

**PENGKATEGORIAN PESAN SINGKAT
BERBAHASA INDONESIA
PADA JEJARING SOSIAL TWITTER
DENGAN METODE KLASIFIKASI NAÏVE BAYES**

**SKRIPSI
KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI**

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh :
RIZAL SETYA PERDANA
NIM. 0910680030

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
PROGRAM STUDI INFORMATIKA/ILMU KOMPUTER
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
MALANG
2013**

LEMBAR PERSETUJUAN

**PENGKATEGORIAN PESAN SINGKAT BERBAHASA INDONESIA
PADA JEJARING SOSIAL TWITTER
DENGAN METODE KLASIFIKASI NAÏVE BAYES**

SKRIPSI

Diajukan untuk memenuhi persyaratan
memperoleh gelar Sarjana Komputer



Disusun oleh :

RIZAL SETYA PERDANA
NIM. 0910680030

Telah diperiksa dan disetujui untuk diuji dengan dosen pembimbing:

1. Suprpto, ST., MT
 2. Rekyan Regasari Mardi Putri, ST., MT
- Malang, 4 Januari 2013

Dosen Pembimbing I

Dosen Pembimbing II

Suprpto, ST., MT

Rekyan Regasari Mardi Putri, ST., MT

NIP. 197107271996031001

NIP. 77041406120253

KATA PENGANTAR

Puji dan syukur saya ucapkan kepada Allah SWT karena atas rahmat dan hidayah-Nya saya dapat menyelesaikan Proposal Tugas Akhir ini sebagai salah satu persyaratan untuk menyelesaikan studi di Program Studi Teknik Informatika Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.

Saya menyadari bahwa Proposal Tugas Akhir ini dapat terselesaikan atas bantuan, petunjuk, dan bimbingan dari berbagai pihak yang telah membantu proses penyelesaiannya. Oleh karena itu, saya mengucapkan terima kasih banyak kepada :

1. Orang tua saya, Irfan, SE. dan Dra. Dwi Orbaningsih, MM. yang tak henti-hentinya memberikan dorongan moril dan materil serta Meutia Tamimi Auli yang telah membantu hingga terselesaikanya proposal skripsi ini.
2. Bapak Suprpto, ST., MT dan Ibu Rekyan Regasari MP, ST., MT selaku pembimbing I dan pembimbing II yang telah memberikan bimbingan dan arahan dengan penuh kesabaran sehingga proposal ini dapat terselesaikan.
3. Drs. Marji, MT selaku ketua Program Studi Teknik Informatika.
4. Seluruh dosen pembina pada Program Studi Teknik Informatika dan Ilmu Komputer yang telah membimbing penulis selama menjadi mahasiswa Universitas Brawijaya Malang.
5. Seluruh rekan kerja dan supervisor PPTI UB yang telah membimbing dan memberi saran masukan kepada penulis hingga terselesaikanya skripsi ini.
6. Teman-teman yang telah banyak membantu dalam pelaksanaan penelitian.

Semoga Tugas Akhir ini dapat memberikan manfaat bagi pembaca sekaligus dapat menjadi bahan acuan untuk penelitian selanjutnya.

Malang, 18 Januari 2013

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Pembahasan.....	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	
2.1 Twitter.....	6
2.2 <i>Text Mining</i>	7
2.3 Tahapan <i>Text Mining</i>	7
2.3.1 <i>Text Preprocessing</i>	8
2.3.2 <i>Text Transformation (future generation)</i>	8
2.3.3 <i>Pattern Discovery</i>	10
2.4 <i>Stemming</i> pada Bahasa Indonesia.....	12
2.4.1 Struktur morfologi kata bahasa Indonesia.....	12

2.4.2 Metode Stemming Arifin dan Setiono.....	15
2.5 <i>Naive Bayes Classifier</i>	19
2.6 Contoh Implementasi Algoritma <i>Naive Bayes Classifier</i>	22
2.7 Evaluasi	28

BAB III METODOLOGI PENELITIAN

3.1 Studi Literatur.....	30
3.2 Analisis Kebutuhan Sistem.....	30
3.3 Merancang Aplikasi.....	32
3.4 Implementasi Pembuatan Aplikasi	34
3.4.1 Lingkungan Sistem	34
3.4.2 Implementasi Aplikasi	35
3.4.2.1 Perancangan <i>Text Preprocessing</i>	35
3.4.2.2 Perancangan <i>Text Transformation</i>	36
3.4.2.3 Perancangan Perhitungan Frekuensi Kata	37
3.4.2.4 Perancangan <i>Pattern Discovery</i>	37
3.5 Pengujian dan Analisis.....	42
3.6 Pengambilan Kesimpulan	42

BAB IV ANALISIS KEBUTUHAN DAN PERANCANGAN

4.1 Analisis Kebutuhan Perangkat Lunak	43
4.2 Perancangan Perangkat Lunak.....	44
4.2.1 Diagram Blok Pemrosesan Data Latih	45
4.2.2 Diagram Blok Sistem Pengkategorian.....	46
4.2.3 Perancangan Sistem Manajemen Data	47
4.2.4 Diagram <i>Use Case</i>	52
4.2.5 Diagram Kelas	53
4.2.6 Diagram Blok Sistem Keseluruhan	57
4.2.7 Diagram <i>Sequence</i>	58
4.2.8 <i>Case Folding</i> dan <i>Parsing</i>	61
4.2.9 Algoritma <i>Stemming</i> Arifin	62

4.2.10 Algoritma <i>Naive Bayes</i>	69
4.2.11 Contoh Implementasi Algoritma <i>Naive Bayes Classifier</i>	70
4.2.12 Perancangan Pengujian	75

BAB V IMPLEMENTASI

5.1 Lingkungan Implementasi	78
5.1.1 Lingkungan Perangkat Keras	78
5.1.2 Lingkungan Perangkat Lunak	78
5.2 Implementasi Perangkat Lunak	79
5.2.1 Pemrosesan Data Latih	79
5.2.2 Sistem Pengkategorian	82
5.2.3 Twitter API	87
5.2.4 Tampilan Aplikasi	91

BAB VI PENGUJIAN DAN ANALISA

BAB VII KESIMPULAN DAN SARAN

7.1 Kesimpulan	118
7.2 Saran	119
DAFTAR PUSTAKA	DP-1

DAFTAR GAMBAR

	Halaman
Gambar 3.1 Diagram Blok Penelitian.....	31
Gambar 3.2 Diagram Alir Keseluruhan Sistem.....	33
Gambar 3.3 Diagram Alir <i>Preprocessing</i>	33
Gambar 3.4 Diagram Alir <i>Transformation</i>	34
Gambar 3.5 Diagram Alir Proses <i>Learn Naive Bayes</i>	39
Gambar 3.6 Diagram Alir Proses <i>Classify Naive Bayes</i>	41
Gambar 4.1 Diagram Blok Sistem Pemrosesan Data Latih.....	45
Gambar 4.2 Diagram Blok Sistem Pengkategorian.....	46
Gambar 4.3 ERD Aplikasi.....	48
Gambar 4.4 Diagram <i>Use Case</i> Aplikasi.....	52
Gambar 4.5 Diagram Kelas Aplikasi.....	54
Gambar 4.6 Diagram Blok Sistem Keseluruhan.....	57
Gambar 4.7 <i>Sequence Diagram</i> Proses Login.....	58
Gambar 4.8 <i>Sequence Diagram</i> Preproses Teks.....	59
Gambar 4.9 <i>Sequence Diagram</i> Transformasi Teks.....	60
Gambar 4.10 <i>Sequence Diagram</i> Perhitungan Probabilitas.....	61
Gambar 4.11 <i>Flowchart Stemming</i> Arifin.....	63
Gambar 4.12 <i>Flowchart</i> Kombinasi Kata.....	65
Gambar 5.1 <i>Source Code</i> Pengambilan RSS.....	80
Gambar 5.2 <i>Source Code</i> Pembersihan RSS.....	81

Gambar 5.3 <i>Source Code</i> Penyimpanan RSS pada Database.....	82
Gambar 5.4 <i>Source Code</i> Pembentukan Tabel Frekuensi	82
Gambar 5.5 <i>Source Code Case Folding</i>	83
Gambar 5.6 <i>Source Code</i> Fungsi <i>Explode</i>	83
Gambar 5.7 <i>Source Code</i> Fungsi Penghapusan Karakter	84
Gambar 5.8 <i>Source Code</i> Fungsi <i>Stemming</i>	84
Gambar 5.9 <i>Source Code</i> Perhitungan $P(w_k v_j)$ dan $P(v_j)$	86
Gambar 5.10 <i>Source Code</i> Perhitungan $V_{nb} = \text{argmax } P(v_j) \prod P(a_i v_j)$	87
Gambar 5.11 <i>Source Code</i> Pendefinisian <i>Key</i>	88
Gambar 5.12 <i>Source Code</i> Kelas <i>Library OAuth</i>	88
Gambar 5.13 <i>Source Code Library</i> Twitter API	89
Gambar 5.14 <i>Source Code</i> Fungsi-fungsi Akses Pengguna	89
Gambar 5.15 <i>Source Code</i> Pembuatan Koneksi pada Twitter	90
Gambar 5.16 <i>Source Code</i> Penggunaan <i>Webservice</i>	90
Gambar 5.17 <i>Source Code</i> Kode Javascript GUI	91
Gambar 5.18 Tampilan Aplikasi <i>Client</i> Saat Dibuka Pertama dan Halaman Login ..	93
Gambar 5.19 Tampilan Aplikasi <i>Client</i> pada saat Mengkategorikan	94
Gambar 6.1 Grafik Waktu Rata-rata Pengkategorian	102
Gambar 6.2 Grafik Waktu Evaluasi Rata-rata Pengujian dengan <i>Stemming</i>	114
Gambar 6.3 Grafik Waktu Evaluasi Rata-rata Pengujian Tanpa <i>Stemming</i>	115
Gambar 6.4 Grafik Perbandingan Eksekusi dengan Menggunakan F_1 Measure	116

DAFTAR TABEL

	Halaman
Tabel 2.1 Pasangan Konfiks yang tidak Diperbolehkan.....	14
Tabel 2.2 Urutan Prefiks Ganda	15
Tabel 2.3 Daftar Kata Dokumen Pertama	23
Tabel 2.4 Pengetahuan Dokumen Pertama	24
Tabel 2.5 Daftar Kata Dokumen Kedua	25
Tabel 2.6 Pengetahuan Dokumen Pertama dan Kedua.....	26
Tabel 2.7 Daftar Kata Dokumen Ketiga	27
Tabel 2.8 Matriks <i>Confusion</i>	28
Tabel 4.1 Daftar Awalan	66
Tabel 4.2 Daftar Akhiran.....	67
Tabel 4.3 Daftar Frekuensi Kata dari <i>Database</i>	71
Tabel 4.4 Daftar Statistik Kata dan Dokumen.....	71
Tabel 4.5 Perhitungan Kategori Olahraga.....	72
Tabel 4.6 Perhitungan Kategori Hiburan.....	72
Tabel 4.7 Perhitungan Kategori Berita	72
Tabel 4.8 Perhitungan Kategori Otomotif.....	73
Tabel 4.9 Perhitungan Kategori Keuangan.....	73
Tabel 4.10 Perhitungan Kategori Teknologi	73
Tabel 4.11 Total Nilai Perhitungan Kategori	74

Tabel 4.12 Pembagian Jumlah Data Latih.....	75
Tabel 4.13 Perbandingan Pengkategorian oleh Sistem dan Manual	76
Tabel 4.14 Perbandingan Rekap Hasil oleh Sistem dan Manual.....	76
Tabel 5.1 Daftar Alamat URL Sumber RSS.....	79
Tabel 6.1 Data Pengujian.....	96
Tabel 6.2 Komposisi Data Latih yang Digunakan pada Pengkategorian	97
Tabel 6.3 Data Hasil Pengkategorian dengan Sistem <i>stemming</i> Pada berbagai Jumlah Data Latih	98
Tabel 6.4 Data Hasil Pengkategorian dengan Sistem tanpa <i>stemming</i> Pada berbagai Jumlah Data Latih.....	99
Tabel 6.5 Waktu Eksekusi Proses Pengkategorian dengan Melalui Proses <i>Stemming</i> dalam <i>Microtime</i>	100
Tabel 6.6 Waktu Eksekusi Proses Pengkategorian Tanpa Melalui Proses <i>Stemming</i> dalam <i>Microtime</i>	101
Tabel 6.7 Jumlah <i>Term</i> dan Dokumen Pada Masing-masing Stage dengan Proses <i>Stemming</i>	104
Tabel 6.8 Jumlah <i>Term</i> dan Dokumen Pada Masing-masing Stage Tanpa Proses <i>Stemming</i>	105
Tabel 6.9 Hasil Pengkategorian Secara Manual Oleh Responden	106
Tabel 6.10 <i>Recall, Precision, F₁ measure</i> pada Stage 1 Melalui Proses <i>Stemming</i> .	107
Tabel 6.11 <i>Recall, Precision, F₁ measure</i> pada Stage 2 Melalui Proses <i>Stemming</i> .	107
Tabel 6.12 <i>Recall, Precision, F₁ measure</i> pada Stage 3 Melalui Proses <i>Stemming</i> .	107
Tabel 6.13 <i>Recall, Precision, F₁ measure</i> pada Stage 4 Melalui Proses <i>Stemming</i> .	107
Tabel 6.14 <i>Recall, Precision, F₁ measure</i> pada Stage 5 Melalui Proses <i>Stemming</i> .	108

Tabel 6.15 *Recall, Precision, F_1 measure* pada Stage 6 Melalui Proses *Stemming* . 108

Tabel 6.16 *Recall, Precision, F_1 measure* pada Stage 7 Melalui Proses *Stemming* . 108

Tabel 6.17 *Recall, Precision, F_1 measure* pada Stage 8 Melalui Proses *Stemming* . 109

Tabel 6.18 *Recall, Precision, F_1 measure* pada Stage 9 Melalui Proses *Stemming* . 109

Tabel 6.19 *Recall, Precision, F_1 measure* pada Stage 10 Melalui Proses *Stemming* 109

Tabel 6.20 *Recall, Precision, F_1 measure* pada Stage 1 Tanpa Proses *Stemming*.... 110

Tabel 6.21 *Recall, Precision, F_1 measure* pada Stage 2 Tanpa Proses *Stemming*.... 110

Tabel 6.22 *Recall, Precision, F_1 measure* pada Stage 3 Tanpa Proses *Stemming*.... 110

Tabel 6.23 *Recall, Precision, F_1 measure* pada Stage 4 Tanpa Proses *Stemming*.... 111

Tabel 6.24 *Recall, Precision, F_1 measure* pada Stage 5 Tanpa Proses *Stemming*.... 111

Tabel 6.25 *Recall, Precision, F_1 measure* pada Stage 6 Tanpa Proses *Stemming*.... 111

Tabel 6.26 *Recall, Precision, F_1 measure* pada Stage 7 Tanpa Proses *Stemming*.... 111

Tabel 6.27 *Recall, Precision, F_1 measure* pada Stage 8 Tanpa Proses *Stemming*.... 112

Tabel 6.28 *Recall, Precision, F_1 measure* pada Stage 9 Tanpa Proses *Stemming*.... 112

Tabel 6.29 *Recall, Precision, F_1 measure* pada Stage 10 Tanpa Proses *Stemming*.. 112

Tabel 6.30 Evaluasi Rata-rata Efektifitas Sistem Menggunakan Proses *Stemming* . 113

Tabel 6.31 Evaluasi Rata-rata Efektifitas Sistem Menggunakan Tanpa *Stemming* . 113