

**KLASIFIKASI SENTIMEN PENGUNJUNG YOUTUBE
BERDASAR KOMENTAR MENGGUNAKAN
ALGORITMA ROCCHIO
STUDI KASUS : VIDEO COVERING**

Tugas Akhir



Oleh :

NORBERTUS YUNENDRA ISTY WUSANA

22104866

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA

2015

**KLASIFIKASI SENTIMEN PENGUNJUNG YOUTUBE
BERDASAR KOMENTAR MENGGUNAKAN
ALGORITMA ROCCHIO
STUDI KASUS : VIDEO COVERING**

Tugas Akhir



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh :

NORBERTUS YUNENDRA ISTY WUSANA
22104866

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA**

2015

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

KLASIFIKASI SENTIMEN PENGUNJUNG YOUTUBE BERDASAR KOMENTAR MENGGUNAKAN ALGORITMA ROCCHIO STUDI KASUS : VIDEO COVERING

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 12 Januari 2015



NORBERTUS YUNENDRA ISTY W
22104866

HALAMAN PERSETUJUAN

Judul Skripsi : KLASIFIKASI SENTIMEN PENGUNJUNG
YOUTUBE BERDASAR KOMENTAR
MENGUNAKAN ALGORITMA ROCCHIO
STUDI KASUS : VIDEO COVERING

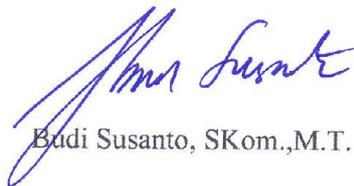
Nama Mahasiswa : NORBERTUS YUNENDRA ISTY W
N I M : 22104866
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2014/2015

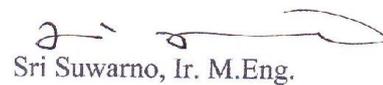
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 12 Januari 2015

DUTA WACANA

Dosen Pembimbing I

Dosen Pembimbing II


Budi Susanto, SKom.,M.T.


Sri Suwarno, Ir. M.Eng.

HALAMAN PENGESAHAN

KLASIFIKASI SENTIMEN PENGUNJUNG YOUTUBE BERDASAR KOMENTAR MENGGUNAKAN ALGORITMA ROCCHIO STUDI KASUS : VIDEO COVERING

Oleh: NORBERTUS YUNENDRA ISTY W / 22104866

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 8 Januari 2015

Yogyakarta, 12 Januari 2015
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom.,M.T.
2. Sri Suwarno, Ir. M.Eng.
3. Prihadi Beny Waluyo, SSi., MT.
4. Willy Sudiarto Raharjo, SKom.,M.Cs



Dekan


(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi


(Gloria Virginia, Ph.D.)

KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Tuhan karena berkat bimbingan-Nya, penulis dapat menyelesaikan skripsi dengan judul “Klasifikasi Sentimen Pengunjung Youtube Berdasar Komentar Menggunakan Algoritma Rocchio Studi Kasus : Video Covering”. Skripsi ini merupakan salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta.

Sejak awal penyusunan skripsi ini hingga selesai, penulis mendapat banyak bantuan dan masukan dari berbagai pihak. Oleh sebab itu, penulis hendak menyampaikan terima kasih kepada :

1. Bapak Budi Susanto, S.Kom., M.T selaku Koordinator Tugas Akhir dan Dosen Pembimbing I, dan Bapak Ir. Sri Suwarno, M.Eng. selaku Dosen Pembimbing II yang telah memberikan bimbingan, arahan, dan masukan.
2. Dekan Fakultas Teknologi Informasi, Ketua Program Studi Teknik Informatika, Wakil Dekan, Dosen, dan Staf Pendukung Akademik di Prodi TI yang telah memberikan bantuan dalam semua proses akademik, dan administrasi selama masa perkuliahan.
3. Bapak Agustinus Mulyadi, Ibu Yohana Sri Sundari, Robertus Setyo, Albertus Windya yang terus memberikan dukungan dan doa.
4. Dias Sihivana Inggita Agriputri yang terus memberikan doa dan dukungan.
5. Para pengisi kuisioner yang telah meluangkan pemikiran dan waktu sehingga memberikan peran dalam terselesainya penelitian ini
6. Pihak-pihak yang telah memberikan doa, dukungan, dan motivasi dalam proses penyelesaian skripsi ini :
 - a. Sahabat dan motivator yang selalu menemani pengerjaan skripsi : Claudia Agraciana Anggraeni Supit, Torry Jatiprakoso
 - b. Teman-teman seperjuangan di proses pengerjaan skripsi : Cornelia Aventina, Deasy Nathalia, Natalia Theofani, Bramasti Pramudyawardani, Stefany Dessy, Theofilus Phetra, Paulus Siagian, William Adi, Elisabet Nur.

- c. Sahabat semasa kuliah : Andar Setiawan Pole, Cornelia Aventina, Yonathan Agung Wijaya, Stefany Dessy, I Gusti Putu Fajar Pala Seraya, Rico Revianto Lemba
 - d. Pengurus Pusat Ikatan Alumni Duta Wacana (IKADUWA), dan Partner Mahasiswa Volunteer Alumni Lounge : Poh Linawati
 - e. Staf Prodi Teknik Informatika, Team dan Mahasiswa Volunteer TIM Akreditasi Prodi Teknik Informatika 2015 : Cornelia Aventina, Stefany Dessy
 - f. Music Director, Manager, Pelatih, Pengurus, dan Anggota Paduan Suara Mahasiswa Duta Voice UKDW, Bass Banget (Dandy, Bryan, Heri, Jordan, Nimo, Teguh, Adi), Tim Pesparawi Ambon, dan Tim Pesparawi Jakarta
 - g. Komunitas “Teman Sebaya” : Theofilus, Cornelia, Jordan, Patrick, Tasya, Christin
 - h. Komunitas “GenkPesparawi” : Tri Samosir, Nala Kusuma, Hendra Wijaya, Kevin Paskalis Datubara, Adi Marianus, Jimmy Gultom, Vincentius Febri Galih, Frivan Triono Pagalu, Naldo, Theresa Vega Toppol, Rangga Aristo Kulas, Mario Gilbert Mandagi, Indra Kurniawan Salama, Resha Reovaldy Tompira, Ezhra Mangera, Muhammad Thamrin Tamtam, Ali Sadikin Fatmaja, Haris Saragih Simarmata
 - i. Teman-teman Paduan Suara Inovatif GKI Gejayan : Putra Andino, Tanta, Agnes, Anna, dan kak Kitty
7. Semua pihak yang telah membantu dalam keseluruhan proses yang tidak dapat disebutkan satu-persatu.

Penulis menyadari bahwa masih ada kekurangan dalam skripsi ini, karena keterbatasan dalam diri penulis. Oleh sebab itu, penulis mengharapkan kritik dan saran yang membangun sehingga skripsi ini dapat semakin berguna bagi semua pihak.

Yogyakarta, Januari 2015

Norbertus Yunendra Isty Wusana
NIM : 22104866

INTISARI

KLASIFIKASI SENTIMEN PENGUNJUNG YOUTUBE BERDASAR KOMENTAR MENGGUNAKAN ALGORITMA ROCCHIO

Studi Kasus : Video Covering

Pengunjung *Youtube* bebas berkomentar dan komentarnya bisa dibaca oleh siapa saja. Komentar dapat dibedakan menjadi tanggapan positif dan negatif dengan membacanya satu per satu. Semakin banyak komentar, maka diperlukan waktu yang lebih lama untuk mendapatkan intisari dari seluruh komentar tersebut.

Untuk mempermudah dalam mengetahui jumlah komentar positif dan negatif, maka dibangun sistem yang mampu mengklasifikasikannya ke kategori positif atau negatif dengan Metode *Rocchio*. Metode *Rocchio* akan mengklasifikasikan komentar yang sudah baku dengan menekankan penggunaan *tf-idf weighting* untuk pembobotan token dan *Euclidean Distance* untuk pengukuran jarak dokumen ke *centroid* kategori.

Sistem akan dapat diakses dan digunakan oleh siapa saja. Dengan adanya penelitian ini, maka akan membantu pengunjung *Youtube* mengklasifikasikan setiap komentar ke dalam klasifikasi komentar positif atau negatif. Metode pengklasifikasian dengan bantuan *tf-idf weighting* dan *Euclidean Distance* ini sangat efektif dan mudah.

Keywords : *tfidf, centroid, euclidean distance, rocchio*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
KATA PENGANTAR	vi
INTISARI	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GRAFIK	xiii
DAFTAR GAMBAR	xiv
DAFTAR LISTING PROGRAM	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Perumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	2
1.5. Metode Penelitian	3
1.6. Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori	6
2.2.1. <i>Text Mining</i>	6
2.2.2. <i>Sentiment Analysis</i>	7
2.2.3. Komentar Pengunjung Video <i>Covering</i> di Youtube	7
A. Teknik <i>Grabbing</i> Komentar	8

B. Deteksi Teks Berbahasa Indonesia menggunakan N-Gram ...	8
2.2.4. <i>Text Classification</i>	9
A. <i>Text Preprocessing</i>	10
B. <i>Text Transformation</i>	10
C. <i>Stopword Removal</i>	11
D. <i>Feature Selection</i> dan Pembobotan Menggunakan Algoritma TF-IDF	11
E. Algoritma <i>Rocchio's Classifier</i>	13
2.2.5. Evaluasi Sistem	14
BAB III ANALISIS DAN PERANCANGAN SISTEM	16
3.1. Spesifikasi Kebutuhan	16
3.1.1. Kebutuhan Fungsional	16
3.1.2. <i>Use Case</i>	18
3.1.3. Spesifikasi Perangkat	19
3.2. Blok Diagram Sistem	20
3.3. Rancangan Proses	21
3.3.1. Proses Klasifikasi Sentimen Pengunjung Youtube Berdasarkan Komentar	21
3.3.2. <i>Grabbing</i> Komentar	23
3.3.3. <i>Text Preprocessing</i>	24
3.3.4. Deteksi Bahasa Indonesia Menggunakan <i>Bigram</i>	25
3.3.5. Tokenisasi dan Penghitungan Nilai TF	27
3.3.6. Penghitungan Nilai DF	28
3.3.7. Penghitungan Nilai TF-IDF	30
3.3.8. Normalisasi	30
3.3.9. <i>Feature Selection</i>	32
3.3.10. <i>Centroid</i> Setiap Kategori	33
3.3.11. Algoritma <i>Rocchio Classifier</i>	34
3.4. Rancangan Database	35
3.4.1. Kamus Data	35

3.4.2. Rancangan Model Relasi	37
3.5. Rancangan Antar Muka	39
3.5.1. Struktur Web	39
3.5.2. Rancangan Halaman	39
A. Halaman <i>Training</i>	39
B. Halaman <i>Testing</i> Bagian Utama	40
C. Halaman <i>Testing</i> Bagian Komentar	41
D. Halaman <i>Testing</i> Bagian <i>Case Folding</i>	41
E. Halaman <i>Testing</i> Bagian Tokenisasi	42
F. Halaman <i>Testing</i> Bagian Deteksi Bahasa Indonesia	43
G. Halaman <i>Testing</i> Bagian Transformasi	43
H. Halaman <i>Testing</i> Bagian <i>Stopword Removal</i>	44
I. Halaman <i>Testing</i> Bagian TF-IDF	45
J. Halaman <i>Testing</i> Bagian Normalisasi	46
K. Halaman <i>Testing</i> Bagian <i>Feature Selection</i>	46
L. Halaman <i>Testing</i> Bagian Pengklasifikasian	47
3.6. Rancangan Pengujian Sistem	48
BAB IV IMPLEMENTASI DAN ANALISIS SISTEM	49
4.1 Implementasi Sistem	49
4.1.1 Antar Muka Sistem	49
A. Halaman Awal	49
B. Halaman <i>Training</i>	49
C. Halaman <i>Testing</i>	51
a. Halaman Utama <i>Testing</i>	51
b. Halaman Daftar Komentar	51
c. Halaman Hasil <i>Case Folding</i>	52
d. Halaman Daftar Token Hasil Tokenisasi	53
e. Halaman Daftar Token Hasil Deteksi Bahasa Indonesia ...	55
f. Halaman Daftar Token Hasil Transformasi	57
g. Halaman Daftar Token Hasil <i>Stopword Removal</i>	58

h. Halaman Hasil Perhitungan TFIDF	59
i. Halaman Hasil Perhitungan Normalisasi Bobot	61
j. Halaman Hasil Proses <i>Feature Selection</i>	63
k. Halaman Bobot Kalimat/Hasil Pengklasifikasian	64
4.1.2 Pengumpulan Dokumen	67
4.2 Analisis dan Evaluasi Sistem	67
4.2.1. Evaluasi <i>Precision Recall</i>	68
A. <i>Feature Selection</i> 10%	68
B. <i>Feature Selection</i> 20%	70
C. <i>Feature Selection</i> 50%	72
D. <i>Feature Selection</i> 70%	74
E. Grafik Hasil Pengujian	75
 BAB V PENUTUP	 78
5.1. Kesimpulan	78
5.2. Saran.....	78
 DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 2.1 Hasil Perhitungan TF-IDF	12
Tabel 2.2 Tabel persyaratan untuk menghitung <i>Precision</i> dan <i>Recall</i>	15
Tabel 3.1 Kamus Data	35
Tabel 4.1 <i>Confusion Matrix</i> Kategori Positif dengan FS 10%	68
Tabel 4.2 <i>Confusion Matrix</i> Kategori Positif dengan FS 20%	70
Tabel 4.3 <i>Confusion Matrix</i> Kategori Positif dengan FS 50%	72
Tabel 4.4 <i>Confusion Matrix</i> Kategori Positif dengan FS 70%	74

DAFTAR GRAFIK

Grafik 4.1 Grafik <i>Precision Recall</i> Kategori Positif	76
Grafik 4.2 Grafik <i>Precision Recall</i> Kategori Negatif	76
Grafik 4.3 Grafik Tingkat Akurasi Sistem	77

DAFTAR GAMBAR

Gambar 3.1 <i>Use Case Diagram</i>	18
Gambar 3.2 Blok Diagram Sistem	20
Gambar 3.3 <i>Flowchart</i> Proses Klasifikasi	22
Gambar 3.4 <i>Flowchart Grabbing</i> Komentar	23
Gambar 3.5 <i>Flowchart Text Preprocessing</i>	24
Gambar 3.6 <i>Flowchart</i> deteksi bahasa Indonesia	26
Gambar 3.7 <i>Flowchart</i> Proses Tokenisasi dan penghitungan nilai TF	27
Gambar 3.8 <i>Flowchart</i> penghitungan nilai DF	29
Gambar 3.9 <i>Flowchart</i> penghitungan nilai TF-IDF	30
Gambar 3.10 <i>Flowchart</i> proses normalisasi bobot	31
Gambar 3.11 <i>Flowchart Feature Selection</i>	32
Gambar 3.12 <i>Flowchart</i> penghitungan <i>Centroid</i>	33
Gambar 3.13 <i>Flowchart</i> pengklasifikasian	34
Gambar 3.14 Relasi Antar Entitas	38
Gambar 3.15 Rancangan struktur web	39
Gambar 3.16 Rancangan Halaman <i>Training</i>	40
Gambar 3.17 Rancangan Halaman <i>Testing</i> Bagian Utama	40
Gambar 3.18 Rancangan Halaman Penampilan Komentar	41
Gambar 3.19 Rancangan Halaman Hasil Proses <i>Case Folding</i>	42
Gambar 3.20 Rancangan Halaman Hasil Proses Tokenisasi	42
Gambar 3.21 Rancangan Halaman Hasil Proses deteksi bahasa Indonesia	43
Gambar 3.22 Rancangan Halaman Hasil Proses Tokenisasi	44
Gambar 3.23 Rancangan Halaman Hasil Proses <i>Stopword Removal</i>	44
Gambar 3.24 Rancangan Halaman Hasil Proses Penghitungan TF-IDF	45
Gambar 3.25 Rancangan Halaman Hasil Proses Normalisasi Bobot	46
Gambar 3.26 Rancangan Halaman Hasil Proses <i>Feature Selection</i>	47
Gambar 3.27 Rancangan Halaman Hasil Proses Pengklasifikasian	48
Gambar 4.1 Implementasi Halaman Awal Sistem	49
Gambar 4.2 Implementasi Halaman Training (bagian a)	50

Gambar 4.3 Implementasi Halaman Training (bagian b)	50
Gambar 4.4 Implementasi Halaman <i>Testing</i>	51
Gambar 4.5 Implementasi Halaman Daftar Komentar	52
Gambar 4.6 Implementasi Halaman Hasil <i>Case Folding</i>	52
Gambar 4.7 Implementasi Halaman Hasil Proses Tokenisasi (bagian a)	54
Gambar 4.8 Implementasi Halaman Hasil Proses Tokenisasi (bagian b)	54
Gambar 4.9 Implementasi Halaman Hasil Proses Deteksi Bahasa Indonesia ...	55
Gambar 4.10 Implementasi Halaman Hasil Proses Transformasi	57
Gambar 4.11 Implementasi Halaman <i>Stopword Removal</i>	58
Gambar 4.12 Implementasi Hasil Proses Perhitungan TF-IDF	60
Gambar 4.13 Implementasi Halaman Hasil Proses Perhitungan Normalisasi	62
Gambar 4.14 Implementasi Halaman Hasil Proses Perhitungan <i>Feature Selection</i>	63
Gambar 4.15 Implementasi Halaman Hasil Proses Pengklasifikasian	65

DAFTAR LISTING PROGRAM

Listing 4.1 <i>Pseudocode</i> Proses <i>Case Folding</i>	53
Listing 4.2 <i>Pseudocode</i> Proses Tokenisasi	55
Listing 4.3 <i>Pseudocode</i> Proses Deteksi Bahasa Indonesia	56
Listing 4.4 <i>Pseudocode</i> Proses Transformasi	58
Listing 4.5 <i>Pseudocode</i> Proses <i>Stopword Removal</i>	59
Listing 4.6 <i>Pseudocode</i> Proses Perhitungan TF-IDF	60
Listing 4.7 <i>Pseudocode</i> Proses Perhitungan Normalisasi Bobot	62
Listing 4.8 <i>Pseudocode</i> Proses Perhitungan <i>Feature Selection</i>	63
Listing 4.9 <i>Pseudocode</i> Proses Perhitungan Pengklasifikasian	65

DAFTAR LAMPIRAN

Lampiran A *Source Code*

Lampiran B *Training Links*

Lampiran C *Testing Links*

Lampiran D *Training Positif*

Lampiran E *Training Negatif*

Lampiran F *Hasil Testing*

©UKDW

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Ketertarikan seseorang untuk menyaksikan video milik orang lain berkembang seiring dengan perkembangan *Youtube* sebagai media *video-sharing*. Selanjutnya, muncul interaksi pengunjung atau penonton *Youtube* melalui komentar yang dituliskan. Mereka bebas berkomentar karena video tersebut sudah dipublikasikan oleh pemiliknya dan bisa diakses oleh siapa saja. Pemilik video dapat membedakan tanggapan positif dan negatif pengunjung dengan membaca komentarnya satu per satu. Semakin banyak komentar, maka diperlukan waktu yang lebih lama untuk mendapatkan intisari dari seluruh komentar tersebut.

Untuk mempermudah dan membantu pemilik video mengetahui jumlah komentar positif dan negatif, maka diperlukan sistem yang mampu mengklasifikasikan komentar. Dengan pengklasifikasian ini, pemilik video tersebut dapat dengan mudah menyimpulkan keseluruhan komentar tersebut condong ke nilai positif atau negatif. Metode yang dapat digunakan untuk membantu proses pengklasifikasian adalah Metode *Rocchio*.

Metode *Rocchio* akan dibantu dengan metode kamus yang terlebih dahulu akan membenahi kata-kata tidak baku dalam komentar tersebut menjadi kata-kata baku, dengan mencocokkannya dengan kamus yang telah disusun. Metode *Rocchio* akan bekerja untuk mengklasifikasikan komentar yang sudah baku dengan menekankan penggunaan *tf-idf weighting* untuk pembobotan token yang akan digunakan untuk klasifikasi dokumen. Dengan adanya penelitian ini, maka akan membantu pemilik video di *Youtube* dan masyarakat secara umum untuk mengklasifikasikan komentar pengunjung *Youtube* ke dalam klasifikasi komentar positif atau komentar negatif.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah diatas, rumusan masalah dalam penelitian ini adalah berapa besar akurasi yang dihasilkan sistem menggunakan algoritma *Rocchio's Classification* dalam mengklasifikasikan komentar berbahasa Indonesia dari *Youtube* ke dalam kategori positif dan negatif?

1.3 Batasan Masalah

Batasan sistem yang akan dibuat dalam penelitian ini adalah sebagai berikut :

1. Dokumen teks yang digunakan merupakan komentar pengunjung *video covering* di *Youtube*. Beberapa contoh *chanel*-nya adalah :
 - a. www.youtube.com/user/gamal1990
 - b. www.youtube.com/user/MeisandriaGuitara
 - c. www.youtube.com/user/kevinikolass
 - d. www.youtube.com/user/cantikabigael
2. Komentar diambil dengan teknik *grabbing* menggunakan *cURL*, sehingga masukan dalam sistem ini adalah berupa *link* halaman *Youtube* tersebut..
3. Komentar yang akan digunakan adalah komentar berbahasa Indonesia dan akan dideteksi menggunakan algoritma *n-gram*.
4. Pengklasifikasian komentar akan menghasilkan dua kelompok komentar, yaitu komentar positif dan negatif.
5. *Stoplist* yang digunakan adalah data *stopwords* yang diambil dari <http://www.ilc.uva.nl/Research/Reports/MoL-2003-02.text.pdf>

1.4 Tujuan Penelitian

Melalui penelitian ini, maka tujuan yang ingin dicapai oleh penulis adalah :

1. Menghasilkan sistem yang dapat mengklasifikasikan teks berbahasa Indonesia dari komentar pengunjung *Youtube* secara otomatis ke dalam klasifikasi positif dan negatif.

2. Mempermudah pemilik video yang diunggah di *Youtube* mengetahui tanggapan pengunjung itu melalui sistem klasifikasi komentar.

1.5 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah :

1. Studi Pustaka
Studi Pustaka dilakukan dengan mempelajari teori dan pemaparan para ahli dari buku, artikel, jurnal, dan bahan tertulis lainnya yang berhubungan dengan metode yang digunakan untuk *text preprocessing* dan klasifikasi teks.
2. Perancangan Sistem
Tahap perancangan sistem berisi perancangan basis data dan *user interface* yang akan dibangun.
3. Pembangunan Sistem
Tahap ini merupakan tahap pembuatan program dan sistem yang sesuai dengan rancangan yang telah dibuat.
4. Implementasi dan Testing
Tahap implementasi dan testing berisi proses pengujian terhadap program pengklasifikasian teks dengan menampilkan hasil penggolongan teks tersebut.
5. Analisis Hasil Percobaan dan Evaluasi
Tahap analisis dan evaluasi adalah tahap yang dilakukan untuk menganalisis dan mengevaluasi program yang dibangun dalam pengklasifikasian teks.

1.6 Sistematika Penulisan

Penulisan laporan tugas akhir ini dibagi menjadi lima bab, yaitu :

Bab 1, PENDAHULUAN, yang akan memberikan gambaran penelitian ini melalui latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

Bab 2, TINJAUAN PUSTAKA, yang akan dibagi menjadi dua bagian, yaitu tinjauan pustaka, dan landasan teori. Tinjauan pustaka akan menguraikan teori-teori tentang klasifikasi yang menggunakan metode *Rocchio* dari berbagai sumber pustaka yang digunakan. Sementara itu, landasan teori akan memuat semua konsep yang akan dijadikan dasar untuk alur penelitian.

Bab 3, ANALISIS DAN PERANCANGAN SISTEM, akan memberikan penjelasan tentang perancangan sistem yang akan dibangun. Perancangan sistem ini adalah kebutuhan *hardware* dan *software*, perancangan *database*, *use case diagram*, algoritma yang akan digunakan, dan perancangan antarmuka yang akan dibangun.

Bab 4, IMPLEMENTASI DAN ANALISIS SISTEM, akan berisi hasil implementasi dan pembahasan tentang proses pengujian sistem yang dibuat berdasarkan perancangannya. Selanjutnya, akan ada pembahasan dan analisis dari sistem yang telah dibuat dan dijalankan.

Bab 5, KESIMPULAN DAN SARAN, berisi kesimpulan dari hasil penelitian dan pembahasan sistem yang didapat, dan saran untuk mengembangkan sistem menjadi sistem yang lebih kompleks dan lebih baik sesuai dengan tuntutan zaman.

BAB V

KESIMPULAN DAN SARAN

5.1. KESIMPULAN

Berdasarkan hasil pembangunan, implementasi, dan analisis sistem klasifikasi sentimen pengunjung *Youtube* berdasar komentar ini, maka diperoleh kesimpulan sebagai berikut :

1. Sistem pengklasifikasian Rocchio memberikan akurasi tertinggi untuk *feature selection* 20%, yaitu sebesar 75,74627 %.
2. Nilai *Precision* tertinggi untuk kategori Positif adalah 0,834394904 pada *feature selection* 50%, sementara itu untuk kategori Negatif adalah 0,323651452 pada *feature selection* 50%.
3. Nilai *Recall* tertinggi untuk kategori Positif adalah 0,913364055 pada *feature selection* 10%, sementara itu untuk kategori Negatif adalah 0,332103321 pada *feature selection* 70%.
4. Peningkatan *feature selection* cukup mempengaruhi nilai *precision recall* dan keakuratan sistem dalam mengklasifikasikan komentar.
5. Hasil pengklasifikasian dipengaruhi oleh data *training* yang telah dibentuk, semakin banyak data *training* yang digunakan, maka akan membuat hasil pengklasifikasian menjadi lebih baik.

5.2. SARAN

Saran yang diajukan penulis untuk perbaikan dan pengembangan sistem adalah sebagai berikut :

1. Dapat ditambahkan proses *stemming* yang akan membantu proses pengklasifikasian menjadi lebih akurat, dengan memperhatikan kata dasar dari setiap data yang diuji.
2. Dapat ditambahkan metode pendeteksi bahasa Indonesia selain menggunakan perhitungan *bigram* agar hasil pendeteksian menjadi lebih optimal.

DAFTAR PUSTAKA

- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks bahasa Indonesia. *Jurnal Link Universitas Narotama Surabaya*, 1-8
- Fadly, F. (2013). *Grabbing Content Website menggunakan cURL*. Jakarta: Universitas Budi Luhur.
- Feldman, R. a. (2007). *The Text Mining Handbook*. New York: Cambridge University Press.
- Hamzah, A. (2010). Deteksi Bahasa Untuk Dokumen Teks Berbahasa Indonesia. *Seminar Nasional Informatika 2010 UPN Veteran Yogyakarta* , A5-A13.
- Harlian, C. M. (2006). *Text Mining CS 391L : Machine Learning Text Categorization*. Austin: University of Texas.
- Joachims, T. (1997). A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. *Proceeding of the Fourteenth Conference on Machine Learning (ICML)* , 143 - 151
- Konchady, M. (2006). *Text Mining Application Programming*. Boston, Massachusetts: Charles River Media.
- Lumbanraja, F. R. (2013). Sistem Pencarian Data Teks dengan Menggunakan Metode Klasifikasi Rocchio (Studi Kasus:Dokumen Teks Skripsi). *Kumpulan Makalah Seminar Semirata 2013 FMIPA Universitas Lampung* , 217-224.
- Manning, C. D. (2008). *Introduction to Information Retrieval*. New York: Cambridge University Press.
- Padjama, S., & Fatima, S. (2013). Opinion Mining and Sentiment Analysis - An Assessment of Peoples' Belief: A Survey. *International Journal of Ad hoc, Sensor & Ubiquitous Computing* , 21-33.

Triawati, C. (2009, November 5). *Digital Library IT Telkom*. Retrieved January 3, 2014, from Digital Library IT Telkom :<http://digilib.itelkom.ac.id/index.php?view=article&catid=20:informatika&id=590:text-mining&tmpl=component&print=1>

©UKDW