

SISTEM ANALISIS SENTIMEN PRODUK TELKOMSEL DARI OPINI TWITTER BERBAHASA INDONESIA MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

Muhamad Roni Firdaus
09.04.1.1.1.00115

Pembimbing I : Mula'ab, S.Si, M.Kom.
Pembimbing II : Firdaus Solihin, S.Kom., M.Kom.

ABSTRAK

Analisis sentimen otomatis berdasarkan opini masyarakat terhadap produk TELKOMSEL berbahasa Indonesia. Opini masyarakat diambil dari sebuah situs microblogging twitter dengan melakukan crawling data opini dengan keyword produk-produk dari sebuah provider di Indonesia yaitu Telkomsel. Sistem akan melakukan pengklasifikasian tweet menjadi sebuah kategori yaitu positif, negatif atau netral. Sebelum data training dan data testing didapatkan dari pemisahan data hasil crawling menjadi 75% untuk training dan 25% untuk data testing. Metode yang digunakan untuk mengklasifikasikan adalah metode *naïve bayes classifier*. Hasil akhir atau output dari aplikasi ini berupa peta yang menunjukkan pada daerah mana user yang menyatakan opini positif, negatif atau netral. Dengan aplikasi ini sebuah perusahaan dapat menganalisa kepuasan pelanggan terhadap produk yang dimilikinya dan dapat mengetahui daerah mana saja user menyatakan kepuasan atau ketidakpuasan.

Hasil ujicoba pada sistem klasifikasi teks dengan menggunakan data hasil crawling diperoleh akurasi sebesar 63% dan error rate sebesar 37%, hasil ini mungkin kurang memuaskan itu dikarenakan data teks yang ditulis oleh user pada twitter lebih banyak menggunakan kata yang mengandung singkatan.

Kata Kunci : Crawler Twitter, Klasifikasi teks, Naïve Bayes Classifier, peta

KATA PENGANTAR

Puji syukur Penulis panjatkan kepada Allah SWT atas berkat dan rahmat-Nya, Penulis dapat menyelesaikan Tugas Akhir yang berjudul **Sistem Analisis Sentimen Produk Telkomsel Dari Opini Twitter Berbahasa Indonesia Menggunakan Metode *Naïve Bayes Classifier***. Pada kesempatan ini, Penulis ingin menyampaikan terima kasih dan penghormatan yang sebesar-besarnya kepada:

1. Bapak Sholikin dan ibu Mutlikah, Bapak Supardji dan Ibu Siti Zuhriyah tercinta yang telah mendukung segenap jiwa raga dan doa untuk selesainya kuliah dan Tugas Akhir penulis.
2. Bapak Mulaab dan Bapak Firdaus Solihin atas semua bimbingan dan nasehat yang diberikan selama masa kuliah dan penyusunan Tugas Akhir. Penulis tidak dapat memberikan balasan apa pun selain ucapan terimakasih dan doa.
3. Segenap staf dan dosen pengajar di Teknik Informatika Fakultas Teknik Universitas Trunojoyo Madura (UTM).
4. Adik-adikku tersayang Muhamad Irfan Firdaus, Milaturrohmah dan seluruh keluarga yang selalu menemani di kala suka dan menghibur di kala duka.
5. Benik Ucista dan keluarga yang telah memberikan dukungan, semangat dan doa untuk selesainya Tugas Akhir Penulis.
6. Teman-teman kontrakan Kang Karjo, Kang darul, Kang Ardi, Mas Doni, Kang Fazza, Lek Apin, Kakak Fajar, Arry Koala, Afif Tole, Bang Rotua, Ali si Trem, Arif si Suneo, Mahfud si Fudli dan Pak de Thoha yang telah menjadi sahabat selama kuliah.
7. Teman seperjuangan SOLITE'09 yang selalu menemani hari-hari di kampus Trunojoyo Madura dari semester Satu hingga semester 8.
8. Keluarga Besar COBS(Community Of Bojonegoro Student) yang dapat menjadi keluarga saat aku kuliah.
9. Warga laboratorium Informatika, LPM SAINT, UKMF ITC, JARBIS Indonesia yang telah memberikan pengalaman dan pengetahuan baik mengenai informatika atau pun organisasi.

10. Dan semua pihak yang telah membantu dalam penyelesaian Tugas Akhir penulis yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa pengerjaan Tugas Akhir ini tidak luput dari berbagai kesalahan. Oleh karena itu, Penulis memohon maaf atas segala kekurangan yang ada. Dengan segala kerendahan hati, berbagai saran dan kritik yang membangun sangat Penulis harapkan dari rekan-rekan pembaca.

Bangkalan, Juli 2013

P e n u l i s

DAFTAR ISI

	Hal
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	v
LEMBAR PERSEMBAHAN	vii
ABSTRAK.....	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR ISTILAH	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan dan Manfaat.....	2
1.4 Batasan Masalah.....	2
1.5 Metodologi Penelitian.....	3
1.6 Sistematika Penulisan.....	5
BAB II KAJIAN PUSTAKA	7
2.1 Penelitian Sebelumnya	7
2.1.1 Sistem Analisis Opini Microblogging Berbahasa Indonesia	7
2.1.2 Text Mining Dengan Metode Naïve Bayes Classifier Dan	
Support Vector Machines Untuk Sentiment Analysis	7
2.1.3 Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen	
pada Opini Terhadap Tokoh Publik	8
2.1.4 Opinion Mining di Twitter Untuk Costumer Feedback	
Smartphone dengan Pembelajaran Mesin.....	8
2.2 Landasan Teori	8
2.2.1 Data Mining.....	8
2.2.2 Teks Mining	9
2.2.3 Web Mining	10
2.2.4 <i>Preprocessing</i>	11
2.2.5 Metode <i>Ekstraksi</i>	12
2.2.6 Metode Klasifikasi.....	14
2.2.7 Evaluasi Model	17

BAB III PERANCANGAN SISTEM	19
3.1 Deskripsi Umum Sistem.....	19
3.1.1 Gambaran Umum Sistem.....	20
3.2 Persiapan Dokumen.....	22
3.2.1 Data	22
3.2.2 Filtering	23
3.2.3 Stemming	24
3.2.3 Metode Klasifikasi Teks	25
3.3 Perancangan Database	26
BAB IV IMPLEMENTASI SISTEM	37
4.1 Batasan Implementasi	37
4.2 Lingkungan Implementasi	37
4.3 Pembuatan Program	38
4.3.1 Pengaturan XAMPP	38
4.3.2 Crawler Twitter 140dev	38
4.3.3 Pemisahan Data <i>Training</i> dan Data <i>Testing</i>	39
4.3.4 <i>Preprocessing</i>	39
4.3.5 Nilai Probabilitas	40
4.3.6 Proses Klasifikasi	41
4.4 Hasil Implementasi.....	43
4.5 Uji Coba Sistem Klasifikasi Pertama	46
4.6 Uji Coba Sistem Klasifikasi Kedua	47
4.7 Uji Coba Sistem Klasifikasi Ketiga.....	48
4.8 Analisis Hasil Uji Coba.....	49
BAB V PENUTUP.....	51
5.1 Kesimpulan	51
5.2 Saran.....	51
DAFTAR PUSTAKA.....	53
LAMPIRAN	55
RIWAYAT HIDUP	63

DAFTAR GAMBAR

	Hal
Gambar 2.1 Proses iteratif KDD.....	9
Gambar 3.1 Gambaran umum rancangan sistem.....	20
Gambar 3.2 Flowchart Pemrosesan Data Training	23
Gambar 3.3 Flowchart Proses <i>Preprocessing</i>	24
Gambar 3.4 Flowchart Naïve Bayes Classifier	25
Gambar 3.4 ER diagram database	29
Gambar 4.1 Halaman input keyword tweet	43
Gambar 4.2 Halaman hasil crawling.....	44
Gambar 4.3 Halaman data training	44
Gambar 4.4 Halaman Hasil Klasifikasi Teks	45
Gambar 4.5 Halaman Hasil Klasifikasi dalam Peta.....	45
Gambar 4.6 Grafik Perbandingan Uji Coba	50

DAFTAR TABLE

	Hal
Tabel 2.1 Nilai fitur berdasarkan frequency	12
Tabel 2.2 Nilai fitur berdasarkan presence	12
Tabel 2.3 Nilai fitur berdasarkan frequency normalized	13
Tabel 2.4 Nilai fitur berdasarkan tf-idf.....	13
Tabel 2.5 Term document matrix dengan frequency.....	13
Tabel 2.6 Term document matrix dengan presence	14
Tabel 2.7 Term document matrix dengan frequency normalized	14
Tabel 2.8 Term document matrix dengan tf-idf.....	14
Tabel 2.9. Kemunculan term dalam kalimat	16
Tabel 2.10. Tabel Nilai Probabilitas	16
Tabel 2.11. Tabel Confusion Matrix	17
Tabel 3.1 Tabel Database Stopword.....	26
Tabel 3.2 Tabel Database Kata Dasar	27
Tabel 3.3 Tabel Database Data Training	27
Tabel 3.4 Tabel Database Hasil Klasifikasi.....	28
Tabel 3.5 Tabel Database Data Testing.....	28
Tabel 3.6 Tabel Database Preprocessing	29
Tabel 3.7 Tabel Database Preprocessing Training	29
Tabel 3.8 Tabel Database Filtering.....	30
Tabel 3.9 Tabel Database User Tweet	30
Tabel 3.10 Tabel Database Tweet	31
Tabel 3.11 Tabel Database URL Tweet	32
Tabel 3.12 Tabel Database Mention Tweet	32
Tabel 3.13 Tabel Database Cache Tweet	33
Tabel 3.14 Tabel Database User Sistem.....	33
Tabel 3.15 Tabel Nilai Probabilitas Ci	34
Tabel 3.16 Tabel Nilai Probabilitas Wkj	34
Tabel 4.1. Evaluasi Data Hasil Klasifikasi Pertama	46
Tabel 4.1. Evaluasi Data Hasil Klasifikasi Kedua	47
Tabel 4.1. Evaluasi Data Hasil Klasifikasi Ketiga.....	48

DAFTAR ISTILAH

<i>Crawler Twitter</i>	Adalah sebuah cara untuk menghimpun atau mengumpulkan data dari <i>Twitter</i> dengan kata kunci yang telah ditentukan.
<i>Stoword</i>	Kata yang tidak penting atau kata yang tidak digunakan.
<i>Filtering</i>	Proses penghapusan kata yang tidak penting (<i>Stopword</i>)
<i>Stemming</i>	Pemisahan kata dari imbuhan
<i>Term</i>	Kata
<i>Preprocessing</i>	adalah merupakan proses awal mengelola pengolahan data teks tweet sebelum dilakukan proses klasifikasi dan menjadi tampilan peta berdasarkan pengklasifikasian teks
<i>Term Frequency</i>	Nilai kemunculan kata pada sebuah kalimat
<i>Latitude dan Longitude</i>	Koordinat pada peta

DAFTAR LAMPIRAN

Lampiran 1. Data Uji Coba Kategori Positif	55
Lampiran 2. Data Uji Coba Kategori Negatif.....	56
Lampiran 3. Data Uji Coba Kategori Netral	59