

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI.....	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR SIMBOL	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Metode Penelitian	7
1.7 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 <i>Natural Language Processing</i> (NLP)	9
2.2 Karakteristik Bahasa Arab dalam NLP	10
2.3 Teknik <i>Pre-Processing</i>	12
2.3.1 <i>Tokenization</i> (Tokenisasi)	12
2.3.2 <i>Text Cleaning</i> (Pembersihan Teks)	12

2.3.3	<i>Stopword Removal</i> (Penghapusan Kata Umum)	13
2.3.4	<i>Text Normalization</i> (Normalisasi Teks)	13
2.3.5	<i>Lemmatization</i>	14
2.4	<i>Word Embedding</i>	14
2.4.1	Word2Vec	15
2.4.2	FastText	19
2.5	Hyperparameter	20
2.6	<i>N-gram</i> dalam Pemrosesan Teks	21
2.7	<i>Semantic Similarity</i>	23
2.8	<i>Python</i>	24
BAB III METODE PENELITIAN		26
3.1	Dataset	27
3.2	<i>Pre-Processing</i>	27
3.2.1	<i>Tokenization</i> (Tokenisasi)	28
3.2.2	<i>Text Cleaning</i> (Pembersihan Teks)	28
3.2.3	<i>Stopword Removal</i>	29
3.2.4	<i>Text Normalization</i>	29
3.2.5	<i>Lemmatization</i>	29
3.3	Pelatihan Model FastText	30
3.4	Arsitektur Model Skip-gram	32
3.5	Evaluasi <i>Semantic Similarity</i>	35
BAB IV EKSPERIMEN DAN ANALISIS		37
4.1	Validasi Metode	37
4.1.1	Analisis Semantik Kontekstual	37
4.1.2	Analisis Semantik Skalar	48
4.1.3	Visual Progres Kinerja Model	55
4.2	Penerapan Model pada Korpus Utama	58
4.3	<i>Pre-Processing</i>	60

4.4 Training Data	60
BAB V PENUTUP	85
5.1 Kesimpulan	85
5.2 Saran	86
DAFTAR PUSTAKA.....	87
RIWAYAT HIDUP	90
LAMPIRAN.....	92

