

DAFTAR ISI

Halaman

HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN	iv
PRAKATA.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
INTISARI.....	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
BAB II TINJAUAN PUSTAKA.....	4
BAB III LANDASAN TEORI.....	12
3.1 <i>Text Preprocessing</i>	12
3.1.1 <i>Case folding</i>	12
3.1.2 <i>Stemming</i>	12
3.1.3 Menghapus <i>stopwords</i>	12
3.1.4 <i>Tokenization</i>	13
3.2 Doc2Vec	13
3.2.1 Paragraph Vector – Distributed Memory (PV-DM).....	14
3.2.2 Paragraph Vector – Distributed Bag of Words (PV-BDOW).....	16
3.3 Jaccard Coefficient	17
3.4 Cosine Similarity	19
3.5 Parameter Evaluasi	21

BAB IV METODOLOGI PENELITIAN	23
4.1 Deskripsi Umum Penelitian.....	23
4.2 Tahapan Penelitian	24
4.3 Alat dan Bahan	26
4.3.1 Alat.....	26
4.3.2 Bahan.....	26
4.4 Pengumpulan Data	26
4.4.1 Google News.....	26
4.4.2 IndoNLU	30
4.4.3 TaPaCo.....	31
4.5 Rancangan Penelitian	32
4.5.1 Rancangan umum model	32
4.5.2 Rancangan prapemrosesan.....	35
4.5.3 Rancangan Doc2Vec dan Jaccard Coefficient.....	36
4.5.4 Rancangan Doc2Vec dan Cosine Similarity.....	40
4.5.5 Rancangan pengujian dan evaluasi	41
BAB V IMPLEMENTASI.....	45
5.1 Implementasi Pengumpulan Data	45
5.1.1 Pengumpulan dataset Google News.....	45
5.1.2 Pengumpulan <i>dataset</i> IndoNLU.....	47
5.1.3 Pengumpulan <i>dataset</i> TaPaCo	49
5.2 Implementasi Prapemrosesan.....	50
5.3 Implementasi Matriks Perbandingan Label	52
5.4 Implementasi Model Doc2vec	53
5.5 Implementasi Jaccard Coefficient dan Cosine Similarity	55
5.6 Implementasi Evaluasi Model.....	56
BAB VI HASIL DAN PEMBAHASAN	58
6.1 Analisis Hasil Secara Umum	58
6.2 Analisis <i>Dataset</i>	61
6.3 Analisis Parameter	63

6.4 Hasil dengan Dataset Google News	64
6.4.1 Hasil dengan <i>dataset</i> Google News bisnis	67
6.4.2 Hasil dengan <i>dataset</i> Google News teknologi	67
6.4.3 Hasil dengan <i>dataset</i> Google News hiburan	67
6.4.4 Hasil dengan <i>dataset</i> Google News kesehatan.....	68
6.5 Hasil dengan <i>Dataset</i> IndoNLU	68
6.6 Hasil dengan <i>Dataset</i> TaPaCo	69
BAB VII KESIMPULAN DAN SARAN	71
7.1 Kesimpulan	71
7.2 Saran.....	71
DAFTAR PUSTAKA	73
LAMPIRAN	77
Lampiran 1 Hasil Percobaan Dataset Google News	77
Lampiran 2 Hasil Percobaan Dataset Google News - Bisnis.....	80
Lampiran 3 Hasil Percobaan Dataset Google News - Teknologi	83
Lampiran 4 Hasil Percobaan Dataset Google News - Hiburan.....	86
Lampiran 5 Hasil Percobaan Dataset Google News - Kesehatan	89
Lampiran 6 Hasil Percobaan Dataset IndoNLU.....	92
Lampiran 7 Hasil Percobaan Dataset TaPaCo	95

DAFTAR GAMBAR

Gambar 3.1 PV-DM (Le & Mikolov, 2014)	15
Gambar 3.2 PV-DM (Le dan Mikolov, 2014).....	17
Gambar 3.3 Ilustrasi Himpunan	18
Gambar 3.4 Ilustrasi Cosine Similarity	19
Gambar 4.1 Kelompok Berita Google News	27
Gambar 4.2 Isi Kelompok Berita Google News	27
Gambar 4.3 Tampilan Google News pada Topik Tertentu	28
Gambar 4.4 Tampilan Isi Berita.....	28
Gambar 4.5 Tampilan <i>dataset</i> Google News	29
Gambar 4.6 Tampilan <i>dataset</i> IndoNLU pada Huggingface	31
Gambar 4.7 Tampilan <i>dataset</i> TaPaCo	32
Gambar 4.8 Rancangan Model Secara Umum	34
Gambar 4.9 Rancangan Prapemrosesan	35
Gambar 4.10 Rancangan Doc2Vec dan Jaccard Coefficient	38
Gambar 4.11 Rancangan Doc2Vec dan Cosine Similarity	40
Gambar 4.12 Skema Pengujian dan Evaluasi	42
Gambar 4.13 Contoh Pengujian dan Evaluasi.....	42
Gambar 5.1 Pengumpulan <i>Dataset</i> GoogleNews.....	46
Gambar 5.2 <i>Library</i> yang Digunakan	47
Gambar 5.3 Membaca <i>Dataset</i> Google News.....	47
Gambar 5.4 Pengumpulan <i>Dataset</i> IndoNLU	48
Gambar 5.5 Implementasi Baca <i>Dataset</i> IndoNLU	49
Gambar 5.6 Implementasi Mengubah Arrow jadi CSV	49
Gambar 5.7 Pengumpulan <i>Dataset</i> TaPaCo	50
Gambar 5.8 Mengambil Teks Berita atau Kalimat	51
Gambar 5.9 Implementasi Prapemrosesan	51
Gambar 5.10 Hasil Implementasi Prapemrosesan	52
Gambar 5.11 Implementasi Tagged Data.....	52
Gambar 5.12 Implementasi Matriks Perbandingan Label.....	53

Gambar 5.13 Implementasi Model Doc2Vec.....	54
Gambar 5.14 Implementasi Jaccard Coefficient	55
Gambar 5.15 Implementasi Cosine Similarity	56
Gambar 5.16 Implementasi Evaluasi Jaccard Coefficient	57
Gambar 5.17 Implementasi Evaluasi Cosine Similarity	57
Gambar 6.1 Ilustrasi Jaccard Coefficient dengan Cosine Similarity	60
Gambar 6.2 Ilustrasi Lain Jaccard Coefficient dengan Cosine Similarity	60
Gambar 6.3 Vektor Berita Google News Bisnis	61

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Terdahulu	9
Tabel 3.1 Confusion Matrix	21
Tabel 4.1 Cuplikan Matriks Label	30
Tabel 4.2 Contoh Hasil <i>Similarity</i>	33
Tabel 4.3 Contoh Hasil <i>Similarity</i> yang Dibulatkan	33
Tabel 4.4 Contoh Prapemrosesan	36
Tabel 4.5 Tabel Parameter yang Digunakan	39
Tabel 4.6 Contoh Perbandingan Label Model dan Label Asli	43
Tabel 5.1 Tabel Detail <i>Dataset</i> Google News	46
Tabel 6.1 Rata-rata Kinerja	59
Tabel 6.2 Kinerja Terbaik <i>Dataset</i>	63
Tabel 6.3 Parameter Terbaik	64
Tabel 6.4 Ringkasan Hasil <i>Dataset</i> Google News	66
Tabel 6.5 Ringkasan Hasil <i>Dataset</i> IndoNLU	69
Tabel 6.6 Ringkasan Hasil <i>Dataset</i> TaPaCo	70