

DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
PERNYATAAN BEBAS PLAGIASI	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
PRAKATA.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
ABSTRAK	xvii
ABSTRACT.....	xviii
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
BAB III LANDASAN TEORI.....	8
3.1. Evaluasi Pembelajaran	8
3.2. Soal Esai	8
3.3. <i>Text mining</i>	9
3.4. <i>Text Preprocessing</i>	9
3.4.1. Tokenisasi	10

3.4.2.	<i>Stop-words removal</i>	11
3.4.3.	<i>Stemming</i>	11
3.5.	<i>Term Frequency</i>	12
3.6.	<i>N-gram</i>	13
3.7.	<i>Term Document Matrix</i>	14
3.8.	<i>Cosine Similarity</i>	15
3.9.	Evaluasi Performa	16
3.9.1.	Akurasi	17
3.9.2.	Presisi	17
3.9.3.	Recall.....	17
3.9.4.	<i>F1-Measure</i>	17
BAB IV ANALISIS DAN PERANCANGAN		18
4.1.	Alat dan Bahan	18
4.2.	Analisis Kebutuhan	18
4.3.	Tahapan Penelitian	21
4.4.	Penggunaan <i>Dataset</i>	24
4.5.	Rancangan Algoritma.....	25
4.5.1.	Anonimisasi Data	26
4.5.2.	<i>Preprocessing</i>	26
4.5.3.	Pembobotan Kata	34
4.5.4.	Pembagian <i>Dataset</i>	35
4.5.5.	Menghitung Nilai Kemiripan	36
4.5.6.	Menentukan <i>Threshold</i>	36
4.6.	Rancangan Pengujian	38
4.6.1.	Pengujian <i>Threshold</i>	38

4.6.2. Perbandingan Nilai Performa Model	38
BAB V IMPLEMENTASI.....	39
5.1. Anonimisasi Data	39
5.2. Implementasi <i>Preprocessing</i>	41
5.2.1. <i>Case Folding</i>	41
5.2.2. <i>Non-alpha-numeric Removal</i>	41
5.2.3. <i>Stopword Removal</i>	42
5.2.4. <i>Stemming</i>	43
5.2.5. Penyeragaman Kata Kunci	44
5.3. Pembobotan Kata	46
5.4. Pembobotan Kata	46
5.4.1. <i>Term Frequency</i>	47
5.4.2. <i>N-gram</i>	48
5.5. Pembagian <i>Dataset</i>	49
5.6. Perhitungan Kemiripan.....	50
5.7. Penentuan <i>Threshold</i>	51
5.7.1. Menentukan Rentang Nilai <i>Threshold</i>	51
5.7.2. Menghitung Nilai Akurasi dan F1	52
5.7.3. Menentukan Nilai <i>Threshold</i>	53
5.8. Pengujian	53
5.8.1. Menghitung Nilai Kemiripan Data <i>Testing</i>	54
5.8.2. Implementasi Nilai <i>Threshold</i>	54
5.8.3. Perbandingan Nilai Performa Model	55
BAB VI HASIL DAN PEMBAHASAN	56
6.1. Hasil Anonimisasi Data.....	56

6.2.	Hasil <i>Preprocessing</i>	57
6.3.	Hasil Pembobotan Kata	58
6.4.	Hasil Pembagian Dataset	59
6.5.	Hasil Perhitungan Kemiripan	61
6.6.	Hasil Penentuan <i>Threshold</i>	62
6.6.1.	Model <i>Term Frequency</i>	63
6.6.2.	Model <i>N-gram</i>	65
6.7.	Hasil Pengujian.....	69
6.7.1.	Hasil implementasi Nilai <i>Threshold</i>	69
6.7.2.	Hasil Perbandingan Nilai Performa	70
BAB VII: KESIMPULAN DAN SARAN		74
7.1.	Kesimpulan.....	74
7.2.	Saran	75
DAFTAR PUSTAKA		76