

DAFTAR ISI

HALAMAN PENGESAHAN PEMBIMBING	iii
HALAMAN PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR	v
PERNYATAAN BEBAS PLAGIASI.....	vi
ABSTRAK.....	vii
ABSTRACT	viii
KATA PENGANTAR	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian.....	3
1.6 Tahapan Penelitian.....	3
1.6.1 Rencana dan Tahapan Penelitian	3
1.6.2 Metodologi Pengembangan Sistem	4
1.6.3 Pengujian Sistem	4
1.6.4 Pengujian Penelitian	4
1.7 Sistematika Penulisan	4
BAB II TINJAUAN LITERATUR.....	6
2.1 <i>Natural Language Processing (NLP)</i>	6
2.2 <i>Topic Modeling</i>	6
2.3 <i>Text Preprocessing</i>	7
2.3.1 <i>Case Folding</i>	7
2.3.2 <i>Data Cleaning</i>	7
2.3.3 <i>Tokenizing</i>	7
2.4 <i>Latent Dirichlet Allocation (LDA)</i>	8
2.5 Konsep Probabilistik Generatif	8
2.6 Ekstraksi Fitur <i>Bag-of-Word (BoW)</i>	11
2.7 <i>Vocabulary / Dictionary</i>	13
2.8 Evaluasi Model	14
2.9 <i>Coherence Score</i>	15
2.10 <i>Topic Diversity</i>	16
2.11 Visualisasi PyLDAvis (Intertopic Distance Map)	16
2.12 <i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i>	17
2.13 Penelitian Terdahulu	19

BAB III METODOLOGI PENELITIAN.....	23
3.1 Alur Penelitian.....	23
3.2 <i>Business Understanding</i>	24
3.3 <i>Data Understanding</i>	24
3.3.1 Distribusi Kategori Pekerjaan.....	25
3.3.2 Analisis Keseimbangan Dataset	26
3.3.3 Relevansi Penggunaan <i>Topic Modeling</i> LDA.....	26
3.3.4 Proses <i>Web Scraping</i>	27
3.4 <i>Data Preparation</i>	28
3.4.1 <i>Deduplication</i>	29
3.4.2 <i>Data Cleaning</i>	30
3.4.3 <i>Case Folding</i>	31
3.4.4 <i>Text Normalization</i>	31
3.4.5 <i>Compound Terms Handling</i>	32
3.4.6 <i>Tokenizing</i>	33
3.4.7 <i>Whitelist-Based Filtering</i>	34
3.4.8 Filtering Dokumen.....	35
3.5 Modeling.....	36
3.5.1 Representasi Teks dengan <i>Bag-of-Words</i> (BOW).....	36
3.5.2 <i>Vocabulary / Dictionary</i>	38
3.5.3 <i>Document-Term Matrix</i> (DTM).....	39
3.5.4 Model <i>Latent Dirichlet Allocation</i> (LDA).....	41
3.5.5 Penentuan Jumlah Topik (K)	44
3.5.6 Hyperparameter Setting (α , β , passes, iterations).....	45
3.5.7 Output Penentuan Topik Dominan dan Distribusi.....	46
3.6 Evaluasi Model	48
3.6.1 <i>Coherence Score</i>	48
3.6.2 <i>Topic Diversity</i>	50
3.6.3 Visualisasi PyLDAvis (<i>Intertopic Distance Map</i>)	51
3.7 Validasi Ekstraksi Fitur dan Pelabelan Topik	52
3.8 Metode Analisis Hasil Pemodelan.....	53
3.8.1 Analisis Frekuensi Skill (<i>Skill Frequency</i>).....	53
3.8.2 Analisis Dominansi Klaster Skill (<i>Skill Cluster Dominance</i>)	53
3.9 Deployment	54
BAB IV HASIL PENGUJIAN DAN PEMBAHASAN.....	55
4.1 Implementasi	55
4.1.1 Implementasi Pengumpulan Data.....	55

4.1.2	Implementasi <i>Preprocessing</i>	56
4.1.3	Implementasi Ekstraksi Fitur <i>Bag of Word</i>	57
4.1.4	Implementasi LDA (<i>Latent Dirichlet Allocation</i>)	58
4.1.5	Implementasi Evaluasi.....	60
4.1.6	Implementasi Visualisasi PyLDAvis	61
4.1.7	Implementasi Sistem	62
4.1.8	Implementasi Sistem pada Data Baru.....	65
4.2	Hasil.....	66
4.2.1	Hasil Pengumpulan Data dan Preprocessing	66
4.2.2	Hasil Penentuan Jumlah Topik (K) Optimal.....	66
4.2.3	Hasil Ekstraksi Topik dan Top Words	67
4.2.4	Hasil Evaluasi Algoritma.....	69
4.2.5	Hasil Visualisasi PyLDAvis.....	71
4.2.6	Hasil Pengujian Model terhadap Data Maret 2026	72
4.3	Pembahasan	73
BAB V PENUTUP		75
5.1	Kesimpulan.....	75
5.2	Saran	75