

## DAFTAR ISI

ABSTRAK .....	i
ABSTRACT .....	ii
KATA PENGANTAR.....	iii
DAFTAR ISI .....	v
DAFTAR GAMBAR .....	viii
DAFTAR TABEL.....	ix
DAFTAR ISTILAH .....	x
DAFTAR LAMPIRAN .....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Kerangka Pemikiran Penelitian.....	6
1.6 Metodologi Penelitian .....	7
1.7 Sistematika Penulisan .....	8
BAB II Kajian Literatur .....	9
2.1 Tinjauan Pustaka .....	9
2.2 Landasan Teori .....	16
2.2.1 Pemodelan Topik.....	16
2.2.2 <i>Clustering</i> .....	16
2.2.3 Python .....	17
2.2.4 <i>Natural Language Processing (NLP)</i> .....	18
2.2.5 <i>Latent Dirichlet Allocation (LDA)</i> .....	18

2.2.6	BERTopic .....	19
2.2.7	<i>Knowledge Discovery in Database Process (KDD)</i> .....	20
2.2.8	<i>Data Scraping: Tweet-Harvest</i> .....	20
2.2.9	<i>Topic Coherence: Coherence Score (<math>C_v</math>)</i> .....	21
2.2.10	<i>Topic Diversity: Jaccard Similarity</i> .....	22
BAB III METODE PENELITIAN.....		23
3.1	<i>Knowledge Discovery Goal</i> .....	23
3.2	<i>Data Integration</i> .....	25
3.3	<i>Data Preprocessing</i> .....	26
3.3.1	<i>Data Cleaning</i> .....	27
3.3.2	<i>Case Folding</i> .....	28
3.3.3	<i>Tokenizing</i> .....	29
3.3.4	<i>Stopword Removing</i> .....	30
3.3.5	Normalisasi .....	30
3.3.6	<i>Stemming</i> .....	31
3.4	<i>Modelling</i> .....	32
3.4.1	Algoritma LDA.....	32
3.4.2	Algoritma BERTopic.....	33
3.5	Pengujian Model .....	35
3.6	Evaluasi Model.....	35
BAB IV HASIL DAN PEMBAHASAN.....		36
4.1	Evaluasi <i>Topic Coherence</i> .....	36
4.2	Evaluasi <i>Topic Diversity</i> .....	37
4.3	Evaluasi Interpretasi Topik Manual .....	38
4.3.1	Model LDA 20 <i>Passes</i> .....	38
4.3.2	Model LDA 50 <i>Passes</i> .....	39

4.3.3	Model LDA 100 <i>Passes</i> .....	41
4.3.4	Model BERTopic.....	42
4.4	Perbandingan LDA dan BERTopic .....	44
BAB V KESIMPULAN .....		46
5.1.	Kesimpulan .....	46
5.2.	Saran.....	47
DAFTAR PUSTAKA .....		48
LAMPIRAN.....		54
Lampiran A: Source Code Tweet Harvest.....		54
Lampiran B: Source Code Data Preprocessing.....		55
Lampiran C: Source Code Latent Dirichlet Allocation.....		59
Lampiran D: Source Code BERTopic .....		61

