

DAFTAR ISI

Halaman

TESIS	ii
HALAMAN PERSETUJUAN	iii
PRAKATA	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
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	4
1.5 Keaslian Penelitian	4
1.6 Metodologi Penelitian	5
1.7 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	11
3.1 PubMed	11
3.2 <i>Text Mining</i>	12
3.3 <i>Feature Extraction</i>	13
3.4 <i>Word2vec</i>	13
3.6 ADAM	17
3.7 <i>RNN</i>	18
3.8 BERT, BioBERT	20
BAB IV METODOLOGI PENELITIAN	25

4.1	Gambaran Umum Penelitian	25
4.2	Alat & Bahan	26
4.2.1	Alat	26
4.2.2	Bahan	27
4.3	Rancangan Data	27
4.3.1	Deskripsi Dataset	27
4.4	<i>Preprocessing</i>	29
4.4.1	<i>Data Preparation</i>	29
4.4.2	<i>Translating</i> (Translate dataset)	36
4.4.3	<i>Remove Unnecessary Symbol</i>	37
4.4.4	<i>Lower Casing</i>	38
4.4.5	<i>Filtering</i>	38
4.4.6	Tokenisasi	39
4.5	<i>Word2Vec</i>	39
4.6	Rancangan Arsitektur Model	40
4.7	Implementasi Sistem	44
4.8	Skema Eksperimen Pengujian	44
4.9	Pengujian	45
BAB V IMPLEMENTASI		49
5.1	Deskripsi Implementasi	49
5.2	Data Implementasi	49
5.3	<i>Preprocessing</i>	50
5.3.1	<i>Data Preparation</i>	50
5.3.2	<i>Translating</i>	54
5.3.3	Noise Removal	55
5.3.4	<i>Lower Casing</i>	55
5.3.5	<i>Filtering</i>	56
5.3.6	Tokenisasi	57
5.4	Pemisahan Data <i>Testing</i> dan <i>Training</i>	58

5.5	<i>Word Embedding</i>	58
5.6	Implementasi Model LSTM	60
5.7	Implementasi Model BioBERT	62
5.8	Implementasi Pengujian dan Evaluasi	65
BAB VI		69
HASIL DAN PEMBAHASAN		69
6.1	Hasil Pengumpulan Data	69
6.2	Hasil <i>Preprocessing</i>	71
6.2.1	Hasil <i>Preprocessing</i> Tokenisasi	72
6.3	<i>Tuning Hyperparameter</i>	72
6.4	<i>Training Model</i>	73
6.4.1	Pengujian <i>Epoch</i>	73
6.4.2	Pengujian <i>Batch</i>	75
6.5	Hasil Pengujian Model dan Evaluasi Model	76
6.5.1	Model LSTM	77
6.5.2	Model <i>Pre-trained</i> BioBERT	86
6.6	Perbandingan Hasil	92
6.6.1	Perbandingan Skor <i>Training Model</i>	92
6.6.2	Perbandingan Nilai Evaluasi	97
BAB VII		100
KESIMPULAN DAN SARAN		100
7.1	Kesimpulan	100
7.2	Saran	100
DAFTAR PUSTAKA		102