



DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN	ii
HALAMAN PERSEMBAHAN	iii
HALAMAN MOTTO	iv
PRAKATA	v
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xii
ABSTRACT	xiii
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	4
II TINJAUAN PUSTAKA	6
III LANDASAN TEORI	10
3.1 Plagiarisme	10
3.2 <i>Block of sentences</i>	10
3.3 Stilometri	11
3.4 <i>Corpus</i>	11



3.5	<i>n-gram</i>	11
3.6	<i>Natural Language Tool Kit</i>	12
3.7	<i>Part-Of-Speech Tagging</i>	12
3.8	Word Frequencies	12
3.9	Gradient Boosting Regressor	13
3.10	TextTilling	13
3.10.1	<i>Tokenization</i>	14
3.10.2	<i>Lexical Score Determination</i>	14
3.10.3	<i>Boundary Identification</i>	15
3.11	Evaluasi	15
3.12	<i>Accuracy, F1 Score, Precision and Recall</i>	16
3.12.1	<i>Accuracy</i>	16
3.12.2	<i>Precision dan Recall</i>	17
3.12.3	<i>F1 Score</i>	18
3.13	<i>Outlier</i>	18
IV	ANALISIS DAN PERANCANGAN	19
4.1	Deskripsi Umum Penelitian	19
4.1.1	<i>Dataset</i>	19
4.1.2	Tokenisasi dan ekstraksi fitur	21
4.1.3	Vektorisasi	23
4.1.4	Data Training	24
4.1.5	<i>Tresholding</i>	25
4.1.6	<i>Plagiarism Detection</i>	26
4.2	Alat dan Bahan	29
V	IMPLEMENTASI	30
5.1	Hierarki file Implementasi Kode	30
5.2	File ekstraksi_fitur.py	30
5.2.1	Meng- <i>import</i> pustaka yang dibutuhkan	31
5.2.2	Tokenisasi	32
5.2.3	Pelabelan Kalimat	33
5.2.4	<i>Punctuation count</i> dan <i>Part-Of-Speech tagging count</i>	35
5.2.5	<i>Lexical Score</i>	37
5.2.6	<i>Word Frequency</i>	38
5.2.7	Vektorisasi	42



5.3	File <code>truth_label.py</code>	45
5.4	File <code>treshold.py</code>	47
5.4.1	Meng- <i>import</i> pustaka yang dibutuhkan	48
5.4.2	<i>Training Data</i>	48
5.4.3	Menghitung Nilai Kalimat	49
5.4.4	Tresholding	50
5.5	File <code>plagiarsm.py</code>	51
5.5.1	Meng- <i>import</i> pustaka yang dibutuhkan	51
5.5.2	<i>Training Data</i>	52
5.5.3	Inisialisasi nilai threshold	53
5.5.4	Menghitung Nilai Kalimat	54
5.5.5	Membuat Figur	55
5.5.6	Menghitung <i>f-1 score</i>	56
VI HASIL DAN PEMBAHASAN		58
6.1	Hasil Pencarian Nilai Threshold	58
6.2	Grafik Perubahan Nilai Kalimat pada Korpus	59
6.3	Hasil Perbandingan Algoritme	61
6.4	T-test	65
VII KESIMPULAN DAN SARAN		67
7.1	Kesimpulan	67
7.2	Saran	67
DAFTAR PUSTAKA		68