

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN BEBAS PLAGIASI.....	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	x
INTISARI.....	xi
ABSTRACT.....	xii
BAB 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
BAB II TINJAUAN PUSTAKA.....	4
BAB III LANDASAN TEORI.....	9
3.1 <i>Information Retrieval</i>	9
3.2 <i>Non Standard Words</i>	9
3.3 <i>Text Mining</i>	9
3.4 <i>Text Preprocessing</i>	10
3.4.1 <i>Tokenization</i>	11
3.4.2 <i>Case Folding</i>	11
3.4.3 <i>Text Normalization</i>	11
3.4.4 <i>Stopword Removal</i>	13
3.4.5 <i>Stemming</i>	14
3.5 <i>Similarity Measure</i>	14
3.6 <i>Jaro-Winkler Distance</i>	16
3.7 <i>Levenshtein Distance</i>	17
3.8 <i>N-gram</i>	18

3.9	Smith-Waterman.....	19
3.10	TF-IDF.....	21
3.11	<i>Support Vector Machines</i>	21
3.12	Naive Bayes.....	22
3.12.1	Multinomial Naive Bayes.....	22
3.13	Pengujian.....	23
3.13.1	Evaluasi Normalisasi Teks.....	23
3.13.2	Evaluasi Waktu dan <i>Resource</i>	23
3.13.3	<i>K-Fold Cross Validation</i>	24
3.13.4	Evaluasi Peforma.....	25
BAB IV	ANALISIS DAN PERANCANGAN	26
4.1	Deskripsi Umum Penelitian.....	26
4.2	Analisis Sistem.....	27
4.3	Tahapan Penelitian.....	28
4.4	Tahapan Eksperimen.....	30
4.4.1	Pengumpulan Data.....	30
4.4.2	<i>Preprocessing</i>	32
4.4.3	Perancangan Sistem.....	33
4.4.4	Rancangan Pengujian.....	36
BAB V	IMPLEMENTASI	39
5.1	Deskripsi Implementasi.....	39
5.2	Pengumpulan Data.....	39
5.2.1	Analisis Data <i>Tweet</i>	39
5.2.2	Implementasi <i>Web Scraping</i>	41
5.3	<i>Data Preprocessing</i>	43
5.3.1	Penghapusan URL, Id, dan Simbol.....	44
5.3.2	<i>Case Folding</i>	45
5.3.3	Tokenisasi.....	46
5.3.4	Normalisasi Kata.....	47
5.3.5	Penghapusan <i>Stopword</i>	52
5.3.6	<i>Stemming</i>	53

5.4	Perancangan Sistem.....	53
5.4.1	Sistem Normalisasi Kata.....	54
5.4.2	Pembobotan Kata.....	57
5.4.3	Sistem Klasifikasi Sentimen.....	58
5.5	Pengujian.....	59
5.5.1	<i>Preprocessing</i> Data Normalisasi.....	59
5.5.2	Pengukuran Waktu Normalisasi.....	60
5.5.3	Pengukuran Tingkat Akurasi Algoritma.....	61
5.5.4	Pengukuran Hasil Klasifikasi.....	61
BAB VI	HASIL DAN PEMBAHASAN	62
6.1	Hasil Pengumpulan Data.....	62
6.1.1	Hasil Pengumpulan Data Kuisioner.....	62
6.1.2	Hasil Pengumpulan Data Twitter.....	63
6.2	Hasil <i>Preprocessing</i>	63
6.2.1	Hasil Penghapusan URL, Id, dan Simbol.....	63
6.2.2	Hasil <i>Case Folding</i>	64
6.2.3	Hasil Tokenisasi.....	65
6.2.4	Hasil Normalisasi Kata.....	65
6.2.5	Hasil Penghapusan <i>Stopword</i> dan <i>Stemming</i>	73
6.3	Hasil Pembangunan Sistem.....	73
6.3.1	Hasil Pengukuran Jumlah Kata Unik.....	73
6.3.2	Hasil Pengukuran Waktu dan Akurasi Normalisasi.....	77
6.3.3	Hasil TF-IDF.....	79
6.3.4	Hasil Klasifikasi.....	80
BAB VII	KESIMPULAN DAN SARAN.....	91
7.1	Kesimpulan.....	91
7.2	Saran.....	92
DAFTAR PUSTAKA	93

DAFTAR TABEL

Tabel 2.1	: Perbandingan penelitian yang sudah ada.....	7
Tabel 3.1	: Contoh tokenisasi teks.....	11
Tabel 3.2	: Contoh <i>case folding</i>	11
Tabel 3.3	: Contoh normalisasi kata – <i>spell correction</i>	13
Tabel 3.4	: Contoh <i>stopword removal</i>	14
Tabel 3.5	: Contoh <i>stemming</i>	14
Tabel 3.6	: Contoh perhitungan ukuran kesamaan.....	15
Tabel 3.7	: Contoh ukuran kesamaan dengan <i>Jaro-Winkler Distance</i>	15
Tabel 3.8	: Aturan pemberian nilai pada matriks Smith-Waterman.....	20
Tabel 3.9	: Ilustrasi <i>k-fold cross validation</i> dengan k=5.....	25
Tabel 6.1	: Hasil pengumpulan data kuisioner.....	62
Tabel 6.2	: Hasil pengumpulan data twitter.....	63
Tabel 6.3	: Hasil penghapusan URL, id, dan simbol.....	64
Tabel 6.4	: Hasil <i>case folding</i> data kuisioner	64
Tabel 6.5	: Hasil <i>case folding</i> pada data twitter.....	64
Tabel 6.6	: Hasil tokenisasi data kuisioner	65
Tabel 6.7	: Hasil tokenisasi pada data twitter.....	65
Tabel 6.8	: Hasil normalisasi kata data kuisioner.....	66
Tabel 6.9	: Ringkasan normalisasi kata data kuisioner.....	67
Tabel 6.10	: Hasil normalisasi kata data twitter.....	67
Tabel 6.11	: Ringkasan normalisasi data twitter.....	68
Tabel 6.12	: Jumlah kata unik pada data kuisioner.....	69
Tabel 6.13	: Jumlah kata unik pada data twitter.....	70
Tabel 6.14	: Hasil penghapusan <i>stopword</i> data kuisioner.....	71
Tabel 6.15	: Hasil penghapusan <i>stopword</i> data twitter.....	72
Tabel 6.16	: Hasil <i>stemming</i> data kuisioner.....	72
Tabel 6.17	: Hasil <i>stemming</i> data twitter.....	73
Tabel 6.18	: Hasil pengukuran jumlah kata unik data kuisioner.....	74
Tabel 6.19	: Hasil pengukuran jumlah kata unik data twitter.....	76

Tabel 6.20	: Hasil pengukuran waktu dan normalisasi data kuisisioner.....	78
Tabel 6.21	: Hasil pengukuran waktu dan normalisasi data twitter.....	79
Tabel 6.22	: Hasil klasifikasi SVM data kuisisioner.....	81
Tabel 6.23	: Hasil klasifikasi SVM data twitter.....	82
Tabel 6.24	: Hasil klasifikasi Naive Bayes data kuisisioner.....	82
Tabel 6.25	: Hasil klasifikasi Naive Bayes data twitter.....	83
Tabel 6.26	: Pengaruh normalisasi terhadap klasifikasi SVM data kuisisioner.	84
Tabel 6.27	: Pengaruh normalisasi terhadap klasifikasi SVM data twitter.....	86
Tabel 6.28	: Pengaruh normalisasi terhadap klasifikasi NB data kuisisioner....	87
Tabel 6.29	: Pengaruh normalisasi terhadap klasifikasi NB data twitter.....	89

DAFTAR GAMBAR

Gambar 3.1	: <i>Support Vector Machines</i>	22
Gambar 4.1	: Diagram alur penelitian.....	27
Gambar 4.2	: Diagram alur <i>scraping</i> data twitter.....	31
Gambar 4.3	: Diagram alur pengumpulan data kuisioner.....	32
Gambar 4.4	: Diagram alur <i>preprocessing</i>	33
Gambar 4.5	: Diagram alur normalisasi teks.....	34
Gambar 4.6	: Tahapan normalisasi	35
Gambar 4.7	: Diagram alur klasifikasi	36
Gambar 5.1	: Objek tweet.....	40
Gambar 5.2	: Objek Tweet <i>Extended Mode</i>	41
Gambar 5.3	: Konfigurasi autentikasi twitter.....	42
Gambar 5.4	: Kode <i>web scraping</i>	42
Gambar 5.5	: Kode <i>penghapusan URL</i>	44
Gambar 5.6	: Kode penghapusan id.....	44
Gambar 5.7	: Kode penghapusan simbol.....	45
Gambar 5.8	: Kode <i>case folding</i>	45
Gambar 5.9	: Kode tokenisasi.....	46
Gambar 5.10	: Kode <i>Levenshtein Distance</i>	48
Gambar 5.11	: Kode mencari karakter yang sama.....	49
Gambar 5.12	: Kode menghitung jumlah transposisi.....	49
Gambar 5.13	: Kode <i>Jaro-Winkler Distance</i>	50
Gambar 5.14	: Kode Smith-Waterman.....	51
Gambar 5.15	: Kode Bigram.....	51
Gambar 5.16	: Kode <i>similarity</i> Bigram.....	52
Gambar 5.17	: Kode penghapusan <i>stopword</i>	52
Gambar 5.18	: Kode <i>stemming</i>	53
Gambar 5.19	: Kode pencocokan pada kamus <i>slang</i>	54
Gambar 5.20	: Kode pencocokan pada kamus formal.....	55
Gambar 5.21	: Kode perhitungan nilai similaritas.....	56

Gambar 5.22 : Kode menghitung jumlah kata benar.....	57
Gambar 5.23 : Kode pembobotan kata.....	57
Gambar 5.24 : Kode klasifikasi SVM.....	58
Gambar 5.25 : Kode klasifikasi Naive Bayes.....	59
Gambar 5.26 : Kode implementasi <i>cleansing</i> , <i>case folding</i> , dan tokenisasi.....	60
Gambar 5.27 : Kode menghitung waktu.....	60
Gambar 5.28 : Kode implementasi pengukuran klasifikasi.....	61
Gambar 6.1 : Jumlah kata unik pada data kuisisioner.....	75
Gambar 6.2 : Jumlah kata unik pada data twitter.....	76
Gambar 6.3 : Hasil vektorisasi kata.....	80
Gambar 6.4 : Pengaruh normalisasi terhadap klasifikasi SVM data kuisisioner.....	85
Gambar 6.5 : Pengaruh normalisasi terhadap klasifikasi SVM data twitter.....	87
Gambar 6.6 : Pengaruh normalisasi terhadap klasifikasi NB data kuisisioner.....	88
Gambar 6.7 : Pengaruh normalisasi terhadap klasifikasi NB data twitter.....	90