

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iv
PERNYATAAN BEBAS PLAGIASI	iv
PRAKATA	v
DAFTAR ISI	vii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
INTISARI	xvii
ABSTRACT	xviii
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Laporan	4
II TINJAUAN PUSTAKA	5
2.1 Augmentasi Data Teks	5
2.2 Normalisasi Teks	13
III DASAR TEORI	15
3.1 <i>User-Generated Content</i> dan Informalitasnya	15
3.2 Augmentasi Data	16
3.3 <i>Recurrent Neural Network</i>	19

3.4	<i>Sequence to Sequence Model</i>	21
3.5	Mekanisme <i>Attention</i>	23
3.6	Arsitektur <i>Transformers</i>	24
3.6.1	<i>Encoder dan Decoder</i>	26
3.6.2	<i>Attention</i>	26
3.6.3	<i>Embedding</i>	28
3.6.4	<i>Positional Encoding</i>	28
3.7	Pengembangan Model Klasifikasi Teks	29
3.8	Prapemrosesan Data Teks	29
3.9	Representasi Teks	30
3.9.1	<i>Bag of Words</i>	31
3.9.2	<i>Word Embedding</i>	32
3.10	Model Klasifikasi Teks	33
3.10.1	<i>Naive Bayes</i>	34
3.10.2	<i>Long Short Term Memory (LSTM)</i>	37
3.10.3	Arsitektur <i>Bidirectional Encoder Representation from Transformers</i> (BERT)	39
3.10.4	Arsitektur <i>Deep Learning</i> untuk Klasifikasi Teks	40
3.11	Evaluasi Model Normalisasi Teks	44
3.11.1	<i>BLEU Score</i>	44
3.11.2	<i>Character error rate</i>	46
3.11.3	Nilai Akurasi	46
3.12	Evaluasi Model Klasifikasi Teks	47
IV	METODE PENELITIAN	49
4.1	Deskripsi Penelitian	49
4.2	Tahapan Penelitian	50
4.3	Dataset	52
4.4	Normalisasi Teks	55
4.4.1	Alur Pengembangan Model Normalisasi Teks	56
4.4.2	Alur Normalisasi Teks pada Level Kata	58
4.4.3	Proses Normalisasi Teks pada Dataset	59

4.5	Desain Skenario Augmentasi Data Teks	61
4.5.1	<i>Offline Data Augmentation</i>	61
4.5.2	<i>Online Data Augmentation</i> dengan Modifikasi	62
4.6	Pembuatan dan Pengujian Model Klasifikasi Teks	64
V	IMPLEMENTASI	68
5.1	Spesifikasi Perangkat Keras dan Perangkat Lunak	68
5.2	Pengaturan untuk Reproduksi Hasil	70
5.3	Pembangunan Model Normalisasi Teks	71
5.3.1	Implementasi Akuisisi dan Pemrosesan Data	71
5.3.2	Implementasi Pembuatan <i>Character Encoder</i>	72
5.3.3	Implementasi Pembuatan <i>Pipeline</i> untuk Memuat Data	73
5.3.4	Implementasi Inisiasi Arsitektur Transformers	77
5.3.5	Implementasi Proses Pelatihan Model	81
5.3.6	Implementasi Inferensi Model Normalisasi Teks	88
5.3.7	Implementasi Evaluasi Model	90
5.4	Pembuatan Alur Proses Augmentasi Data Teks	91
5.4.1	Implementasi Fungsi Normalisasi Teks	92
5.4.2	Implementasi <i>Command-Line Interface</i> (CLI) Augmentasi Data	96
5.5	Pelatihan Model Klasifikasi Teks	100
5.5.1	Implementasi Inisiasi Objek <i>Tokenizer</i>	100
5.5.2	Implementasi Penyiapan Dataset Pelatihan	102
5.5.3	Implementasi Inisiasi Model Bi-LSTM	108
5.5.4	Implementasi Inisiasi Model BERT	109
5.5.5	Implementasi Alur Pelatihan Model	111
5.5.6	Implementasi Evaluasi Model Klasifikasi Teks	126
VI	HASIL DAN PEMBAHASAN	128
6.1	Pelatihan Model Normalisasi Teks	128
6.1.1	Hasil Evaluasi Model Normalisasi Teks	129
6.1.2	Hasil Luaran Normalisasi Teks	131
6.2	Hasil Evaluasi Pelatihan Model Klasifikasi Teks	134
6.2.1	Augmentasi Data secara <i>Offline</i>	135
6.2.2	Augmentasi Data secara <i>Online</i> dengan Modifikasi	141

6.2.3	Pengukuran Durasi Inferensi Model	147
6.3	Analisis Hasil Model Klasifikasi Teks	148
6.3.1	Skenario Model dengan Performa Terbaik	148
6.3.2	Perbandingan dengan Penelitian Sebelumnya	149
VII KESIMPULAN DAN SARAN		151
7.1	Kesimpulan	151
7.2	Saran	151
A Hasil Evaluasi Model Normalisasi Teks pada Proses <i>Hyperparameter Tuning</i>		159
B Salinan <i>File</i> Lingkungan Anaconda		163