

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
Halaman Pernyataan	iv
PRAKATA	v
DAFTAR ISI	viii
DAFTAR TABEL	ix
DAFTAR GAMBAR	xi
INTISARI	xii
ABSTRACT	xiii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	4
II TINJAUAN PUSTAKA	6
III LANDASAN TEORI	10
3.1 <i>Word Embedding</i>	10
3.2 Word2Vec	10
3.2.1 <i>Continuous Bag-of-Words (CBOW)</i>	11
3.2.2 <i>Skip Gram</i>	15
3.3 FastText	16
3.4 Analisis Sentimen	17
3.5 <i>Preprocessing</i>	18
3.5.1 Tokenisasi	18
3.5.2 Normalisasi Teks	19
3.5.3 <i>Case Folding</i>	20
3.5.4 Algoritma Jaro-Winkler	20
3.6 Hubungan Semantik	21

3.6.1	Sinonim	21
3.6.2	Antonim	22
3.6.3	Polisemi	22
3.6.4	Hiponim	22
3.6.5	Bentuk Turunan	23
3.7	Evaluasi Intrinsik	23
3.7.1	Kemiripan Kata	24
3.7.2	Analogi Kata	24
3.7.3	Deteksi Sinonim Kata	24
3.7.4	Kategorisasi Kata	25
3.8	Evaluasi Ekstrinsik	25
3.9	<i>Recurrent Neural Network</i>	25
3.9.1	<i>Long Short-Term Memory</i>	27
IV	ANALISIS DAN PERANCANGAN MODEL	30
4.1	Deskripsi Umum Penelitian	30
4.2	Tahapan Penelitian	32
4.3	Pengumpulan Data	32
4.4	Rancangan Penelitian	34
4.4.1	Rancangan <i>Preprocessing</i>	34
4.4.2	Rancangan <i>Word Embedding</i>	36
4.4.3	Rancangan Pengujian	39
V	IMPLEMENTASI	47
5.1	Lingkungan Implementasi	47
5.1.1	Spesifikasi Perangkat Lunak	47
5.1.2	Spesifikasi Perangkat Keras	48
5.2	Implementasi Pengumpulan Data	48
5.3	Implementasi <i>Preprocessing</i>	50
5.3.1	<i>Lower Casing</i>	50
5.3.2	<i>Filtering</i>	50
5.3.3	Menghapus Karakter Berulang	50
5.3.4	Tokenisasi	51
5.4	Normalisasi	51
5.4.1	Kamus Slang	51
5.4.2	Jaro Winkler	52
5.5	Implementasi <i>Query Inventory</i>	54

5.6	Implementasi Pembuatan Model <i>Word Embedding</i>	55
5.7	Implementasi Evaluasi Intrinsik	58
5.8	Implementasi Evaluasi Ekstrinsik	59
5.8.1	Pelabelan Ulasan Hotel	59
5.8.2	<i>Preprocessing</i>	61
5.8.3	<i>Split Dataset</i>	61
5.8.4	Konversi Teks Ulasan	62
5.8.5	Konversi Vektor	63
5.8.6	Pelatihan LSTM	64
5.8.7	Pengujian LSTM	65
VI	HASIL DAN PEMBAHASAN	67
6.1	Hasil <i>Scraping</i> Ulasan Hotel	67
6.2	Hasil Pembuatan Word Embedding	68
6.2.1	<i>Hyperparameter Tuning</i> Word2Vec	68
6.2.2	Hasil <i>Hyperparameter Tuning</i> FastText	70
6.3	Hasil Evaluasi Intrinsik	71
6.4	Hasil Evaluasi Ekstrinsik	80
VII	KESIMPULAN DAN SARAN	82
7.1	Kesimpulan	82
7.2	Saran	83
	DAFTAR PUSTAKA	87

DAFTAR TABEL

2.1	Perbandingan dengan Penelitian Sebelumnya	8
3.1	Kosakata teks	13
4.1	Sebaran <i>Rating</i> pada Dataset Ulasan Hotel	33
4.2	Contoh Hasil Pembangunan Konteks	38
6.1	Contoh Ulasan Hotel Hasil <i>Scraping</i>	67
6.2	Nilai <i>Hyperparameter</i>	68
6.3	Hasil <i>Hyperparameter Tuning</i> Word2Vec	69
6.4	Hasil <i>Hyperparameter Tuning</i> FastText	71
6.5	Contoh <i>Query Inventory</i>	72
6.6	Hasil Evaluasi Intrinsik	73
6.7	Hasil Evaluasi Intrinsik - Domain Spesifik Word2Vec	74
6.8	Hasil Evaluasi Intrinsik - <i>Pretrained</i> Word2Vec	75
6.9	Hasil Evaluasi Intrinsik - <i>Pretrained</i> +Domain Spesifik Word2Vec	76
6.10	Hasil Evaluasi Intrinsik - Domain Spesifik FastText	77
6.11	Hasil Evaluasi Intrinsik - <i>Pretrained</i> FastText	78
6.12	Hasil Evaluasi Intrinsik - <i>Pretrained</i> +Domain Spesifik FastText	79
6.13	Hasil Evaluasi Ekstrinsik	80

DAFTAR GAMBAR

3.1	Arsitektur CBOW dan Skip Gram (Mikolov dkk., 2013)	11
3.2	Ilustrasi model CBOW (Zhang dkk., 2020)	12
3.3	Arsitektur CBOW	13
3.4	Ilustrasi model Skip-Gram (Zhang dkk., 2020)	15
3.5	Arsitektur Skip-Gram	16
3.6	Arsitektur RNN (Neapolitan, 2018)	26
3.7	<i>Memory Block</i> pada LSTM (Zhang dkk., 2020)	28
4.1	Alur penelitian	31
4.2	Pembagian Dataset Ulasan Hotel	33
4.3	Tahap <i>Preprocessing</i>	34
4.4	Tahap Pembuatan Domain Spesifik <i>Word Embedding</i>	36
4.5	Pemodelan Domain Spesifik <i>Word Embedding</i>	37
4.6	Contoh Representasi Kata	39
4.7	Tahap Pembuatan <i>Query Inventory</i>	40
4.8	Evaluasi Intrinsik Domain Spesifik <i>Word Embedding</i>	41
4.9	Evaluasi Intrinsik <i>Pre-trained Word Embedding</i>	42
4.10	Evaluasi Intrinsik <i>Word Embedding</i> Kombinasi	43
4.11	Konversi teks menjadi vektor	44
4.12	Arsitektur LSTM	45
4.13	Alur Evaluasi Ekstrinsik	46
5.1	Kode <i>Scraping</i> Ulasan Hotel	49
5.2	Kode Preprocessing-Filtering	50
5.3	Kode Preprocessing-Menghapus Karakter Berulang	50
5.4	Kode Preprocessing-Tokenisasi	51
5.5	Kode Normalisasi dengan Kamus Slang	52
5.6	Kode Ekstraksi Kata Tidak Baku pada Dataset	53
5.7	Kode Normalisasi dengan Jaro Winkler	53
5.8	Kode POS Tagging	54
5.9	Kode Frekuensi Kemunculan Kata	54
5.10	Kode Pembuatan Model Word2Vec	56
5.11	Kode Pembuatan Model FastText	57
5.12	Kode Pembuatan <i>Word Embedding</i> Kombinasi - Word2vec	58

5.13 Kode Pembuatan <i>Word Embedding</i> Kombinasi - FastText	58
5.14 Kode Evaluasi Intrinsik	60
5.15 Kode Pembobotan Kelas Sentimen	61
5.16 Kode Menghapus <i>stopwords</i>	61
5.17 Kode <i>Split Dataset</i>	62
5.18 Kode Konversi Teks Ulasan	63
5.19 Kode Konversi Vektor	64
5.20 Kode Pelatihan LSTM	65
5.21 Kode Pengujian LSTM	66