

DAFTAR ISI

TESIS	i
HALAMAN PENGESAHAN	ii
PERNYATAAN	iii
PRAKATA	iv
DAFTAR ISI	i
DAFTAR GAMBAR	viii
DAFTAR TABEL	iv
INTISARI	xi
ABSTRACT	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Keaslian Penelitian	4
1.6 Manfaat Penelitian	5
1.7 Metode Penelitian	5
1.8 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
BAB III LANDASAN TEORI	11
3.1 Analisis Sentimen	11
3.1.1 Native Bayes Clasifier	12
3.2 Clustering	13
3.2.1 Algoritma Suffix Tree Clustering (STC)	17
3.3 Framework Carrot2	20
3.3.1 Arsitektur	21
3.3.2 Format Input XML Carrot2	21
3.3.3 Menambahkan Framework Carrot2 ke dalam Aplikasi Java	22
BABIV ANALISIS SISTEM	23
4.1 Gambaran Umum Sistem	23
4.2 Pre-processing Tweet	25
4.2.1 Case Folding	26
4.2.2 Konversi Emoticon	26
4.2.3 Stopword Removal	27
4.2.4 Normalisasi ke Kata Baku	28
4.2.5 Stemming	28
4.3 Ekstraksi Fitur	29
4.4 Proses Training	30
4.4.1 Pembelajaran 1 untuk Sentiment Positif	31
4.4.2 Pembelajaran 2 untuk Sentimen Netral	32
4.4.3 Pembelajaran 3 untuk Sentimen Negatif	34

4.5 Clustering Tweet	36
4.6 Klasifikasi Tweet	40
BAB V IMPLEMENTASI	48
5.1 Lingkungan Implementasi	48
5.2 Proses Pengambilan Tweet	48
5.3 Pre-processing Tweet	49
5.3.1 Case Folding Tweet	50
5.3.2 Konversi Emoticon Tweet	50
5.3.3 Stopword Removal Tweet	51
5.3.4 Stemming Tweet	52
5.4 Seleksi Fitur	53
5.5 <i>Clustering</i> Tweet dan Analisis Sentimen	53
5.5.1 Clustering Tweet	54
5.5.2 Analisis Sentimen	55
BAB VI HASIL DAN PEMBAHASAN	57
6.1 Pengujian Akurasi Analisis Sentimen	57
6.2 Pengujian Clustering	61
6.3 Pembahasan	62
BABVII KESIMPULAN SARAN	66
7.1 Kesimpulan	66
7.2 Saran	67
DAFTAR PUSTAKA	68

DAFTAR GAMBAR

Gambar 3.1 Contoh dendogram	16
Gambar 3.2 Pembentukan suffix tree	18
Gambar 3.3 Base cluster graph	20
Gambar 3.4 Format input XML carrot2	21
Gambar 4.1 Alur sistem	24
Gambar 4.2 <i>Pre-Processing</i>	25
Gambar 4.3 <i>Suffix tree</i> yang terbentuk	38
Gambar 5.1 <i>Source code</i> untuk pengambilan tweet	49
Gambar 5.2 <i>Pre-processing</i> tweet	49
Gambar 5.3 <i>Source code case folding</i> tweet	50
Gambar 5.4 <i>Source code</i> konversi <i>emoticon</i>	51
Gambar 5.5 <i>Source code stopword removal</i>	52
Gambar 5.6 <i>Source code stemming</i>	52
Gambar 5.7 Tampilan proses ekstraksi fitur	53
Gambar 5.8 <i>Source code</i> untuk penghitungan frekuensi term	53
Gambar 5.9 Tampilan proses <i>clustering</i> dan analisis sentimen	54
Gambar 5.10 <i>Source code clustering</i> dengan algoritma STC	54
Gambar 5.11 <i>Source code</i> analisis sentimen	55
Gambar 5.12 <i>Pie chart</i> hasil analisis sentimen topik tertentu	56
Gambar 6.1 Pengujian akurasi dengan data uji sebanyak 50 tweet	58
Gambar 6.2 Pengujian akurasi dengan data uji sebanyak 100 tweet	58
Gambar 6.3 Pengujian akurasi dengan data uji sebanyak 150 tweet	59
Gambar 6.4 Pengujian akurasi dengan data uji sebanyak 200 tweet	59
Gambar 6.5 Contoh hasil <i>clustering</i> dengan 10 <i>cluster</i>	63
Gambar 6.6 Hasil analisis sentimen untuk topik “Setya Novanto”	63

DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian	9
Tabel 3.1 Base cluster yang terbentuk	18
Tabel 4.1 Daftar kata kunci untuk kueri tweet bertema “E-KTP”	23
Tabel 4.2 Contoh tweet kotor	25
Tabel 4.3 Contoh hasil tahapan <i>case folding</i>	26
Tabel 4.4 Daftar <i>emoticon</i>	26
Tabel 4.5 Contoh hasil tahapan <i>stopword removal</i>	27
Tabel 4.6 Daftar contoh padanan kata	28
Tabel 4.7 Contoh hasil tahapan normalisasi ke kata baku	28
Tabel 4.8 Contoh hasil tahapan <i>stemming</i>	29
Tabel 4.9 Daftar contoh kata sentimen	29
Tabel 4.10 Daftar tweet <i>training</i> data	30
Tabel 4.11 Daftar term sentimen positif	31
Tabel 4.12 Daftar probabilitas term sentimen positif	31
Tabel 4.13 Daftar term sentimen netral	32
Tabel 4.14 Daftar probabilitas term sentimen netral	33
Tabel 4.15 Perubahan probabilitas term sentimen positif pertama	33
Tabel 4.16 Daftar term sentimen negatif	34
Tabel 4.17 Probabilitas term sentimen negatif	35
Tabel 4.18 Perubahan probabilitas term sentimen netral	35
Tabel 4.19 Perubahan probabilitas term sentimen positif	36
Tabel 4.20 Daftar contoh tweet kotor untuk proses <i>clustering</i>	37
Tabel 4.21 Daftar contoh tweet bersih untuk proses <i>clustering</i>	37
Tabel 4.22 Contoh penentuan <i>base cluster</i> algoritma STC	38
Tabel 4.23 Contoh perhitungan skor <i>base cluster</i>	38
Tabel 4.24 Data uji proses klasifikasi tweet	41
Tabel 4.25 Nilai probabilitas term pada sentimen positif utk tweet 6	41
Tabel 4.26 Nilai probabilitas term pada sentimen positif utk tweet 7	42
Tabel 4.27 Nilai probabilitas term pada sentimen positif utk tweet 6 & tweet 7 ...	42
Tabel 4.28 Nilai probabilitas term pada sentimen negatif utk tweet 6	43
Tabel 4.29 Nilai probabilitas term pada sentimen negatif utk tweet 7	43
Tabel 4.30 Nilai probabilitas term pd sentimen negatif utk tweet 6 & tweet 7 ..	44
Tabel 4.31 Nilai probabilitas term pada sentimen netral utk tweet 6	44
Tabel 4.32 Nilai probabilitas term pada sentimen netral utk tweet 7	45
Tabel 4.33 Nilai probabilitas term pada sentimen netral utk tweet 6 & tweet 7	45



Tabel 4.34 Perbandingan nilai V_{MAP} antar kelas sentimen	46
Tabel 4.35 Hasil klasifikasi tweet pada <i>cluster</i> / topik “korupsi ektp”	47
Tabel 6.1 Deskripsi data uji	57
Tabel 6.2 Hasil proses pengujian akurasi.....	60
Tabel 6.3 Data hasil pengujian <i>clustering</i>	61
Tabel 6.4 Contoh tweet bergaya bahasa satire	64