

DAFTAR ISI

ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR SIMBOL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Metodologi Penelitian	6
1.6 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 <i>Natural Language Processing</i>	9
2.2 Karakteristik Bahasa Arab dalam NLP	10
2.2.1 Kompleksitas Morfologi	10
2.2.2 Karakteristik Ortografi	11
2.3 Prapemrosesan Data	12
2.4 Korpus Wikipedia Bahasa Arab	13
2.5 <i>Word Embedding</i>	14
2.6 Word2Vec	16
2.7 <i>Pointwise Mutual Information</i> (PMI)	18
2.8 <i>Singular Value Decomposition</i> (SVD)	20
2.8.1 Penurunan Matematis Matriks U , Σ , dan V	21
2.8.2 Contoh Perhitungan <i>Singular Value Decomposition</i> (SVD)	22

2.9	Estimasi Tingkat <i>Noise</i>	26
2.10	<i>Universal Singular Value Thresholding</i> (USVT)	27
2.11	<i>Pairwise Inner Product</i> (PIP) dan <i>PIP Loss</i>	28
2.12	Teori Perturbasi Matriks	32
2.13	<i>Bias–Variance Trade-off</i>	33
2.14	Teorema Penentuan Dimensi Optimal	35
2.15	Simulasi Monte Carlo	39
2.16	Evaluasi Intrinsik <i>Word Embedding</i>	41

BAB III OPTIMISASI DIMENSI *WORD EMBEDDINGS* BAHASA ARAB MENGGUNAKAN METODE PIP LOSS BERBASIS

TEORI PERTURBASI MATRIKS	43
3.1 Alur Penelitian	43
3.2 Pengumpulan dan Pembentukan Korpus	47
3.3 Prapemrosesan Data	52
3.4 Konstruksi Matriks PMI	54
3.5 Estimasi <i>Noise</i> dengan <i>Count-Twice Trick</i>	55
3.6 Dekomposisi Spektral dan Penyaringan USVT	56
3.6.1 Dekomposisi Nilai Singular	56
3.6.2 Penyaringan Nilai Singular Menggunakan USVT	57
3.7 Penentuan Dimensi Optimal Berdasarkan Teorema 3	58
3.8 Validasi Numerik Menggunakan Simulasi Monte Carlo	59
3.9 Evaluasi Kualitas <i>Word Embedding</i>	60

BAB IV HASIL DAN PEMBAHASAN 62

4.1 Konfigurasi dan Pelaksanaan Eksperimen	62
4.2 Statistik Korpus	62
4.3 Estimasi <i>Noise</i> Statistik	63
4.4 Penyaringan Spektral Menggunakan USVT	64
4.5 Penentuan Dimensi Terbaik dengan Simulasi Monte Carlo	65
4.6 Penentuan Dimensi Berdasarkan Teorema 3	75
4.7 <i>Word Similarity</i>	82
4.7.1 Korpus Utama	82
4.7.2 Korpus Geografi	84
4.7.3 Korpus Sejarah	86

BAB V KESIMPULAN DAN SARAN 89

5.1 Kesimpulan	89
--------------------------	----

5.2	Saran	90
DAFTAR PUSTAKA		91
LAMPIRAN		96

