

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
PERNYATAAN KEASLIAN SKRIPSI	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI	iv
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	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Metodologi Penelitian	5
1.6 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Al-Quran dan Kajian Eskatologi	9
2.2 Penambangan Data (<i>Data Mining</i>) dan Penambangan Teks (<i>Text Mining</i>)	10
2.3 <i>Natural Language Processing</i>	11
2.4 Pra-pemrosesan	13
2.5 <i>Word Embedding</i>	14
2.6 <i>Sentence Embedding</i>	15
2.7 <i>Proximity Measure</i>	16
2.8 Metode Klasterisasi	17

2.9	<i>Hierarchical Clustering</i>	19
2.10	<i>Hierarchical Agglomerative Clustering (HAC)</i>	20
2.11	Evaluasi Kinerja Klaster	23
2.12	Penelitian Terkait Klasterisasi Menggunakan HAC	24

BAB III KLASTERISASI TEMA ESKATOLOGI PADA TEKS AL-QUR'AN BERBAHASA INDONESIA MENGGUNAKAN HIERARCHICAL AGGLOMERATIVE CLUSTERING (HAC) BERBASIS WORD2VEC		26
3.1	Pengumpulan Data dan Penyaringan Eskatologi (<i>Rule-Based</i>) .	27
3.2	<i>Pre-processing</i>	31
3.3	Pelatihan Model <i>Word2Vec</i>	36
3.3.1	Arsitektur <i>Skip-gram</i>	37
3.3.2	Konfigurasi Pelatihan	42
3.3.3	Alur Pelatihan Model	42
3.4	Representasi <i>Embedding</i> Tiap Kalimat	43
3.4.1	Probabilitas Kemunculan Kata	44
3.4.2	<i>Smooth Inverse Frequency (SIF)</i>	44
3.4.3	<i>Common Component Removal (CCR)</i>	45
3.4.4	Normalisasi Norma- L_2	46
3.5	Visualisasi Representasi Kalimat dengan <i>Principal Component Analysis (PCA)</i>	47
3.6	Pengukuran Kedekatan Semantik (<i>Cosine Similarity</i>)	48
3.7	<i>Hierarchical Agglomerative Clustering (HAC)</i>	50
3.8	Pelabelan Tema Menggunakan TF-IDF	55
3.9	Evaluasi Kinerja Klaster	57
3.9.1	<i>Silhouette Score</i>	57
3.9.2	<i>Cophenetic Correlation Coefficient (CCC)</i>	58
3.9.3	Kohesi dan <i>Inter-cluster Similarity</i>	59

BAB IV ANALISIS STUDI KASUS PADA DATA TERJEMAHAN AL-QUR'AN BAHASA INDONESIA		60
4.1	Dataset	60
4.1.1	Dataset Ayat Eskatologi	60
4.1.2	Segmentasi Ayat Menjadi Kalimat	62
4.2	Representasi Vektor Kata Menggunakan <i>Word2Vec</i>	64
4.3	Konstruksi <i>Sentence Embedding</i> dengan SIF dan CCR	67
4.4	Pengukuran Kedekatan Semantik dengan <i>Cosine Similarity</i> . .	70

4.5	Penyajian Hasil <i>Clustering</i>	76
4.6	Evaluasi Kualitas Hasil <i>Clustering</i>	81
4.7	Analisis Kualitas Klaster Berdasarkan Kesamaan Semantik	82
4.8	Perbandingan Hasil Eksperimen	84
4.8.1	Skenario dan Rancangan Perbandingan	84
4.8.2	Hasil pada Jumlah Klaster Optimal	86
4.8.3	Hasil pada Jumlah Klaster Seragam	91
BAB V	KESIMPULAN DAN SARAN	96
5.1	Kesimpulan	96
5.2	Saran	98
	DAFTAR PUSTAKA	100
	LAMPIRAN	104

