

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>HALAMAN PENGESAHAN</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN</b>	<b>iii</b>
<b>HALAMAN PERSEMBAHAN</b>	<b>iv</b>
<b>HALAMAN MOTTO</b>	<b>v</b>
<b>PRAKATA</b>	<b>vi</b>
<b>DAFTAR ISI</b>	<b>viii</b>
<b>DAFTAR TABEL</b>	<b>xii</b>
<b>DAFTAR GAMBAR</b>	<b>xiii</b>
<b>INTISARI</b>	<b>xiv</b>
<b>ABSTRACT</b>	<b>xv</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Pembatasan Masalah	4
1.4. Tujuan dan Manfaat Penelitian	6
1.4.1. Tujuan Penelitian	6
1.4.2. Manfaat Penelitian	6
1.5. Tinjauan Pustaka	7
1.6. Metodologi Penelitian	10
1.7. Sistematika Penulisan	12
<b>II LANDASAN TEORI</b>	<b>15</b>
2.1. Aljabar Vektor dan Matriks	15
2.1.1. Vektor	15
2.1.2. Matriks	16
2.2. Ruang Sampel dan Kejadian	18
2.3. Probabilitas	19
2.4. Variabel Acak	21
2.4.1. Variabel Acak Diskret	21
2.4.2. Variabel Acak Kontinu	23
2.4.3. Sifat-Sifat Variabel Acak	25
2.5. Distribusi Peluang	26
2.5.1. Fungsi Gamma	26

2.5.2.	Distribusi Beta . . . . .	27
2.5.3.	Distribusi Dirichlet . . . . .	29
2.5.4.	Distribusi Multinomial . . . . .	32
2.5.5.	Distribusi Normal . . . . .	34
2.6.	Model Probabilistik . . . . .	38
2.6.1.	Faktorisasi dan Marginalisasi . . . . .	38
2.6.2.	Model Generatif dan Variabel Laten . . . . .	38
2.6.3.	Estimasi Parameter . . . . .	39
2.6.4.	Inferensi Posterior . . . . .	40
2.6.5.	Inferensi Variasional . . . . .	41
2.6.6.	Algoritma Ekspektasi–Maksimisasi . . . . .	42
2.7.	Representasi Vektor dan Embedding . . . . .	43
2.7.1.	Konsep Ruang Vektor Semantik . . . . .	43
2.7.2.	<i>Word Embedding</i> . . . . .	43
2.7.3.	<i>Contextualized Embedding</i> . . . . .	44
2.7.4.	Ruang Vektor Dokumen . . . . .	45
2.8.	<i>Natural Language Processing (NLP)</i> . . . . .	45
2.8.1.	Tingkat Pemrosesan Bahasa . . . . .	46
2.8.2.	Pendekatan dalam Pemrosesan Bahasa . . . . .	47
2.8.3.	Model Bahasa Probabilistik . . . . .	47
2.9.	<i>Machine Learning</i> . . . . .	48
2.9.1.	Konsep Dasar Pembelajaran Mesin . . . . .	48
2.9.2.	Model dan Generalisasi . . . . .	50
2.9.3.	Model Diskriminatif . . . . .	50
2.9.4.	Model Generatif . . . . .	51
<b>III METODEDE PROBABILISTIK, NON PROBABILISTIK, DAN NEURAL EMBEDDING DALAM PEMODELAN TOPIK . . . . .</b>		<b>52</b>
3.1.	Pemodelan Topik . . . . .	52
3.1.1.	Text Preprocessing . . . . .	52
3.1.2.	Representasi Teks . . . . .	55
3.1.3.	Integrasi Representasi Teks . . . . .	59
3.1.4.	Model-model untuk Pemodelan Topik . . . . .	60
3.2.	Latent Dirichlet Allocation (LDA) . . . . .	74
3.2.1.	Proses Generatif . . . . .	76
3.2.2.	Inferensi dan Estimasi . . . . .	78
3.3.	Non-negative Matrix Factorization (NMF) . . . . .	84

3.3.1.	Dasar dan Algoritma . . . . .	84
3.3.2.	Interpretasi dan Representasi Akhir . . . . .	87
3.4.	Contextualized Topic Model (CTM) . . . . .	88
3.4.1.	Variational Autoencoder . . . . .	89
3.4.2.	Alur Pemodelan dalam CTM . . . . .	92
3.4.3.	Varian Combined dan Zero-shot CTM . . . . .	94
3.5.	BERTopic . . . . .	95
3.5.1.	Konsep dan Landasan Teoretis . . . . .	95
3.5.2.	Formulasi Matematis dan Mekanisme Pemodelan . . . . .	97
3.5.3.	Representasi Akhir dan Interpretasi Topik . . . . .	98
3.6.	Generative Artificial Intelligence (Gen AI) . . . . .	99
3.6.1.	Konsep Dasar Generative Artificial Intelligence . . . . .	99
3.6.2.	Arsitektur Tranformer . . . . .	100
3.6.3.	Karakteristik Model Gemini . . . . .	104
3.6.4.	<i>Gemini Flash</i> untuk Pemodelan Topik . . . . .	106
3.7.	Evaluasi Model . . . . .	107
3.7.1.	Topic Coherence . . . . .	108
3.7.2.	Topic Diversity . . . . .	109
<b>IV</b>	<b>STUDI KASUS . . . . .</b>	<b>111</b>
4.1.	Deskripsi Data . . . . .	112
4.2.	Proses Pengambilan Data . . . . .	113
4.3.	Text Pre-processing . . . . .	114
4.3.1.	Case Folding . . . . .	115
4.3.2.	Remove Punctuation . . . . .	116
4.3.3.	Normalisasi Teks . . . . .	117
4.3.4.	Tokenisasi . . . . .	118
4.3.5.	Stopword Removal . . . . .	118
4.3.6.	Stemming . . . . .	119
4.4.	Text Representation . . . . .	120
4.4.1.	Bag of Words (BoW) . . . . .	121
4.4.2.	Term Frequency–Inverse Document Frequency (TF-IDF) . . . . .	122
4.4.3.	Embedding . . . . .	123
4.4.4.	Class-based TF-IDF (cTF-IDF) . . . . .	123
4.5.	Pemodelan Topik . . . . .	124
4.5.1.	Latent Dirichlet Allocation (LDA) . . . . .	125
4.5.2.	Non-negative Matrix Factorization (NMF) . . . . .	129

4.5.3. Contextualized Topic Model (CTM) . . . . .	131
4.5.4. BERTopic . . . . .	134
4.6. Evaluasi Model dan Perbandingan Performa . . . . .	143
4.7. Pelabelan Topic dan Interpretasi . . . . .	145
<b>V PENUTUP . . . . .</b>	<b>148</b>
5.1. Kesimpulan . . . . .	148
5.2. Saran . . . . .	149
<b>A Data . . . . .</b>	<b>157</b>
<b>B Syntax Python . . . . .</b>	<b>158</b>