

DAFTAR ISI

LEMBAR PERSEMBAHAN	i
ABSTRAK	iii
<i>ABSTRACT</i>	iv
KATA PENGANTAR.....	v
DAFTAR ISI	vii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
1.6 Kerangka Pemikiran	6
1.7 Sistematika Penulisan.....	7
BAB II KAJIAN LITERATUR.....	9
2.1 Hasil Penelitian Terdahulu.....	9
2.2 Dasar Teori	15
2.2.1 Kepuasan Mahasiswa.....	15
2.2.2 Pelatihan ICT	15
2.2.3 <i>Natural Language Processing</i> (NLP)	16
2.2.4 <i>Text Preprocessing</i>	17
2.2.5 Klasifikasi Teks.....	18
2.2.6 Analisis Sentimen	19

2.2.7 <i>Deep Learning</i>	19
2.2.8 <i>Arsitektur Transformer</i>	20
2.2.9 <i>Bidirectional Encoder Representations from Transformers (BERT)</i> ...	22
2.2.10 <i>IndoBERT</i>	24
2.2.11 <i>Penanganan Imbalanced Dataset</i>	25
2.2.12 <i>Confussion Matrix</i>	26
2.2.13 <i>Metrik Evaluasi Klasifikasi</i>	27
2.2.14 <i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i>	29
BAB III METODOLOGI PENELITIAN	32
3.1 <i>Business Understanding</i>	33
3.1.1 <i>Identifikasi Permasalahan</i>	33
3.1.2 <i>Menetapkan Tujuan Penelitian</i>	34
3.1.3 <i>Menentukan Kebutuhan Sistem Klasifikasi</i>	34
3.2 <i>Data Understanding</i>	34
3.2.1 <i>Pengumpulan Dataset ICT</i>	35
3.2.2 <i>Analisis Struktur Dataset</i>	35
3.2.3 <i>Eksplorasi Karakteristik Dataset</i>	36
3.3. <i>Data Preparation</i>	36
3.3.1 <i>Data Selection</i>	37
3.3.2 <i>Cleaning Data</i>	38
3.3.3 <i>Transformasi Dataset</i>	39
3.3.4 <i>Filtering Noise</i>	40
3.3.5 <i>Labeling</i>	41
3.3.6 <i>Text Processing</i>	41
3.3.7 <i>Label Encoding</i>	43
3.3.8 <i>Splitting Data (Pembagian Data)</i>	44

3.3.9 Penanganan <i>Imbalanced Dataset</i>	45
3.3.10 <i>Tokenization</i>	45
3.4 <i>Modeling</i>	47
3.4.1 Pembentukan <i>Dataset PyTorch</i>	47
3.4.2 Inisialisasi Model IndoBERT.....	48
3.4.2 Penentuan <i>Hyperparameter</i>	49
3.4.3 <i>Fine-Tuning</i> Model	49
3.4.5 Prediksi	49
3.5 <i>Evaluation</i>	50
3.5.1 Metrik Evaluasi.....	50
3.5.2 <i>Confusion Matrix</i>	51
3.5.3 Analisis Hasil Evaluasi	51
3.6 <i>Deployment</i>	52
BAB IV HASIL DAN PEMBAHASAN.....	54
4.1 Hasil Rumusan Masalah 1	54
4.1.1 Hasil <i>Business Understanding</i>	54
4.1.2 Hasil <i>Data Understanding</i>	55
4.1.2.1 Pengumpulan Dataset.....	55
4.1.2.2 Analisis Struktur Dataset	57
4.1.2.3 Eksplorasi Karakteristik Dataset	58
4.1.3 Hasil <i>Data Preparation</i>	61
4.1.3.1 Hasil <i>Data Selection</i>	61
4.1.3.2 Hasil <i>Cleaning Data</i>	63
4.1.3.3 Hasil <i>Data Transformation</i>	66
4.1.3.4 Hasil <i>Filtering Noise</i>	69
4.1.3.5 Hasil <i>Labeling</i>	71

4.1.3.6 Hasil <i>Text Preprocessing</i>	73
4.1.3.7 Hasil <i>Label Encoding</i>	77
4.1.3.8 Hasil <i>Data Splitting</i> (Pembagian Data)	78
4.1.3.9 Hasil Penanganan <i>Imbalance Data</i>	81
4.1.3.10 Hasil Tokenisasi	82
4.1.4. Hasil <i>Modeling</i>	83
4.1.4.1 Pembentukan Dataset PyTorch	83
4.1.4.2 Inisialisasi Model IndoBERT	85
4.1.4.3 Penentuan <i>Hyperparameter</i>	86
4.1.4.4 Hasil <i>Fine Tuning Model</i>	87
4.1.4.5 Prediksi	91
4.1.5. Hasil <i>Deployment</i>	92
4.2 Hasil Rumusan Masalah 2	93
4.2.1 Hasil <i>Evaluation</i>	93
4.2.1.1 Metrik Evaluasi	94
4.2.1.2 <i>Confussion Matrix</i>	95
4.2.1.3 Analisis Hasil Evaluasi Model	97
4.2.1.4 Analisis Kesesuaian Label dengan Teks	98
4.2.2 Eksperimen Pelabelan Berbasis Teks	101
BAB V PENUTUP	107
5.1 Kesimpulan	107
5.2 Saran	108
DAFTAR PUSTAKA	110

DAFTAR GAMBAR

Gambar 1. 1 Kerangka Pemikiran.....	6
Gambar 2. 1 Alur PRISMA Diagram.....	9
Gambar 2. 2 Arsitektur Transformer.....	21
Gambar 2. 3 Arsitektur <i>Pre-training</i> dan <i>Fine-tuning</i> BERT.....	22
Gambar 2. 4 <i>Confusion Matrix</i>	26
Gambar 2. 5 Tahapan Metodologi CRISP-DM.....	30
Gambar 3. 1 Alur Penelitian.....	32
Gambar 3. 2 Alur Tahap <i>Bussines Understanding</i>	33
Gambar 3. 3 Alur Tahap <i>Data Understanding</i>	34
Gambar 3. 4 Alur Tahap <i>Data Preparation</i>	37
Gambar 3.5 Alur Tahap <i>Modeling</i>	47
Gambar 3.6 Alur Klasifikasi Menggunakan IndoBERT.....	48
Gambar 3. 7 Alur <i>Deployment</i>	52
Gambar 4. 1 Data Survei pada Sistem Pelatihan ICT.....	56
Gambar 4. 2 <i>Raw</i> Dataset Hasil Ekspor ke Format CSV.....	56
Gambar 4.3 Informasi Struktur Dataset.....	57
Gambar 4.4 Tampilan Awal Raw Dataset.....	58
Gambar 4.5 Distribusi Jenis.....	59
Gambar 4.6 Karakteristik Pertanyaan Survei berdasarkan Jenis.....	59
Gambar 4.7 Pemeriksaan <i>Missing Value</i>	60
Gambar 4. 8 Jumlah Atribut Sebelum diseleksi.....	61
Gambar 4. 9 Jumlah Atribut Setelah diseleksi.....	61
Gambar 4. 10 <i>Missing Value</i> Sebelum <i>Cleaning</i>	63
Gambar 4. 11 <i>Missing Value</i> Setelah <i>Cleaning</i>	63
Gambar 4. 12 Pemeriksaan Data Duplikat.....	64
Gambar 4. 13 Sebelum Pembersihan Karakter.....	65
Gambar 4. 14 Setelah Pembersihan Karakter.....	65
Gambar 4. 15 Pembersihan Karakter pada Atribut Jawaban.....	65
Gambar 4. 16 Penghapusan Jawaban Kosong.....	66
Gambar 4.17 Sebelum Pengelompokan Data.....	66
Gambar 4.18 Hasil <i>Grouping</i> Data.....	67

Gambar 4. 19 Hasil Penyesuaian Nama Atirbut	67
Gambar 4. 20 Tipe Data Rating Hasil Konversi	68
Gambar 4. 21 Hasil Perhitungan Rating <i>Mean</i>	68
Gambar 4. 22 Hasil Transformasi Dataset	69
Gambar 4.23 Data <i>Noise</i>	69
Gambar 4. 24 Distribusi Label Data <i>Noise</i>	70
Gambar 4. 25 Dataset Setelah <i>Filtering Noise</i>	71
Gambar 4. 26 Rating <i>Mean</i> Sebelum <i>Labeling</i>	71
Gambar 4. 27 Hasil Pembentukan Label.....	71
Gambar 4. 28 Grafik Distribusi Label Tingkat Kepuasan.....	72
Gambar 4. 29 Dataset Akhir Hasil <i>Labeling</i>	73
Gambar 4. 30 Hasil <i>Case Folding</i>	73
Gambar 4. 31 Hasil <i>Cleaning Text</i>	74
Gambar 4. 32 Hasil Penghapusan Penanda.....	74
Gambar 4. 33 Hasil Normalisasi Kata.....	75
Gambar 4. 34 Pemeriksaan Data Kosong	76
Gambar 4. 35 Data Hasil <i>Preprocessing</i>	76
Gambar 4. 36 Hasil <i>Label Encoding</i>	77
Gambar 4. 37 Visualisasi Hasil Data <i>Splitting</i> Skenario 60,20,20.....	78
Gambar 4. 38 Visualisasi Hasil Data <i>Splitting</i> Skenario 70,15,15.....	79
Gambar 4. 39 Visualisasi Hasil Data <i>Splitting</i> Skenario 80,10,10.....	80
Gambar 4.40 Hasil Konversi Tokenisasi.....	83
Gambar 4. 41 Pembuatan Class IndoBERTDataset	84
Gambar 4. 42 Pembentukan <i>train_dataset</i> , <i>valid_dataset</i> , dan <i>test_dataset</i>	85
Gambar 4. 43 Proses <i>Load</i> Model IndoBERT	85
Gambar 4. 44 Hasil <i>Fine-Tuning</i> Skenario 1	87
Gambar 4. 45 Hasil <i>Fine-Tuning</i> Skenario 2	88
Gambar 4. 46 Hasil <i>Fine-Tuning</i> Skenario 3	89
Gambar 4. 47 Implementasi Proses Prediksi.....	91
Gambar 4. 48 Tampilan Awal <i>Streamlit</i>	92
Gambar 4. 49 Hasil Prediksi	93

DAFTAR TABEL

Tabel 2. 1 <i>State Of The Art</i>	11
Tabel 3. 1 Aturan Pembentukan Label dan Distribusi Data Kepuasan	41
Tabel 4. 1 Hasil Tahap <i>Business Understanding</i>	54
Tabel 4. 2 Ringkasan Dataset Penelitian	57
Tabel 4. 3 Seleksi Pertanyaan Survei yang Digunakan	62
Tabel 4.4 Distribusi Label Tingkat Kepuasan	72
Tabel 4. 5 Hasil Skenario 3 Split Data	80
Tabel 4. 6 Hasil <i>Random Oversampling</i> pada Data <i>Training</i>	81
Tabel 4. 7 Contoh Sebelum dan Sesudah Tokenisasi	82
Tabel 4. 8 Konfigurasi Pelatihan	86
Tabel 4. 9 Hasil perbandingan Proses <i>Fine-Tuning</i>	90
Tabel 4. 10 Hasil Perbandingan Metrik Evaluasi pada Tiga Skenario	94
Tabel 4. 11 Hasil <i>Confusion Matrix</i> Tiga Skenario	95
Tabel 4. 12 Hasil Analisis Kesesuaian Label dengan Teks	99
Tabel 4. 13 Perbandingan Label Rating dengan Penilaian Berdasarkan Teks	99
Tabel 4. 14 Aturan Pelabelan Berbasis <i>Keyword Matching</i>	101
Tabel 4. 15 Hasil Evaluasi Model dengan Pelabelan Berbasis Teks	102
Tabel 4. 16 Hasil Evaluasi Model Berdasarkan <i>Classification Report</i> dan <i>Confusion Matrix</i>	103