

DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
PERNYATAAN BEBAS PLAGIASI	iv
PRAKATA.....	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN.....	xvi
INTISARI.....	xvii
ABSTRACT	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah.....	7
1.4 Tujuan	7
1.5 Manfaat	7
BAB II TINJAUAN PUSTAKA.....	8
2.1 <i>Topic Modeling</i>	8
2.2 <i>Word Embedding</i>	10
2.3 <i>Topic Modeling</i> dengan Model <i>Word Embedding</i>	12
BAB III LANDASAN TEORI.....	19
3.1 <i>Big Data Analytic</i>	19
3.2 <i>Natural Language Processing (NLP)</i>	20
3.3 <i>Text Mining</i>	21
3.4 <i>Topic Modeling</i>	22
3.5 <i>Latent Dirichlet Allocation (LDA)</i>	25
3.5.1 <i>Hyperparameter Model</i>	26
3.5.2 Tahapan Pemodelan Topik dengan LDA.....	27
3.6 <i>Hierarchical Dirichlet Process (HDP)</i>	30
3.7 <i>Bag of Words</i>	35

3.8	<i>Word Embedding</i>	36
3.9	Word2Vec	37
3.10	FastText.....	42
3.11	<i>Clustering</i>	43
3.12	K-Means.....	44
3.13	<i>Cosine Similarity</i>	46
3.14	<i>Topic Coherence</i>	47
3.15	<i>Silhouette Score</i>	49
3.16	Visualisasi Data.....	50
BAB IV METODOLOGI PENELITIAN		53
4.1	Alur Penelitian	53
4.2	Alat Penelitian.....	55
4.3	Dataset.....	57
4.4	Persiapan Data.....	60
4.4.1	Penggabungan Data (<i>Data Blending</i>).....	60
4.4.2	Pembersihan Data (<i>Data Cleaning</i>)	60
4.4.3	Normalisasi Bahasa	61
4.4.4	Load Data	61
4.5	<i>Preprocessing</i>	62
4.5.1	<i>Casefolding</i>	62
4.5.2	Tokenisasi	63
4.5.3	<i>Punctual Removal</i>	64
4.5.4	<i>Stopwords Removal</i>	65
4.5.5	<i>Stemming</i>	67
4.5.6	Pembentukan Bigram & Trigram.....	67
4.6	<i>Topic Modeling</i>	68
4.6.1	Pengujian Parameter <i>Topic Modeling</i>	68
4.6.2	<i>Topic Modeling</i> dengan <i>Latent Dirichlet Allocation</i> (LDA).....	69
4.6.3	<i>Topic Modeling</i> dengan <i>Hierarchical Dirichlet Process</i> (HDP)	74
4.7	Pembobotan Kata	79
4.7.1	<i>Bag of Words</i>	79

4.7.2 Word2Vec	80
4.7.3 FastText.....	81
4.8 Clustering Dokumen pada Topik.....	82
4.8.1 LDA+K-Means	82
4.8.2 LDA+Word2Vec+K-Means.....	83
4.8.3 LDA+FastText+K-Means	84
4.8.4 HDP+K-Means.....	84
4.8.5 HDP+Word2Vec+K-Means	85
4.8.6 HDP+FastText+K-Means	86
4.8.7 Clustering Dokumen pada Topik dengan K-Means (<i>similarity</i>).....	87
4.9 Evaluasi Model.....	90
4.9.1 Topic Coherence	90
4.9.2 Silhouette Score.....	93
4.10 Visualisasi Hasil.....	96
4.11 Analisis Perbandingan.....	96
BAB V IMPLEMENTASI.....	98
5.1 Pengumpulan Data	98
5.2 Persiapan Data.....	100
5.3 Preprocessing.....	103
5.4 Topic Modeling	106
5.4.1 Latent Dirichlet Allocation	106
5.4.2 Hierarchical Dirichlet Process	109
5.5 Clustering Dokumen pada Topik.....	112
5.5.1 LDA+K-Means	113
5.5.2 LDA+Word2Vec+K-Means.....	114
5.5.3 LDA+FastText+K-Means	116
5.5.4 HDP+K-Means.....	118
5.5.5 HDP+Word2Vec+K-Means.....	120
5.5.6 HDP+FastText+K-Means	122
BAB VI HASIL DAN PEMBAHASAN	125
6.1 Hasil Pengumpulan Data.....	125

6.2 Hasil Persiapan Data	127
6.3 Hasil <i>Preprocessing</i>	128
6.4 Hasil Pengujian Parameter Pemodelan Topik.....	129
6.4.1 Hasil Pengujian Parameter LDA.....	130
6.4.2 Hasil Pengujian Parameter HDP	133
6.5 Hasil Pemodelan Topik.....	135
6.5.1 Hasil Pemodelan Topik dengan LDA	135
6.5.2 Hasil Pemodelan Topik dengan HDP	138
6.5.3 Perbandingan Hasil Pemodelan Topik Antara LDA dan HDP	142
6.6 Hasil Pembobotan Kata.....	143
6.7 Hasil <i>Clustering</i> Dokumen pada Topik	143
6.7.1 Hasil <i>Clustering</i> LDA+K-Means	144
6.7.2 Hasil <i>Clustering</i> LDA+Word2Vec+K-Means	145
6.7.3 Hasil <i>Clustering</i> LDA+FastText+K-Means.....	145
6.7.4 Hasil <i>Clustering</i> HDP+K-Means	146
6.7.5 Hasil <i>Clustering</i> HDP+Word2Vec+K-Means	147
6.7.6 Hasil <i>Clustering</i> HDP+FastText+K-Means	148
6.7.7 Perbandingan Hasil <i>Clustering</i> Dokumen pada Topik.....	149
6.8 Analisis Hasil Perbandingan.....	151
6.8.1 Analisis Hasil Pemodelan Topik.....	151
6.8.2 Analisis Hasil <i>Clustering</i> Dokumen pada Topik.....	155
BAB VII KESIMPULAN DAN SARAN	158
7.1 Kesimpulan	158
7.2 Saran.....	159
DAFTAR PUSTAKA	160
LAMPIRAN.....	165