

## DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN .....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI .....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR.....	xvi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan dan Manfaat .....	4
1.4 Ruang Lingkup.....	5
1.5 Sistematika Penulisan .....	6
BAB II TINJAUAN PUSTAKA .....	8
2.1 <i>State of the Art</i> .....	8
2.2 <i>Aspect Based Sentiment Analysis (ABSA)</i> .....	12
2.3 <i>Data Scraping</i> .....	13
2.4 Pra-pemrosesan Data.....	13
2.4.1 Deduplikasi Data .....	13
2.4.2 Pembersihan Awal ( <i>Basic Text Cleaning</i> ) .....	14
2.4.3 <i>Hastag Handling</i> .....	14
2.4.4 Konversi Emoji .....	14
2.4.5 <i>Case Folding</i> .....	15
2.4.6 Normalisasi Teks ( <i>Text Normalization</i> ) .....	15
2.4.7 Pra-Pemrosesan untuk Pemodelan Klasifikasi .....	15
2.4.8 Pra-Pemrosesan untuk Pemodelan Topik.....	15

2.5	<i>Topic Modelling</i> .....	16
2.5.1	<i>BERTopic</i> .....	16
2.5.1.1	Pembentukan Representasi Teks ( <i>Embedding</i> ).....	17
2.5.1.2	Reduksi Dimensi dengan <i>UMAP</i> .....	17
2.5.1.3	Klasterisasi Topik dengan <i>HDBSCAN</i> .....	18
2.5.1.4	<i>Class-based TF-IDF (c-TF-IDF)</i> .....	18
2.5.2	Evaluasi Kualitas Topik .....	19
2.5.3	Pemetaan Topik ke Aspek ( <i>Topic-to-Aspect Mapping</i> ) .....	20
2.6	Pelabelan Data.....	20
2.6.1	Pedoman Anotasi dan Konsistensi Keputusan Label.....	21
2.6.2	Pelabelan Ganda dan Reliabilitas Antar Anotator.....	21
2.6.3	Adjudikasi dan Resolusi Konflik Label .....	23
2.6.4	Ketidakeimbangan Kelas ( <i>Class Imbalance</i> ) dalam Klasifikasi Teks.....	24
2.6.4.1	<i>Pos Weight</i> .....	24
2.7	Arsitektur <i>Transformer</i> .....	25
2.7.1	<i>Encoder</i> .....	26
2.7.2	<i>Multi-head attention</i> .....	27
2.7.2.1	<i>Feed Forward Network (FFN)</i> .....	29
2.7.2.2	<i>Decoder</i> .....	30
2.8	Arsitektur <i>BERT</i> .....	30
2.8.1	Arsitektur <i>IndoBERTtweet</i> .....	33
2.8.2	Arsitektur <i>DistilBERTIndo</i> .....	33
2.9	Kerangka Teknik Klasifikasi Teks.....	34
2.9.1	<i>Transfer Learning</i> dan <i>Fine-Tuning</i> .....	34
2.9.2	Klasifikasi <i>Multi-Label</i> dan <i>Multi-Class</i> .....	35
2.10	Optimasi <i>Hyperparameter</i> .....	37
2.10.1	<i>Bayesian Optimization (Tree-structured Parzen Estimator   TPE)</i> .....	37
2.11	Evaluasi Kinerja Model.....	39
2.11.1	<i>Confusion Matrix</i> .....	40

2.11.2	<i>Accuracy</i> dan <i>Subset Accuracy</i> .....	40
2.11.3	<i>Precision</i> .....	41
2.11.4	<i>Recall</i> .....	41
2.11.5	<i>F1-Score</i> dan <i>F1-Macro</i> .....	42
2.11.6	<i>Hamming Loss</i> .....	42
BAB III METODOLOGI PENELITIAN .....		44
3.1	Gambaran Umum Penelitian .....	44
3.2	Pengumpulan Data .....	45
3.3	Pra-pemrosesan Data.....	46
3.3.1	Deduplikasi Data .....	46
3.3.2	Pembersihan Awal.....	47
3.3.3	Penanganan Tagar .....	47
3.3.4	Konversi Emoji ( <i>Demojize</i> ).....	47
3.3.5	<i>Case Folding</i> .....	48
3.3.6	Normalisasi Teks.....	48
3.3.7	Pra-pemrosesan Lanjutan untuk Klasifikasi dan <i>Embedding BERTopic</i> .....	48
3.3.8	Pra-pemrosesan Lanjutan untuk Representasi Topik .....	49
3.4	Ekstraksi Aspek <i>BERTopic</i> .....	50
3.4.1	Pembentukan Representasi Teks ( <i>Embedding</i> ) .....	51
3.4.2	Eksperimen Parameter ( <i>Hyperparameter Tuning</i> ) .....	51
3.4.3	Evaluasi Topik.....	52
3.4.4	Pemodelan Topik Final .....	52
3.4.5	Penanganan <i>Outlier</i> .....	53
3.4.6	Ekstraksi Kata Kunci Topik .....	53
3.4.7	Interpretasi dan Penamaan Aspek .....	53
3.5	Pelabelan Data.....	54
3.5.1	Pengambilan Sampel Data .....	55
3.5.2	Pendefinisian Kelas .....	55
3.5.3	Pelabelan Manual .....	57

3.5.3.1	Contoh Pelabelan Data.....	58
3.5.4	Validasi Pelabelan .....	59
3.5.5	Penyeimbangan Data.....	59
3.6	Pemodelan Aspek ( <i>Aspect Modelling</i> ).....	60
3.6.1	Pembagian Data ( <i>Data Splitting</i> ) .....	61
3.6.2	Tokenisasi ( <i>Tokenization</i> ) .....	62
3.6.3	Pelatihan Model.....	63
3.6.3.1	Inisialisasi Model .....	63
3.6.3.2	Eksperimen ( <i>Hyperparameter Tuning</i> ).....	65
3.6.3.3	Pelatihan Ulang ( <i>Retraining</i> ) .....	66
3.6.3.4	Penyimpanan Model Terbaik.....	66
3.6.4	Evaluasi Model.....	67
3.6.4.1	Confusion Matrix .....	67
3.6.4.2	Akurasi & Subset Akurasi .....	68
3.6.4.3	<i>Precision</i> .....	68
3.6.4.4	<i>Recall</i> .....	69
3.6.4.5	<i>F1-Score</i> .....	69
3.6.4.6	<i>F1-Macro</i> .....	70
3.6.4.7	<i>Hamming Loss</i> .....	70
3.7	Pemodelan Sentimen ( <i>Sentiment Modelling</i> ).....	71
3.7.1	Pembagian Data.....	72
3.7.2	Tokenisasi.....	72
3.7.3	Pelatihan Model.....	73
3.7.3.1	Inisialisasi Model .....	74
3.7.3.2	Eksperimen Hyperparameter ( <i>Hyperparameter Tuning</i> ).....	74
3.7.3.3	Evaluasi Stabilitas Model ( <i>5-Fold Cross-Validation</i> ) .....	75
3.7.3.4	Pelatihan Model Final.....	75
3.7.4	Penyimpanan Model Terbaik .....	76
3.7.5	Evaluasi Model.....	76

BAB IV HASIL DAN PEMBAHASAN.....	77
4. 1 Pengumpulan Data .....	77
4. 2 Pra-Pemrosesan Data .....	78
4.2.1 Deduplikasi Data .....	78
4.2.2 Pembersihan Awal.....	79
4.2.3 Penanganan Tagar .....	80
4.2.4 Konversi Emoji .....	80
4.2.5 <i>Case Folding</i> .....	81
4.2.6 Normalisasi Teks.....	81
4.2.7 Pra-pemrosesan Lanjutan untuk Klasifikasi dan <i>Embedding BERTopic</i> .....	82
4.2.8 Pra-Pemrosesan Lanjutan Untuk Data Pemodelan Topik .....	82
4. 3 Hasil Ekstraksi Aspek Menggunakan <i>BERTopic</i> .....	83
4.3.1 Hasil Eksperimen Parameter ( <i>Hyperparameter Tuning</i> ) .....	83
4.3.2 Penanganan <i>Outlier</i> .....	86
4.3.3 Visualisasi Representasi Kata Kunci.....	86
4.3.4 Interpretasi dan Penamaan Aspek Final .....	88
4.3.5 Kesimpulan Hasil Ekstraksi Aspek .....	91
4. 4 Pelabelan Data.....	92
4.4.1 Hasil Seleksi Data .....	92
4.4.2 Hasil Validasi Anotasi ( <i>Inter-Annotator Agreement</i> ) .....	93
4.4.3 Distribusi Data Final dan Penyeimbangan Kelas .....	94
4.4.4 Analisis Konflik dan Adjudikasi .....	95
4. 5 Hasil Pemodelan Aspek .....	97
4.5.1 Hasil Pembagian Data .....	97
4.5.2 Hasil Pelatihan Model .....	98
4.5.2.1 Hasil Pelatihan Model <i>IndoBERTweet</i> .....	98
4.5.2.2 Hasil Pelatihan Model <i>DistilBERTIndo</i> .....	102
4.5.3 Evaluasi Model.....	105
4.5.4 Kesimpulan Hasil Perbandingan Pemodelan Aspek .....	111

4.6	Hasil Pemodelan Sentimen .....	112
4.6.1	Hasil Pembagian Data .....	113
4.6.2	Hasil Pelatihan Model .....	114
4.6.2.1	Hasil Pelatihan Model <i>IndoBERTtweet</i> .....	114
4.6.2.2	Hasil Pelatihan Model <i>DistilBERTIndo</i> .....	119
4.6.3	Evaluasi Model.....	123
4.6.4	Kesimpulan Hasil Perbandingan Pemodelan Sentimen .....	128
BAB V PENUTUP .....		130
5.1	Kesimpulan .....	130
5.2	Saran .....	131
DAFTAR PUSTAKA.....		132
LAMPIRAN 1: Ilustrasi Proses Pembentukan Topik pada Model <i>BERTopic</i> .....		143
LAMPIRAN 2. Contoh Proses Tokenisasi.....		148
LAMPIRAN 3. Contoh Proses Perhitungan Manual Model Klasifikasi Aspek .....		150
LAMPIRAN 4. Contoh Proses Perhitungan Manual Model Klasifikasi Sentimen.....		157
LAMPIRAN 5. Contoh Proses Perhitungan <i>Cohen's Kappa</i> .....		164
LAMPIRAN 6. Hasil Interpretasi Manual Topik <i>BERTopic</i> .....		167
LAMPIRAN 7. Perbandingan Evaluasi Penanganan <i>Imbalanced Dataset</i> .....		191
LAMPIRAN 8. Panduan Anotasi <i>ABSA</i> – Dataset Opini PBB Pati.....		197