

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN	iv
HALAMAN PRAKATA	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	6
BAB III LANDASAN TEORI.....	12
3.1 <i>Authorship Verification</i>	12
3.2 <i>Emoji Representation</i>	13
3.2.1 <i>Emoji2Text</i> Indonesia.....	13
3.2.2 <i>Emoji2vec</i>	13
3.3 <i>Text Representation</i>	14
3.2.1 BERT	14

3.2.2	IndoBERT	18
3.4	RNN.....	19
3.4.1	LSTM.....	19
3.4.2	Bi-LSTM.....	21
3.5	Metrik Evaluasi	22
3.5.1	<i>Confusion Matrix</i>	22
3.5.2	ROC-AUC.....	23
BAB IV METODOLOGI PENELITIAN		25
4.1	Deskripsi Penelitian.....	25
4.2	Tahapan Penelitian	26
4.3	<i>Dataset</i>	28
4.3.1	Persiapan dan <i>Preprocessing Dataset</i>	29
4.3.2	<i>Pairing</i> dan <i>Labelling Dataset</i>	34
4.4	Perancangan Model	35
4.5	Pengujian dan Evaluasi	43
BAB V IMPLEMENTASI		46
5.1.	Alat dan Bahan	46
5.2.	Hirarki Model Implementasi Kode.....	46
5.3.	Modul Kode.....	47
5.3.1	Modul <i>Library</i> dan <i>Loading Data</i>	47
5.3.2	Modul <i>Pairing Data</i> dan <i>Labelling Data</i>	50
5.3.3	Modul Metode Representasi <i>Tweet</i>	51
5.3.4	Modul Fungsi	54
5.3.5	Modul Pengujian	57
BAB VI HASIL DAN PEMBAHASAN		69

6.1.	<i>Hyperparameter</i>	69
6.2.	Hasil Pengujian.....	71
6.2.1.	Hasil Pengujian 1	72
6.2.2.	Hasil Pengujian 2	76
6.2.3.	Hasil Pengujian 3	79
6.3.	Pembahasan Keseluruhan Hasil Penelitian	83
BAB VII KESIMPULAN DAN SARAN.....		86
DAFTAR PUSTAKA		88

DAFTAR GAMBAR

Gambar 3.1 Contoh <i>Dataset Emoji2text</i>	13
Gambar 3.2 <i>Pre-Training</i> dan <i>Fine-Tuning</i> BERT (Devlin et al., 2019)	15
Gambar 3.3 Proses pada MLM (Alammar, 2018).....	17
Gambar 3.4 Representasi <i>Input</i> pada BERT (Devlin et al., 2019).....	17
Gambar 3.5 Sel Tunggal LSTM dan Bi-LSTM (Shahid et al., 2020).....	22
Gambar 4.1 Tahapan Penelitian	26
Gambar 4.2 Perancangan Model.....	35
Gambar 4.3 Skenario Representasi dengan <i>Emoji2text</i> + IndoBERT	38
Gambar 4.4 Skenario Representasi dengan <i>Emoji2vec</i> + IndoBERT	40
Gambar 5.1 Instalasi dan Impor Pustaka	48
Gambar 5.2 Pengunduhan <i>Dataset</i>	48
Gambar 5.3 Penghapusan Kolom yang Tidak Digunakan	49
Gambar 5.4 Penghapusan Data yang Tidak Diperlukan	49
Gambar 5.5 Pemisahan Teks dan Emoji dari <i>Dataset</i>	50
Gambar 5.6 Penghapusan Konten yang Terduplikat.....	50
Gambar 5.7 Pemilihan <i>Subset Dataset</i>	50
Gambar 5.8 <i>Pairing</i> Data dan <i>Labelling</i> Data	51
Gambar 5.9 Proses Impor <i>Emoji2text</i>	52
Gambar 5.10 Contoh Pengubahan Makna Emoji.....	52
Gambar 5.11 Proses Impor Model <i>Emoji2vec</i>	53
Gambar 5.12 Impor IndoBERT	54
Gambar 5.13 Fungsi <i>Preprocessing</i> untuk <i>Dataset</i>	54
Gambar 5.14 Fungsi Pembagian <i>Dataset</i>	55
Gambar 5.15 Fungsi Pembuatan Model.....	56
Gambar 5.16 Fungsi Mendapatkan <i>Callback</i>	57
Gambar 5.17 Pemilihan dan <i>Preprocessing Dataset</i> pada Pengujian 1	57
Gambar 5.18 Representasi Teks dengan IndoBERT pada Pengujian 1	57
Gambar 5.19 Pembuatan Data Pasangan pada Pengujian 1	58

Gambar 5.20 Pembagian <i>Dataset</i> pada Pengujian 1	58
Gambar 5.21 Klasifikasi pada Pengujian 1	59
Gambar 5.22 Pengujian pada Pengujian 1	59
Gambar 5.23 Evaluasi pada Pengujian 1	60
Gambar 5.24 Pemilihan dan <i>Preprocessing Dataset</i> serta Representasi Emoji dengan <i>Emoji2text</i> pada Pengujian 2	61
Gambar 5.25 Representasi Teks dengan IndoBERT pada Pengujian 2	61
Gambar 5.26 Pembuatan Data Pasangan pada Pengujian 2	62
Gambar 5.27 Pembagian <i>Dataset</i> pada Pengujian 2	62
Gambar 5.28 Klasifikasi pada Pengujian 2	63
Gambar 5.29 Pengujian pada Pengujian 2	63
Gambar 5.30 Evaluasi pada Pengujian 2	64
Gambar 5.31 Pemilihan dan <i>Preprocessing Dataset</i> pada Pengujian 3	64
Gambar 5.32 Representasi Emoji dengan <i>Emoji2vec</i> pada Pengujian 3	65
Gambar 5.33 Representasi Teks dengan IndoBERT pada Pengujian 3	65
Gambar 5.34 Pembuatan Data Pasangan pada Pengujian 3	66
Gambar 5.35 Pembagian <i>Dataset</i> pada Pengujian 3	66
Gambar 5.36 Klasifikasi pada Pengujian 3	67
Gambar 5.37 Pengujian pada Pengujian 3	67
Gambar 5.38 Evaluasi pada Pengujian 3	68
Gambar 6.1 Grafik <i>Training Loss</i> dan <i>Validation Loss</i> pada Pengujian 1	73
Gambar 6.2 Hasil Kecepatan Prediksi Pengujian 1	73
Gambar 6.3 Hasil Evaluasi Pengujian 1	74
Gambar 6.4 <i>Classification Report</i> Pengujian 1	75
Gambar 6.5 Kurva ROC Pengujian 1	75
Gambar 6.6 Grafik <i>Training Loss</i> dan <i>Validation Loss</i> pada Pengujian 2	76
Gambar 6.7 Hasil Kecepatan Prediksi Pengujian 2	77
Gambar 6.8 Hasil Evaluasi Pengujian 2	77
Gambar 6.9 <i>Classification Report</i> Pengujian 2	78
Gambar 6.10 Kurva ROC Pengujian 2	79
Gambar 6.11 Grafik <i>Training Loss</i> dan <i>Validation Loss</i> pada Pengujian 3	80

Gambar 6.12 Hasil Kecepatan Prediksi Pengujian 3	80
Gambar 6.13 Hasil Evaluasi Pengujian 3	81
Gambar 6.14 <i>Classification Report</i> Pengujian 3	82
Gambar 6.15 Kurva ROC Pengujian 3	82

DAFTAR TABEL

Tabel 2.1 Perbandingan Pustaka Terhadap Penelitian yang Dilakukan.....	10
Tabel 3.1 Tabel <i>Confusion Matrix</i> (Ayman et al., 2020)	23
Tabel 4.1 Contoh <i>Dataset</i> Asli.....	29
Tabel 4.2 Contoh <i>Dataset</i> Setelah Penghapusan Kolom yang Tidak Relevan	30
Tabel 4.3 Contoh <i>Dataset</i> Setelah Penghapusan Data yang Kurang Informatif...	31
Tabel 4.4 Contoh <i>Dataset</i> Setelah Pemisahan Teks dan Emoji	31
Tabel 4.5 Contoh <i>Dataset</i> Setelah Penghapusan Konten yang Terduplikasi	32
Tabel 4.6 Contoh <i>Preprocessing Dataset</i>	33
Tabel 4.7 Contoh Hasil <i>Pairing</i> dan <i>Labelling Dataset</i>	34
Tabel 6.1 <i>Hyperparameter</i> yang Digunakan.....	69
Tabel 6.2 Tabel Perbandingan Hasil Pengujian	83