

## DAFTAR ISI

HALAMAN DEPAN .....	i
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN MOTIVASI.....	v
PRAKATA.....	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xiii
DAFTAR TABEL.....	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian .....	4
1.7 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
BAB III LANDASAN TEORI.....	11
3.1 Penandaan Otomatis ( <i>Auto-Tagging</i> ) .....	11
3.2 <i>Text Mining</i> .....	11
3.3 <i>Text Pre-Processing</i> .....	12
3.3.1 <i>Case Folding</i> .....	13
3.3.2 <i>Tokenizing</i> (Tokenisasi).....	13
3.3.3 <i>Filtering</i> .....	13
3.3.4 <i>Stemming</i> .....	14
3.4 Pemodelan Topik ( <i>Topic Modelling</i> ).....	15
3.4.1 <i>Latent Dirichlet Allocation</i> (LDA) .....	16

3.4.2	<i>Gibbs Sampling</i> .....	21
3.4.3	<i>Coherence Score</i> .....	23
3.5	Evaluasi Performa Hasil .....	26
3.5.1	Akurasi ( <i>Accuracy</i> ) .....	27
3.5.2	Presisi ( <i>Precision</i> ).....	27
3.5.3	<i>Recall</i> .....	27
3.5.4	<i>F-Measure</i> .....	27
BAB IV ANALISIS DAN PERANCANGAN SISTEM .....		28
4.1	Deskripsi Umum Penelitian.....	28
4.2	Tahap-Tahap Penelitian.....	29
4.2.1	Studi Literatur .....	29
4.2.2	Perancangan Sistem .....	29
4.2.3	Implementasi Sistem.....	31
4.2.4	Pengujian dan Evaluasi Sistem .....	31
4.2.5	Penyelesaian Laporan .....	31
4.3	Akuisisi Data .....	31
4.3.1	Data Latih .....	31
4.3.2	Data Uji.....	33
4.4	Perancangan Sistem.....	34
4.4.1	<i>Pre-Processing</i> (Prapemrosesan).....	36
4.4.2	Membuat <i>Dictionary</i> .....	39
4.4.3	Rancangan Pemodelan Topik dengan LDA .....	40
4.4.4	Rancangan Validasi Model LDA.....	42
4.4.5	Rancangan Implementasi Model LDA sebagai Metode <i>Auto-Tagging</i> .....	43
4.5	Evaluasi Hasil.....	44
BAB V IMPLEMENTASI.....		46
5.1	Lingkungan Implementasi .....	46
5.2	Implementasi <i>Input</i> Teks Artikel Berita .....	46
5.3	Implementasi <i>Pre-Processing</i> (Prapemrosesan) .....	47
5.3.1	<i>Case Folding</i> .....	48

5.3.2	<i>Tokenizing</i> (Tokenisasi).....	48
5.3.3	<i>Filtering</i> .....	49
5.3.4	<i>Stemming</i> .....	50
5.4	Implementasi Membuat <i>Dictionary</i> .....	50
5.5	Implementasi Pemodelan Topik dengan LDA .....	51
5.6	Implementasi Validasi Model LDA .....	52
5.7	Implementasi Model LDA Sebagai <i>Auto-Tagging</i> .....	53
5.8	Implementasi Visualisasi Model LDA .....	56
5.9	Perhitungan Evaluasi Hasil.....	56
BAB VI HASIL IMPLEMENTASI DAN PEMBAHASAN .....		58
6.1	Hasil <i>Pre-Processing</i> (Prapemrosesan) .....	58
6.1.1	Hasil <i>Case Folding</i> .....	58
6.1.2	Hasil <i>Tokenizing</i> (Tokenisasi) .....	59
6.1.3	Hasil <i>Filtering</i> .....	60
6.1.4	Hasil <i>Stemming</i> .....	62
6.2	Hasil Membuat <i>Dictionary</i> .....	63
6.3	LDA.....	66
6.3.1	Penentuan Jumlah <i>Passes</i> dan Jumlah Topik .....	66
6.3.2	Validasi Model dengan <i>Coherence Score</i> .....	66
6.3.4	Hasil Pemodelan Topik LDA .....	68
6.3.5	Visualisasi Topik dengan pyLDAvis.....	70
6.4	Implementasi LDA sebagai <i>Auto-Tagging</i> .....	71
6.5	Evaluasi Hasil.....	74
BAB VII PENUTUP .....		83
7.1	Kesimpulan .....	83
7.2	Saran .....	83
DAFTAR PUSTAKA .....		84
LAMPIRAN.....		87
I.	DAFTAR TEKS ARTIKEL BERITA .....	87
II.	HASIL DARI PROSES TOKENIZING PADA TEKS ARTIKEL BERITA (artikel10) .....	95

i.	<i>Case Folding</i> .....	95
ii.	<i>Tokenizing</i> .....	98
iii.	<i>Filtering</i> .....	101
iv.	<i>Stemming</i> .....	103
III. DAFTAR KATA YANG MELALUI PROSES <i>STEMMING</i> UNTUK “artikel10”.....		
		105
IV. HASIL VISUALISASI DENGAN pyLDAvis UNTUK SETIAP TOPIK.....		
		107
i.	<i>Selected Topic: 2</i> .....	107
ii.	<i>Selected Topic: 3</i> .....	107
iii.	<i>Selected Topic: 4</i> .....	108
iv.	<i>Selected Topic: 5</i> .....	108
v.	<i>Selected Topic: 6</i> .....	109
vi.	<i>Selected Topic: 7</i> .....	109
vii.	<i>Selected Topic: 8</i> .....	110
viii.	<i>Selected Topic: 9</i> .....	110
ix.	<i>Selected Topic: 10</i> .....	111
x.	<i>Selected Topic: 11</i> .....	111
xi.	<i>Selected Topic: 12</i> .....	112
xii.	<i>Selected Topic: 13</i> .....	112
xiii.	<i>Selected Topic: 14</i> .....	113
xiv.	<i>Selected Topic: 15</i> .....	113
xv.	<i>Selected Topic: 16</i> .....	114