

DAFTAR ISI

HALAMAN JUDUL SKRIPSI	i
PERNYATAAN KEASLIAN KARYA SKRIPSI	iii
PERSETUJUAN DOSEN PEMBIMBING SKRIPSI.....	iv
PENGESAHAN TIM PENGUJI SKRIPSI	v
ABSTRAK	vii
ABSTRACT.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian	3
1.5.1 Bagi Akademik.....	3
1.5.2 Bagi Masyarakat.....	3
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 Analisis Sentimen	6
2.2 Text Mining.....	9
2.3 <i>BERT</i>	10
2.4 <i>IndoBERT</i>	11
2.5 Indonesia Emas 2045	13
2.6 TikTok.....	14
2.7 Penelitian Terdahulu	16
BAB III METODOLOGI	20
3.1 Flowchart Penelitian.....	20

3.2	Pengumpulan Data	21
3.3	<i>Preprocessing Dataset</i>	25
3.3.1	Penghapusan data duplikat dan baris yang kosong	25
3.3.2	Pembersihan <i>Special Character</i> , URL, Tanda Baca, <i>Emoji</i> , Spasi Berlebih, dan Baris yang Kosong.	26
3.3.3	Case Folding	27
3.3.4	Pengecekan frekuensi kemunculan kata.....	28
3.3.5	Normalisasi	29
3.3.6	Menghapus baris yang memiliki satu kata.....	30
3.3.7	Menghapus kembali spasi yang berlebihan.....	31
3.4	Pelabelan <i>Dataset</i>	32
3.5	Melakukan <i>Balancing</i> Terhadap <i>Dataset</i>	35
3.6	Membagi <i>Dataset</i>	37
3.7	<i>Training, Validation</i> dan <i>Testing</i>	37
3.8	<i>Push Model</i>	40
BAB IV HASIL DAN PEMBAHASAN		41
4.1	<i>Confusion Matrix</i>	41
4.2	Evaluasi Metriks.....	45
4.3	Perbandingan Model	47
BAB V KESIMPULAN DAN SARAN		50
5.1	Kesimpulan	50
5.2	Saran.....	51
DAFTAR PUSTAKA.....		52
LAMPIRAN A : SOURCE CODE		A-1

DAFTAR GAMBAR

Gambar 2.1 Statistik dari tasks.....	12
Gambar 2.2 Perbandingan performa IndoBERT dengan model lain.....	13
Gambar 2.3 Contoh Komentar Positif.....	15
Gambar 2.4 Contoh Komentar Negatif.....	15
Gambar 2.5 Contoh Komentar Netral.....	15
Gambar 3.1 Flowchart tahapan pada penelitian.....	20
Gambar 3.2 File program yang digunakan untuk scrapping komentar TikTok....	21
Gambar 3.3 Langkah – langkah scrapping komentar TikTok.....	21
Gambar 3.4 Dataset hasil Scrapping	24
Gambar 3.5 Perhitungan sebelum dan sesudah penghapusan data duplikat dan baris yang kosong pada kolom "Comment Text"	25
Gambar 3.6 Dataset sebelum dilakukan pembersihan karakter yang tidak diperlukan dan baris yang kosong.....	26
Gambar 3.7 Dataset sesudah dilakukan pembersihan karakter yang tidak diperlukan dan baris yang kosong.....	27
Gambar 3.8 Dataset setelah <i>Case Folding</i>	28
Gambar 3.9 Output pengecekan kata yang berulang	29
Gambar 3.10 Dataset hasil normalisasi.....	30
Gambar 3.11 Dataset setelah penghapusan baris yang hanya memiliki satu kata	31
Gambar 3.12 Code untuk menghilangkan spasi yang berlebih.....	31

Gambar 3.13 Visualisasi pelabelan secara otomatis	33
Gambar 3.14 Visualisasi pelabelan Secara manual.....	33
Gambar 3.15 Perbandingan hasil evaluasi pelabelan secara otomatis dan manual.	34
Gambar 3.16 Pelabelan dataset secara otomatis	35
Gambar 3.17 Jumlah masing - masing sentimen sebelum balancing.....	35
Gambar 3.18 Teknik Undersampling.....	36
Gambar 3.19 Jumlah masing - masing sentimen setelah dilakukan balancing	36
Gambar 3.20 Hasil dari training dan validation	38
Gambar 3.21 Hasil testing.....	40
Gambar 3.22 Model Indonesia-Emas-2045-Sentiment-Classification	40
Gambar 4.1 <i>Confusion matrix validation dataset imbalance</i>	41
Gambar 4.2 <i>Confusion matrix validation dataset balance</i>	42
Gambar 4.3 <i>Confusion matrix testing dataset imbalance</i>	43
Gambar 4.4 <i>Confusion matrix testing dataset balance</i>	44
Gambar 4.5 Evaluasi performa <i>validation dataset imbalance</i>	45
Gambar 4.6 Evaluasi performa <i>validation dataset balance</i>	45
Gambar 4.7 Evaluasi performa <i>testing dataset imbalance</i>	46
Gambar 4.8 Evaluasi performa <i>testing dataset balance</i>	46

DAFTAR TABEL

Tabel 2. 1 Rangkuman penelitian terdahulu 18

Tabel 4. 1 Perbandingan Model 48

