

DAFTAR ISI

DAFTAR ISI.....	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vii
INTISARI.....	vi
ABSTRACT	1
BAB I PENDAHULUAN	2
1.1 Latar Belakang Masalah	2
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	4
1.5 Manfaat.....	4
BAB II TINJAUAN PUSTAKA.....	5
BAB III LANDASAN TEORI.....	10
3.1 Data Mining	10
3.2 Text Mining	11
3.3 Natural Language Processing	11
3.4 Deep Learning	12
3.5 Analisis Sentimen.....	12
3.6 Preprocessing Data	13
3.7 Word Embedding	13
3.8 Word2Vec CBOW (Continuous Bag-of-Word).....	16
3.9 Bidirectional Encoder Representations from Transformers (BERT)	17
3.10 Long Short-Term Memory (LSTM)	22
3.11 Batch Size dan Epoch	24
3.12 Confusion Matrix.....	24
BAB IV METODOLOGI PENELITIAN	26

4.1	Alat & Bahan	26
4.2	Pengumpulan Data	26
4.3	<i>Annotation</i>	27
4.4	Prosedur Kerja	27
BAB V IMPLEMENTASI.....		33
5.1	Spesifikasi Perangkat	33
5.2	Pengumpulan Data	33
5.3	Dataset	34
5.4	Implementasi Word2Vec	34
5.4.1	Teks Masukan	35
5.4.2	Preprocessing Data	35
5.4.2.1	Data Cleaning	35
5.4.2.2	<i>Case Folding</i>	36
5.4.2.3	<i>Stopword Removal</i>	36
5.4.2.4	<i>Tokenizing</i>	37
5.4.3	Pembagian Dataset.....	38
5.4.4	<i>Generate Model Word2Vec</i>	39
5.4.5	<i>Training and Validation Accuracy</i>	42
5.4.6	Test Prediksi Model	42
5.4.7	Evaluasi.....	43
5.5	Implementasi BERT	44
5.5.1	<i>Preparation Library Transformers</i>	44
5.5.2	<i>Repository IndoNLU</i>	45
5.5.3	<i>Preparation Data Pre-Train</i>	46
5.5.4	Data Teks Masukan	46
5.5.5	<i>Preprocessing Data</i>	47
5.5.5.1	<i>Case Folding</i>	47
5.5.5.2	<i>Punctuation</i>	47
5.5.5.3	<i>Tokenizing</i>	48
5.5.6	Pembagian Dataset.....	48

5.5.7	<i>Testing Model Data Pre-Train</i>	49
5.5.8	<i>Fine Tuning Model</i>	49
5.5.9	Evaluasi Validasi	51
5.5.10	Evaluasi <i>Testing</i>	52
5.5.11	<i>Testing Fine Tuning Model</i>	53
5.5.12	Evaluasi Akurasi	53
BAB VI HASIL DAN PEMBAHASAN		55
6.1	Hasil Sentimen Analisis Tanpa Ekstraksi Fitur.....	55
6.1.1	Evaluasi Hasil	56
6.2	Hasil Sentimen Analisis Menggunakan <i>Word2Vec</i>	59
6.2.1	Evaluasi Hasil	60
6.3	Hasil Sentimen Analisis Menggunakan <i>BERT</i>	62
6.3.1	Evaluasi Hasil	63
6.4	Validasi Hasil Metode <i>Word2Vec</i> dan <i>BERT</i>	65
BAB VII KESIMPULAN DAN SARAN		67
7.1	Kesimpulan.....	67
7.2	Saran	67
DAFTAR PUSTAKA		69

DAFTAR GAMBAR

Gambar 3.1 Arsitektur CBOW	17
Gambar 3.2 Visualisasi sederhana kinerja encoder dan decoder	18
Gambar 3.3 Ilustrasi Proses Fine-Tuning pada BERT	20
Gambar 3.4 Representasi input pada BERT	20
Gambar 3.5 Visualisasi proses fine-tuning pada BERT.....	22
Gambar 3.6 Pengulang RNN yang berisi satu layer	23
Gambar 3.7 Pengulang dalam LSTM berisi empat layer	23
Gambar 4.1 Rancangan Sistem	27
Gambar 4.2 Preprocessing	28
Gambar 5.1 Gambar Dataset	35
Gambar 5.2 Source Code Pembacaan Dataset.....	36
Gambar 5.3 Source Code Data Cleaning	37
Gambar 5.4 Source Code Case Folding	37
Gambar 5.5 Source Code Stopword Removal	38
Gambar 5.6 Source Code Library Tokenizing	38
Gambar 5.7 Source Code Tokenizing	39
Gambar 5.8 Source Code Membagi Data Latih dan Data Uji	39
Gambar 5.9 Source Code Menggunakan Korpus	40
Gambar 5.10 Source Code Case Folding	40
Gambar 5.11 Source Code Case Folding	41
Gambar 5.12 Source Code Similarity Kata	41
Gambar 5.13 Source Code Similarity Weight	41
Gambar 5.14 Source Code Model Word2Vec	42
Gambar 5.15 Source Code Proses Training	42
Gambar 5.16 Source Code Training and Validation Accuracy	43
Gambar 5.17 Source Code Test Prediksi Model	43
Gambar 5.18 Source Code Evaluasi Model	44
Gambar 5.19 Source Code Confusion Matrix	44
Gambar 5.20 Source Code Evaluasi Perfoma	45

Gambar 5.21 Source Code Library Transformers	46
Gambar 5.22 Source Code Repository IndoNLU	46
Gambar 5.23 Source Code Library dan Function	46
Gambar 5.24 Source Code Data Pre-Train	47
Gambar 5.25 Source Code Load dan Visualisasi Dataset.....	47
Gambar 5.26 Source Code Case Folding	48
Gambar 5.27 Source Code Punctuation	49
Gambar 5.28 Source Code Tokenizing	49
Gambar 5.29 Source Code Pembagian Dataset	49
Gambar 5.30 Source Code Testing Model on Sample Sentences Positive	50
Gambar 5.31 Source Code Testing Model on Sample Sentences Negative	50
Gambar 5.32 Source Code Optimizer	51
Gambar 5.33 Source Code Training Model	51
Gambar 5.34 Source Code Evaluasi Validasi.....	52
Gambar 5.35 Source Code Visualisasi Training dan Evaluasi	53
Gambar 5.36 Source Code Testing Model	53
Gambar 5.37 Source Code Test Model on Sample Sentences Positive	54
Gambar 5.38 Source Code Test Model on Sample Sentences Negative	54
Gambar 5.39 Source Code Memanggil dan Menyalin Label Pred	54
Gambar 5.40 Source Code Evaluasi Confusion Matrix	55
Gambar 6.1 Hasil Training dan Validasi Tanpa Ekstraksi Fitur	57
Gambar 6.2 Hasil Confusion Matrix Tanpa Ekstraksi Fitur	58
Gambar 6.3 Hasil Training dan Validasi Word2Vec	60
Gambar 6.4 Hasil Confusion Matrix Word2Vec	61
Gambar 6.5 Hasil Training dan Validasi BERT	63
Gambar 6.6 Hasil Confusion Matrix BERT	64

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	7
Tabel 3.1 Confusion Matrix	25
Tabel 4.1 Contoh Data Cleaning	29
Tabel 4.2 Contoh Case Folding.....	29
Tabel 4.3 Contoh Stopword	30
Tabel 6.1 Hasil Confusion Matrix Tanpa Ekstraksi Fitur	59
Tabel 6.2 Hasil Confusion Matrix Word2Vec	62
Tabel 6.3 Hasil Confusion Matrix BERT	65
Tabel 6.4 Hasil Evaluasi dari Confusion Matrix	66