

DAFTAR ISI

COVER	i
HALAMAN PERNYATAAN	ii
HALAMAN PERSETUJUAN TESIS	iii
HALAMAN PENGESAHAN TESIS	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
ABSTRAK	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
BAB I_PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
BAB II_TINJAUAN PUSTAKA	7
2.1 Penelitian Terkait	7
2.2 Landasan Teori	10
2.2.1 Kekerasan Verbal	10
2.2.2 NLP (Natural Language Preprocessing)	10
2.2.3 Analisis Sentimen	11
2.2.4 BERT	12
2.2.5 Deep Learning	21
2.2.6 Bi-LSTM	23
2.2.7 Semi-Supervised Learning	24
2.2.8 INDOBERT	25
BAB III_METODOLOGI PENELITIAN	28
3.1 Arsitektur Umum	28
3.2 Data Crawling	29
3.3 Labeling Data	31

3.3.1	Labeling Data Manual.....	31
3.3.2	Labeling Data Otomatis	31
3.4	Pre-Processing Data	33
3.4.1	Case Folding	33
3.4.2	Data Cleaning	34
3.4.3	Tokenisasi	34
3.4.4	Stopwords Removal	35
3.4.5	Stemming	35
3.4.6	Normalisasi	36
3.5	Implementasi INDOBERT	37
3.6	Evaluasi	37
BAB IV PEMBAHASAN.....		40
4.1	Implementasi Sistem	40
4.1.1	Spesifikasi Perangkat Keras	40
4.1.2	Spesifikasi Perangkat Lunak	41
4.2	Implementasi Twitter Crawling.....	41
4.3	Preparing Data	45
4.3.1	Shaping Dataset.....	45
4.3.2	Cleaning Dataset	48
4.3.3	<i>Data Sampling and Preparation</i>	51
4.4	Labeling Data	55
4.4.1	Labeling Data Manual.....	56
4.4.2	Labeling Data Otomatis	59
4.5	Implementasi INDOBERT	67
4.6	Evaluasi	73
4.6.1	Labeling Supervised Learning	73
4.6.2	Implementasi INDOBERT.....	74
BAB V KESIMPULAN DAN SARAN.....		77
5.1	Kesimpulan.....	77
5.2	Saran	79
DAFTAR PUSTAKA		80
LAMPIRAN.....		84

DAFTAR GAMBAR

Gambar 2.1 Encoder (kiri) dan Decoder (kanan) (Vaswani et al., 2016)	13
Gambar 2.2 Proses pada <i>Self-attention Layer</i>	14
Gambar 2. 3 Proses pada <i>Encoder</i>	15
Gambar 2.4 Perbedaan Ukuran BERTBASE dan BERTLARGE	16
Gambar 2.5 Arsitektur BERT	17
Gambar 2. 6 Perbedaan Arsitektur BERT Dengan OpenAI GPT dan ELMo.....	17
Gambar 2.7 Proses <i>Masked Language Modelling</i>	19
Gambar 2.8 Proses <i>Pre-training</i> pada BERT.....	19
Gambar 2.9 Representasi Input pada BERT (Devlin et al., 2019).....	20
Gambar 2.10 Prosedur <i>Pre-training</i> dan <i>Fine-tuning</i> (Devlin et al., 2019)	21
Gambar 2. 11 Ilustrasi <i>Fine-tuning</i> pada Tugas dengan Single Sentence (Devlin et al., 2019).....	21
Gambar 3. 1 Langkah Penelitian	28
Gambar 4.1 Hasil crawling twint format file .csv	42
Gambar 4.2 Hasil crawling twint format file .txt	42
Gambar 4.3 Proses crawling tweet-harvest.....	43
Gambar 4.4 Hasil crawling tweet-harvest.....	44
Gambar 4.5 Struktur asli dataset	46
Gambar 4.6 Struktur dataset setelah di shape	47
Gambar 4.7 <i>Summary</i> data setelah tahap <i>shaping</i> dataset.....	48
Gambar 4.8 Hasil sebelum <i>Cleaning</i> data.....	49
Gambar 4.9 Hasil sebelum <i>Cleaning</i> data.....	50
Gambar 4.10 <i>Summary</i> data setelah <i>cleaning</i> data.....	50
Gambar 4.11 Perhitungan 10% data untuk labeling manual.....	51
Gambar 4.12 Hasil perhitungan porsi setiap database	52
Gambar 4.13 Chart porsi setiap database	53
Gambar 4.14 Data sampling.....	54
Gambar 4.15 Sisa data dari Langkah Data Sampling	55

Gambar 4.16 Data terlabel annotator 1	57
Gambar 4.17 Data terlabel annotator 2	57
Gambar 4.18 Data terlabel annotator 3	58
Gambar 4.19 Komparasi Data Labeled.....	58
Gambar 4.20 Proses Preprocessing data	60
Gambar 4.21 Hasil Preprocessing data	61
Gambar 4.22 Distribusi data berdasarkan Sentimen	62
Gambar 4.23 Percobaan Labeling pertama	64
Gambar 4.24 Percobaan Labeling kedua	65
Gambar 4.25 Distribusi label sentimen.....	66
Gambar 4.26 Hasil akurasi percobaan labeling ketiga.....	66
Gambar 4.27 Hasil akurasi 5 Epoch.....	68
Gambar 4.28 Hasil akurasi 10 Epoch.....	68
Gambar 4.29 Hasil akurasi 15 Epoch.....	69
Gambar 4.30 Kurva performa training model.....	70
Gambar 4.31 Diagram confusion matrix	70
Gambar 4. 32 Diagram confusion matrix Percobaan ke 4.....	71
Gambar 4. 33 Diagram confusion matrix Percobaan ke 5.....	72
Gambar 4.34 Hasil akurasi model BERT.....	75
Gambar 4. 35 Diagram confusion matrix BERT	75

DAFTAR TABEL

Tabel 1 Penelitian Terkait	7
Tabel 2 Kategori Kekerasan Verbal	30
Tabel 3 Contoh Dataset.....	32
Tabel 4 Perbandingan hasil Case Folding.....	34
Tabel 5 Perbandingan hasil Data Cleaning	34
Tabel 6 Perbandingan hasil Tokenisasi.....	35
Tabel 7 Perbandingan hasil Stopwords Removal.....	35
Tabel 8 Perbandingan hasil Stemming.....	35
Tabel 9 Perbandingan hasil Normalisasi	36
Tabel 10 Confussion Matrix.....	38