

DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN KARYA	iii
LEMBAR PERSEMBAHAN	iv
ABSTRAK	vi
ABSTRACT.....	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	4
1.5 Manfaat.....	5
1.6 Kerangka Pemikiran	6
1.7 Sistematika Penulisan.....	7
BAB II KAJIAN LITERATUR.....	9
2.1 Tinjauan Pustaka	9
2.2 Landasan Teori	18
2.2.1 Analisis Sentimen	18
2.2.2 Pemrosesan Bahasa Alami (<i>Natural Language Processing-NLP</i>)	18
2.2.3 Model BERT dan IndoBERT	19

2.2.4	Evaluasi Model dalam Analisis Sentimen	21
BAB III METODOLOGI PENELITIAN		23
3.1	Pemahaman Bisnis (<i>Business Understanding</i>).....	23
3.2	Pemahaman Data (<i>Data Understanding</i>)	23
3.3	Persiapan Data (<i>Data preparation</i>)	24
3.4	Pembangunan Model (<i>Modeling</i>)	25
3.5	Evaluasi Model	27
3.6	Prediksi Data Non Label	28
3.7	Visualisasi Hasil	28
3.8	Bahan dan Peralatan Yang Digunakan	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Hasil <i>Business Understanding</i>	30
4.2	Hasil <i>Data Understanding</i>	31
4.2.1	<i>Load Dataset</i>	32
4.2.2	Deskripsi Statistik	32
4.2.3	Distribusi Kelas Sentimen	33
4.2.4	Distribusi Panjang Komentar	34
4.3	Hasil <i>Data Preparation</i>	35
4.3.1	Menormalkan Kolom Dataset.....	36
4.3.2	Pembersihan Teks (<i>Text Cleaning</i>).....	36
4.3.3	Jumlah total data berlabel	37
4.3.4	Pembagian Dataset.....	37
4.3.5	Tokenisasi dengan IndoBERT	38
4.4	Modelling	39
4.4.1	Arsitektur Model.....	40
4.4.2	Proses Pelatihan (<i>Training</i>)	41

4.4.3	<i>Hyperparameter Tuning</i>	42
4.4.4	Proses Validasi.....	53
4.5	<i>Evaluation</i>	53
4.6	<i>Predicting Unlabeled Data</i>	54
4.7	Visualisasi Hasil	55
BAB V PENUTUP.....		62
5.1	Kesimpulan.....	62
5.2	Saran	63
DAFTAR PUSTAKA		64
LAMPIRAN.....		67
RIWAYAT HIDUP		82



DAFTAR GAMBAR

Gambar 1. 1 Kerangka Pemikiran	6
Gambar 2. 1 Arsitektur BERT	20
Gambar 3. 1 Flowchart Metodologi Penelitian	26
Gambar 4. 1 <i>Load Dataset</i>	32
Gambar 4. 2 Deskripsi Statistik	33
Gambar 4. 3 Visualisasi Distribusi Kelas Sentimen	34
Gambar 4. 4 Distribusi Panjang Komentar	35
Gambar 4. 5 menormalisasi kolom dataset	36
Gambar 4. 6 <i>Text Cleaning</i>	36
Gambar 4. 7 Jumlah data berlabel.....	37
Gambar 4. 8 Skenario 1 80%-10%-10%	37
Gambar 4. 9 Skenario 2 70%-20%-10%	38
Gambar 4. 10 Tokenisasi dengan IndoBERT	39
Gambar 4. 11 Arsitektur Model	40
Gambar 4. 12 Proses Pelatihan.....	41
Gambar 4. 13 <i>Splitting</i> 80%-10%-10% LR 1e-4	43
Gambar 4. 14 <i>Confusion Matrix</i> 80%-10%-10% LR 1e-4	44
Gambar 4. 15 <i>Splitting</i> 80%-10%-10% LR 1e-5	45
Gambar 4. 16 <i>Confusion Matrix</i> 80% - 20% - 10% LR 1e-5	45
Gambar 4. 17 <i>Splitting</i> 80% - 10% - 10% LR 5e-5	46
Gambar 4. 18 <i>Confusion Matrix</i> 80% - 10% - 10% LR 5e-5	47
Gambar 4. 19 <i>Splitting</i> 70%-20%-10% LR 1e-4	48
Gambar 4. 20 <i>Confusion Matrix</i> 70%-20%-10% LR 1e-4	49
Gambar 4. 21 <i>Splitting</i> 70%-20%-10% LR 1e-5	50
Gambar 4. 22 <i>Confusion Matrix</i> 70%-20%-10% LR 1e-5	50
Gambar 4. 23 <i>Splitting</i> 70%-10%-10% LR 5e-5	51
Gambar 4. 24 <i>Confusion Matrix</i> 70%-20%-10% LR 5e-5	52
Gambar 4. 25 Kode Prediksi Non-Label.....	55
Gambar 4. 26 Hasil Prediksi	56
Gambar 4. 27 <i>Stopword remover</i> dan <i>pre-processing</i>	57

Gambar 4. 28 <i>WordCloud</i> Positif.....	59
Gambar 4. 29 <i>WordCloud</i> Negatif	60
Gambar 4. 30 Hasil Prediksi	61



DAFTAR TABEL

Tabel 2. 1 <i>State Of The Art</i>	9
Tabel 2. 2 <i>Komponen Utama Confusion Matrix</i>	22

