 <i>Arsitektur Transformer</i>	22
2.7.1 <i>Encoder</i>	23
2.7.1.1 <i>Multi-Head Attention</i>	24
2.7.1.2 <i>Feed Forward Network (FFN)</i>	25

2.7.2	<i>Decoder</i>	26
2.8	Arsitektur IndoBERT	26
2.9	<i>Recurrent Neural Network (RNN)</i>	28
2.9.1	<i>Long Short-Term Memory (LSTM)</i>	29
2.9.2	<i>Bidirectional Long Short-Term Memory (BiLSTM)</i>	31
2.9.3	<i>Gated Recurrent Unit (GRU)</i>	32
2.9.4	<i>Bidirectional Gated Recurrent Unit (BiGRU)</i>	34
2.10	<i>Fully Connected Layer dan Gaussian Error Linear Unit (GELU)</i>	35
2.11	<i>Conditional Random Field (CRF)</i>	37
2.12	<i>Optimasi Hyperparameter</i>	38
2.12.1	<i>Grid Search</i>	38
2.12.2	<i>Tree-structured Parzen Estimator (TPE) pada Library Optuna</i>	39
2.13	<i>Evaluasi Berbasis Entitas (Entity-Level Evaluation)</i>	40
2.13.1	<i>Skema Pencocokan Kriteria Entitas</i>	40
2.13.2	<i>Confusion Matrix</i>	42
2.13.3	<i>Precision</i>	42
2.13.4	<i>Recall</i>	43
2.13.5	<i>F1-Score</i>	44
BAB III METODOLOGI PENELITIAN		45
3.1	Pengumpulan Data	47
3.2	Pembersihan Data	47
3.2.1	Ekstraksi Teks Pertanyaan	47
3.2.2	Perbaikan <i>Encoding</i> dan <i>Unicode Escape</i>	47
3.2.3	Penghapusan Tag HTML dan Simbol Tidak Diperlukan	48
3.2.4	Normalisasi Karakter dan Spasi	48
3.2.5	Konversi Format Akhir	48
3.3	Anotasi Data	48
3.3.1	Definisi Entitas Medis	49
3.3.2	Pedoman Anotasi	52
3.3.3	Proses Anotasi Berbasis <i>Span</i> Menggunakan Doccano	53
3.3.4	Pengujian <i>Intra-annotator Agreement</i>	54
3.4	Konversi Data ke Format BIO	54
3.4.1	Tokenisasi Menggunakan IndoBERT <i>Tokenizer</i>	55

3.4.2	Span Entitas ke Label BIO	55
3.4.3	Penyelarasan Token dan Label	56
3.5	Pembagian Data.....	56
3.6	Penanganan Ketidakseimbangan Distribusi Entitas	57
3.6.1	Strategi Augmentasi	57
3.6.2	Augmentasi Data Menggunakan <i>Mention Replacement</i>	59
3.6.3	Augmentasi Data Menggunakan <i>Contextual Word Replacement</i>	59
3.7	Pembentukan Input Model	60
3.7.1	Penambahan Token Khusus [CLS] dan [SEP]	60
3.7.2	Konversi Token ke <i>Input IDs</i>	60
3.7.3	Pembuatan <i>Attention Mask</i>	61
3.7.4	<i>Padding</i> dan Penyeragaman Panjang Urutan	61
3.7.5	Penyesuaian Label Target.....	61
3.8	Pelatihan dan Pengembangan Model	61
3.8.1	Skenario 1: Optimasi Model Dasar (<i>Baseline</i>).....	62
3.8.1.1	Arsitektur Model Dasar.....	62
3.8.1.2	Optimasi <i>Hyperparameter</i> Menggunakan Optuna.....	65
3.8.1.3	Proses Pelatihan Model	67
3.8.1.4	Pelatihan Final Model	68
3.8.2	Skenario 2: Modifikasi Arsitektur Model.....	68
3.8.2.1	Skenario 2a: Peningkatan Dimensi <i>Hidden</i> BiGRU	69
3.8.2.2	Skenario 2b: Implementasi <i>Stacked Layer</i> BiGRU dan <i>Dropout</i>	71
3.8.2.3	Skenario 2c: Optimasi Lapisan Klasifikasi dengan GELU	73
3.9	Pemilihan Model Terbaik.....	76
3.10	Evaluasi Tingkat Entitas (<i>Entity-Level Evaluation</i>).....	77
3.10.1	Evaluasi Menggunakan <i>Precision</i> , <i>Recall</i> , dan <i>F1-score</i>	77
3.10.2	Analisis Melalui <i>Confusion Matrix</i> dan <i>Error Analysis</i>	78
3.11	Lingkungan Implementasi dan Pengujian	78
BAB IV HASIL DAN PEMBAHASAN.....		80
4.1	Hasil Pengumpulan Data	80
4.2	Hasil Pembersihan Data	81
4.3	Hasil Anotasi Data.....	82
4.3.1	Distribusi Entitas Medis	82

4.3.2	Contoh Hasil Pelabelan	83
4.3.3	Hasil Pengujian <i>Intra-annotator Agreement</i>	84
4.4	Hasil Konversi ke BIO	85
4.5	Hasil Pembagian Data	86
4.6	Hasil Augmentasi Data.....	86
4.7	Hasil Persiapan Input Model	88
4.8	Hasil Pengembangan dan Skenario Eksperimen	89
4.8.1	Skenario 1: Optimasi Model Dasar (<i>Baseline</i>).....	90
4.8.2	Skenario 2: Modifikasi Arsitektur Model.....	95
4.8.2.1	Skenario 2a: Peningkatan Dimensi <i>Hidden</i> BiGRU	95
4.8.2.2	Skenario 2b: Implementasi <i>Stacked Layer</i> BiGRU dan <i>Dropout</i>	97
4.8.2.3	Skenario 2c: Optimasi Lapisan Klasifikasi dengan GELU.....	99
4.9	Perbandingan Performa Antar Skenario.....	103
4.10	Analisis Hasil Prediksi Model Terbaik	106
4.10.1	Evaluasi Kinerja Keseluruhan (<i>Nevaluate</i>).....	106
4.10.2	Evaluasi Kinerja per Kategori Kelas (<i>Classification Report</i>)	107
4.10.3	Analisis Matriks Kesalahan (<i>Confusion Matrix</i>).....	109
4.10.4	Analisis Kesalahan Prediksi (<i>Error Analysis</i>).....	110
4.10.5	Implikasi terhadap Ekstraksi Informasi Kesehatan Ibu dan Anak	112
BAB V PENUTUP		114
5.1	Kesimpulan.....	114
5.2	Saran.....	115
DAFTAR PUSTAKA.....		116
LAMPIRAN 1. Perhitungan Manual <i>Intra-Annotator Agreement</i>		124
LAMPIRAN 2. Perhitungan Manual Evaluasi Kinerja Model Berbasis <i>Entity-Level</i>		127