

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN.....	ii
PERNYATAAN ORISINALITAS SKRIPSI.....	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian.....	4
1.7 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Kajian Pustaka.....	6
2.1.1 Jurnal Klasifikasi <i>E-mail Spam</i> Berbahasa Inggris.....	6
2.2 Emosi Lagu.....	6
2.2.1 Definisi Emosi Lagu.....	6
2.2.2 Jenis Lagu.....	7

2.2.3	Bagian Lagu	7
2.3	Teks Mining.....	9
2.4	Tahap Teks Mining	10
2.4.1	Tahap <i>Preprocessing</i>	10
2.4.1.1	<i>Preprocessing</i> Tahap <i>Tokenizing</i>	10
2.4.1.2	<i>Preprocessing</i> Tahap <i>Stemming</i>	11
2.4.1.3	<i>Preprocessing</i> Tahap <i>Stopword</i>	11
2.4.1.4	<i>Preprocessing</i> Tahap Frekuensi Kata	11
2.4.2	Tahap Teks <i>Transformation</i> dan <i>Feature Generation</i>	12
2.4.3	Tahap Data Mining atau <i>Pattern Discovery</i>	12
2.5	<i>Stemming</i>	13
2.5.1	Algoritma <i>Porter Stemmer</i>	13
2.5.2	Algoritma <i>Porter Stemmer</i> untuk Lirik Bahasa Inggris.....	15
2.6	<i>Term Frequency-Inverse Document Frequency</i>	20
2.7	<i>Transformation</i> Data dengan Distribusi Frekuensi.....	21
2.8	<i>Decision Tree</i>	22
2.9	<i>Iterative Dichotomizer Tree</i>	24
2.10	Evaluasi	25
2.11	<i>Accuracy</i>	27
BAB III METODELOGI DAN PERANCANGAN		28
3.1	Analisa Data	28
3.2	Deskripsi Umum Sistem	29
3.3	Perancangan Proses.....	31
3.3.1	Perancangan <i>Preprocessing</i>	31
3.3.1.1	Perancangan Proses <i>Tokenizing</i>	32

3.3.1.2	Perancangan Proses <i>Stopword</i>	33
3.3.1.3	Perancangan Proses <i>Stemming</i>	34
3.3.2	Perancangan Proses <i>tf-idf</i>	35
3.3.3	Perancangan Transformasi Data.....	36
3.3.4	Pembelajaran	39
3.3.5	Proses Klasifikasi.....	43
3.4	Perhitungan Manual.....	44
3.4.1	Sumber Data.....	44
3.4.2	Proses Pembobotan.....	45
3.4.3	Proses <i>Data Transformation</i>	48
3.4.4	Perancangan Pohon Keputusan.....	50
3.4.5	Perancangan <i>Rule Tree</i>	54
3.4.6	Perancangan Proses Klasifikasi.....	55
3.5	Perancangan Antarmuka.....	57
3.6	Metode Pengujian.....	58
3.6.1	Perhitungan <i>Accuracy</i>	58
BAB IV IMPLEMENTASI.....		59
4.1	Lingkungan Implementasi.....	59
4.1.1	Lingkungan Perangkat Keras.....	59
4.1.2	Lingkungan Perangkat Lunak.....	59
4.2	Implementasi Program.....	59
4.2.1	Implementasi <i>Preprocessing</i>	60
4.2.1.1	<i>Tokenizing</i>	60
4.2.1.2	<i>Stopwords</i>	61
4.2.1.3	<i>Stemming</i>	62

4.2.1.4 Frekuensi Kata	62
4.2.2 Implementasi Pembobotan	63
4.2.3 Implementasi Transformasi Data	64
4.2.4 Implementasi Pembentukan Pohon Keputusan	66
4.2.5 Implementasi Keputusan ID3	69
4.3 Penerapan Aplikasi	70
BAB V ANALISA DAN PEMBAHASAN	71
5.1 Skenario Pengujian	71
5.1.1 Data Pengujian	71
5.1.2 Lingkungan Pengujian	71
5.1.3 Hasil Pengujian	71
5.2 Analisa Hasil	73
5.2.1 Analisa Hasil Klasifikasi <i>Training</i> lebih besar dari <i>Testing</i>	73
5.2.2 Analisa Hasil Klasifikasi <i>Testing</i> lebih besar dari <i>Training</i>	74
5.2.3 Analisa Hasil Klasifikasi <i>Training</i> dan <i>Testing</i> dengan perbandingan sama	75
BAB VI KESIMPULAN DAN SARAN	77
6.1 Kesimpulan	77
6.2 Saran	77
DAFTAR PUSTAKA	78
LAMPIRAN 1	80
LAMPIRAN 2	84

DAFTAR GAMBAR

Gambar 2.1 Tahap dalam Teks Mining	10
Gambar 2.2 <i>Porter Stemmer</i>	14
Gambar 2.3 <i>Decision Tree</i> untuk pembelian komputer.....	22
Gambar 2.4 <i>Decision Tree</i> yang memungkinkan dalam Proses Partisi Data	23
Gambar 3.1 Alur Penelitian	29
Gambar 3.2 <i>Flowchart</i> Deskripsi Umum Sistem.....	31
Gambar 3.3 <i>Flowchart</i> Perancangan <i>Preprocessing</i>	32
Gambar 3.4 <i>Flowchart</i> Proses <i>Tokenizing</i>	33
Gambar 3.5 <i>Flowchart</i> Proses <i>Stopword</i>	34
Gambar 3.6 <i>Flowchart</i> Proses Pembobotan	36
Gambar 3.7 <i>Flowchart</i> Proses Data Transformasi.....	37
Gambar 3.8 <i>Flowchart</i> Proses Distribusi Frekuensi	38
Gambar 3.9 <i>Flowchart</i> Proses Transformasi Bobot.....	39
Gambar 3.10 <i>Flowchart</i> Proses <i>Training ID3</i>	39
Gambar 3.11 <i>Flowchart</i> Proses Pencarian <i>Root</i>	41
Gambar 3.12 <i>Flowchart</i> Proses Pembentukan <i>Node</i>	43
Gambar 3.13 <i>Flowchart</i> Proses Klasifikasi ID3	44
Gambar 3.14 <i>Tree</i> Hasil Perhitungan <i>Root</i>	51
Gambar 3.15 <i>Tree</i> Hasil Cabang Int 1 dari Atribut <i>Screw</i>	52
Gambar 3.16 <i>Tree</i> Hasil Cabang Int 4 dari Atribut <i>Screw</i>	53
Gambar 3.17 <i>Tree</i> Hasil Cabang Int 8 dari Atribut <i>Screw</i>	54
Gambar 3.18 Antarmuka Sistem.....	57
Gambar 4.1 Antarmuka Sistem.....	70

Gambar 5.1 Grafik hasil akurasi klasifikasi *training* lebih besar dari *testing* 73

Gambar 5.2 Grafik hasil akurasi klasifikasi *testing* lebih besar dari *training* 74

Gambar 5.3 Grafik hasil akurasi klasifikasi *training* dan *testing* dengan perbandingan sama 75



DAFTAR TABEL

Tabel 2.1 Jenis Lagu.....	7
Tabel 2.2 Kondisi pada <i>Stem</i>	16
Tabel 2.3 <i>Step</i> 1a: Menghapus akhiran jamak	18
Tabel 2.4 <i>Step</i> 1b: Menghapus kata kerja verbal.....	18
Tabel 2.5 <i>Step</i> 1b1: Lanjutan untuk <i>-ed</i> dan <i>-ing</i>	18
Tabel 2.6 <i>Step</i> 1c: <i>y</i> dan <i>i</i>	18
Tabel 2.7 <i>Step</i> 2: Memisahkan satu akhiran untuk beberapa akhiran.....	19
Tabel 2.8 <i>Step</i> 3.....	19
Tabel 2.9 <i>Step</i> 4: Menghapus akhiran terakhir	20
Tabel 2.10 <i>Step</i> 5a: Hapus <i>e</i>	20
Tabel 2.11 <i>Step</i> 5b: Reduksi	20
Tabel 2.12 Matrik <i>Confusion</i>	26
Tabel 3.1 Data <i>training</i> untuk <i>term frequency</i> , <i>document frequency</i> dan <i>invers document frequency</i>	45
Tabel 3.2 Data <i>testing</i> untuk <i>term frequency</i> , <i>document frequency</i> dan <i>invers document frequency</i>	46
Tabel 3.3 Data <i>training</i> hasil pembobotan kata.....	47
Tabel 3.4 Data <i>testing</i> hasil pembobotan kata.....	47
Tabel 3.5 Hasil pengurutan data pembobotan	49
Tabel 3.6 Data <i>training</i> hasil transformasi data.....	49
Tabel 3.7 Hasil perhitungan <i>entropy</i> dan <i>information gain</i>	51
Tabel 3.8 Data atribut pada cabang Int 1 untuk atribut <i>screw</i>	52
Tabel 3.9 Data atribut pada cabang Int 4.....	53
Tabel 3.10 Data atribut pada cabang Int 8.....	54

Tabel 3.11 Data *testing* hasil generalisasi 55

Tabel 3.12 Perhitungan *accuracy*..... 58

Tabel 4.1 Kelas pada program 60

Tabel 4.2 *Method* pada *preprocessing* 60

Tabel 4.3 *Method* pembentukan pohon keputusan..... 66

Tabel 5.1 Hasil klasifikasi *training* lebih besar dari *testing* 72

Tabel 5.2 Hasil klasifikasi *testing* lebih besar dari *training* 72

Tabel 5.3 Hasil klasifikasi *training* dan *testing* dengan perbandingan sama 73

