

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>HALAMAN PENGESAHAN</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN</b>	<b>iv</b>
<b>DAFTAR ISI</b>	<b>vii</b>
<b>DAFTAR TABEL</b>	<b>ix</b>
<b>DAFTAR GAMBAR</b>	<b>x</b>
<b>INTISARI</b>	<b>xi</b>
<b>ABSTRACT</b>	<b>xii</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	4
1.3 Batasan Masalah . . . . .	4
1.4 Tujuan Penelitian . . . . .	4
1.5 Manfaat Penelitian . . . . .	5
1.6 Metodologi Penelitian . . . . .	5
1.7 Sistematika Penulisan . . . . .	5
<b>II TINJAUAN PUSTAKA</b>	<b>7</b>
<b>III LANDASAN TEORI</b>	<b>10</b>
3.1 <i>Question Answering System</i> . . . . .	10
3.1.1 <i>Automatic Question Answering</i> . . . . .	10
3.1.2 <i>Community Question Answering</i> . . . . .	10
3.2 <i>Question Retrieval</i> . . . . .	11
3.3 <i>Text Mining</i> . . . . .	12
3.4 <i>Web Scraping</i> . . . . .	13
3.5 <i>Natural Language Processing</i> . . . . .	14
3.6 Klasifikasi Statistik . . . . .	14
3.7 Model Statistik . . . . .	15
3.8 Pemodelan Topik . . . . .	16
3.9 <i>Term Frequency - Inverse Document Frequency</i> . . . . .	16
3.9.1 <i>Term Frequency</i> . . . . .	16
3.9.2 <i>Inverse document frequency</i> . . . . .	17

3.9.3	<i>Term Frequency - Inverse document frequency</i>	17
3.10	<i>Latent Dirichlet Allocation</i>	17
3.10.1	<i>Bag-of-Words</i>	18
3.10.2	Asumsi LDA terhadap proses pembuatan dokumen	18
3.10.3	<i>Gibbs Sampling</i>	20
3.11	<i>Non-Negative Matrix Factorization</i>	20
3.12	<i>Jensen-Shannon Divergence</i>	21
3.13	Nilai Koherensi Topik	21
3.14	Pengukuran Akurasi	22
3.14.1	<i>Mean Average Precision</i>	22
3.14.2	<i>Mean Reciprocal Rank</i>	23
<b>IV</b>	<b>ANALISIS DAN PERANCANGAN</b>	<b>25</b>
4.1	Deskripsi Umum Penelitian	25
4.2	Analisis Sistem	25
4.3	Tahapan Penelitian	27
4.4	Tahapan Eksperimen	30
4.4.1	Pengumpulan Data	30
4.4.2	<i>Data Preprocessing</i>	32
4.4.3	Rancangan Pemodelan Topik	35
4.4.4	Rancangan Pengujian	38
<b>V</b>	<b>IMPLEMENTASI</b>	<b>41</b>
5.1	Deskripsi Implementasi	41
5.2	Pengumpulan Data	41
5.2.1	Analisis HTML	42
5.2.2	Implementasi <i>Web scraping</i>	44
5.3	<i>Data Preprocessing</i>	47
5.3.1	<i>Cleansing dan Lowercasing</i>	47
5.3.2	<i>Stopword Removal dan Stemming</i>	48
5.4	Pemodelan Topik	53
5.4.1	Pencarian Banyaknya Topik Optimal	53
5.4.2	Transformasi ke TF dan TFIDF	56
5.4.3	Pembuatan Indeks Kata	57
5.4.4	Aplikasi Pemodelan LDA dan NMF	57
5.4.5	Visualisasi Model	58
5.4.6	Pembuatan Tabel <i>Ranking</i> Per Topik	59

5.5	Pengujian . . . . .	61
5.5.1	Pengambilan Data Tes . . . . .	61
5.5.2	<i>Preprocessing</i> Data Tes . . . . .	61
5.5.3	Pembuatan Rekomendasi . . . . .	62
5.5.4	Evaluasi Kebenaran Rekomendasi . . . . .	64
5.5.5	Penghitungan Tingkat Akurasi . . . . .	65
<b>VI</b>	<b>HASIL DAN PEMBAHASAN</b>	<b>66</b>
6.1	Hasil Pengambilan Data . . . . .	66
6.1.1	Pengambilan Data Judul dan <i>Link</i> . . . . .	66
6.1.2	Pengambilan Data Deskripsi . . . . .	67
6.2	Hasil <i>Preprocessing</i> . . . . .	68
6.2.1	Hasil <i>Cleansing</i> dan <i>Lowercasing</i> . . . . .	68
6.2.2	Hasil <i>Stopword Removal</i> dan <i>Stemming</i> . . . . .	69
6.3	Hasil Proses Pemodelan Topik . . . . .	76
6.3.1	Pencarian Banyaknya Topik Optimal . . . . .	76
6.3.2	Hasil TF, TF-IDF dan Indeks Kata . . . . .	79
6.3.3	Hasil Aplikasi Pemodelan LDA dan NMF . . . . .	79
6.3.4	Hasil Visualisasi Model . . . . .	80
6.3.5	Hasil Tabel <i>Ranking</i> Per Topik . . . . .	81
6.4	Pengetesan Model . . . . .	84
6.4.1	Hasil Evaluasi Kebenaran Rekomendasi . . . . .	92
6.4.2	Hasil Penghitungan Tingkat Akurasi . . . . .	96
<b>VII</b>	<b>KESIMPULAN DAN SARAN</b>	<b>98</b>
7.1	Kesimpulan . . . . .	98
7.2	Saran . . . . .	99
	<b>DAFTAR PUSTAKA</b>	<b>104</b>