

**PENERAPAN SENTIMENT ANALYSIS PADA
HASIL EVALUASI DOSEN DENGAN
METODE SUPPORT VECTOR MACHINE**

Skripsi



oleh
VALONIA INGE SANTOSO
71120004

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2016

**PENERAPAN SENTIMENT ANALYSIS PADA
HASIL EVALUASI DOSEN DENGAN
METODE SUPPORT VECTOR MACHINE**

Skripsi



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

Disusun oleh

VALONIA INGE SANTOSO

71120004

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2016

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE

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, 3 Juni 2016



VALONIA INGE SANTOSO

71120004

HALAMAN PENGESAHAN

PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE

Oleh: VALONIA INGE SANTOSO / 71120004

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 24 Mei 2016

Yogyakarta, 3 Juni 2016
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. Yuan Lukito, S.Kom., M.Cs.
3. Budi Susanto, S.Kom., M.T.
4. Laurentius Kuncoro Probo Saputra, S.T.,
M.Eng.

DUTA WACANA



Dekan

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

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMAKASIH

Terselesaikannya tugas akhir yang berupa sistem *sentiment analysis* pada hasil evaluasi dosen dan laporan tugas akhir ini Penulis mendapatkan bantuan, saran, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, Penulis ingin mengucapkan terima kasih kepada:

1. Tuhan yang Maha Esa atas rahmatNya sehingga laporan Tugas Akhir ini dapat terselesaikan dengan lancar dan tepat waktu.
2. Kedua orang tua saya yang telah mendukung dan memberi semangat serta kepada kakak Penulis, Octavially Santoso yang juga turut memberikan motivasi dan bantuan kepada Penulis.
3. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. sebagai dosen pembimbing I tugas akhir yang telah meluangkan waktu untuk memberikan bimbingan dan memberikan saran serta masukan mengenai penulisan laporan juga analisis sistem.
4. Bapak Yuan Lukito, S.Kom., M.Cs. sebagai dosen pembimbing II tugas akhir yang telah memberikan waktu secara rutin untuk melakukan konsultasi dan memberikan saran dan masukan mengenai pemrograman sistem, algoritma, serta penyelesaian masalah dengan cara yang lebih sederhana.
5. Rekan satu angkatan yakni Jovani, Ireene, Monika, Susy, Melisa, Laksmi, Dina, Evelin, Siene, Tiffany, Vivi, Monica yang selalu mendukung, memberikan semangat dan bantuan dalam pemecahan masalah yang timbul dalam pengerjaan tugas akhir ini.
6. Pihak – pihak lain yang tidak dapat Penulis sebutkan satu persatu yang berperan secara langsung maupun tidak langsung selama pengerjaan tugas akhir.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya, sehingga Penulis mampu menyelesaikan pembuatan sistem dan laporan tugas akhir dengan judul “Penerapan *Sentiment Analysis* terhadap Hasil Evaluasi Dosen dengan Metode *Support Vector Machine*” dengan baik dan tepat waktu. Penyusunan laporan tugas akhir ini berdasarkan hasil penelitian Penulis yang telah mendapatkan bimbingan dan persetujuan dari kedua dosen pembimbing. Isi dari laporan tugas akhir ini juga telah diuji oleh dosen Penguji pada saat dilaksanakan sidang pendadaran.

Penulis berharap laporan yang telah disusun ini dapat bermanfaat untuk pengembangan wawasan bagi Pembaca. “Tak ada gading yang tak retak” seperti itulah laporan ini. Oleh karena itu Penulis memohon maaf apabila dalam penulisan laporan ini, ada kalimat yang kurang berkenan. Akhir kata, Penulis mengucapkan terima kasih.

Yogyakarta, 13 Mei 2016

Penulis

INTISARI

PENERAPAN *SENTIMENT ANALYSIS* PADA HASIL EVALUASI DOSEN DENGAN MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Universitas Kristen Duta Wacana menerapkan sistem kuesioner *online* yang harus diisi mahasiswa untuk menuangkan opini mahasiswa mengenai kinerja dosennya. Opini yang dimasukkan dapat berupa opini positif, negatif, maupun netral. Sebagai tindak lanjut, opini mahasiswa ini harus direkap dan diklasifikasi apakah opini tersebut bersentimen negatif, positif, atau netral agar dapat dijadikan bahan penilaian atas kinerja dosen tersebut. Pengklasifikasian secara manual oleh manusia membutuhkan waktu yang lama dan tidak efektif. Akan lebih mudah apabila pengklasifikasian ini dapat dilakukan secara otomatis.

Pada penelitian ini, Penulis akan membuat sistem yang menerapkan *sentiment analysis* pada hasil evaluasi dosen dengan metode *support vector machine* (SVM). Penulis meneliti unjuk kerja dari metode SVM berdasarkan nilai akurasi. Untuk mendapatkan akurasi terbaik, Penulis juga meneliti faktor apa saja yang dapat mempengaruhi tingkat akurasi. Faktor yang akan diteliti berupa *range* yang tepat untuk membagi hasil persamaan di setiap kelas sentimen, bobot TF-IDF dari data uji, dan komposisi data latih.

Berdasarkan hasil penelitian Penulis, didapatkan akurasi tertinggi yaitu 67,83%. Melalui beberapa pengujian, juga didapatkan kesimpulan bahwa *range* yang terbaik pada kasus penelitian ini adalah < 0 untuk kelas negatif, 0 untuk kelas netral, dan > 0 untuk kelas positif. Bobot TF-IDF dari data uji tidak memberikan perubahan yang cukup signifikan terhadap nilai akurasi.

Selain itu, komposisi data latih juga merupakan faktor yang mempengaruhi nilai akurasi. Hal ini disebabkan karena tidak semua data dapat dikatakan data latih yang baik. Hanya data yang memenuhi beberapa kriteria sajalah yang dapat dikatakan sebagai data latih yang baik. Penentuan data latih yang baik dilakukan dengan metode *k-fold crossvalidation*.

Kata Kunci: [*Support Vector Machine, Sentiment analysis, Penyelesaian sistem persamaan linear, matrix inverse, k-fold crossvalidation, pembobotan TF-IDF*]

©UKDWN

DAFTAR ISI

HALAMAN JUDUL.....	
PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMAKASIH.....	vi
KATA PENGANTAR	vii
INTISARI.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xvii
BAB 1. PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	3
1.5. Metodologi Penelitian	3
1.5.1. Pengumpulan Data	3
1.5.2. <i>Preprocessing</i>	4
1.5.3. Pembobotan.....	4
1.5.4. Metode Klasifikasi	5
1.5.5. Validasi dan Evaluasi	5
1.6. Sistematika Penulisan.....	5

BAB 2. TINJAUAN PUSTAKA	7
2.1. Tinjauan Pustaka	7
2.2. Landasan Teori	8
2.2.1. <i>Text Mining</i> dan <i>Sentiment Analysis</i>	9
2.2.2. Pembobotan TF-IDF	9
2.2.3. <i>Formalization</i>	11
2.2.4. <i>Primal Form</i>	12
2.2.5. <i>Dual Form</i>	12
2.2.6. <i>Support Vector Machines</i>	13
2.2.7. <i>Hyperplans biased</i> dan <i>unbiased</i>	15
2.2.8. Contoh Kasus Proses Pembobotan TF-IDF dan Penggunaan Metode SVM	15
2.2.9. Evaluasi dan Validasi	18
2.2.10. <i>Text Preprocessing</i>	19
2.2.11. Sistem Evaluasi Dosen Berbasis Kuesioner <i>Online</i>	21
BAB 3. ANALISIS DAN PERANCANGAN SISTEM	23
3.1. Analisis Kebutuhan Sistem	23
3.1.1. Kebutuhan Fungsional	23
3.1.2. Kebutuhan Non – Fungsional	24
3.1.3. Kebutuhan Software dan Hardware	24
3.2. Rancangan Sistem	24
3.2.1. Alur Kerja Sistem	24
3.2.2. <i>Use Case Diagram</i>	27
3.2.3. <i>User Interface</i>	29
3.2.4. <i>Database</i>	35
3.2.5. Contoh Format Penggunaan Fitur <i>Import File</i>	37

3.3. Rancangan Pengujian dan Validasi	39
BAB 4. IMPLEMENTASI DAN ANALISIS SISTEM.....	41
4.1. Implementasi Sistem	41
4.1.1. <i>User Interface</i>	41
4.1.2. <i>Database</i>	46
4.2. Hasil Pengujian.....	52
4.3. Analisis Hasil Pengujian	59
BAB 5. KESIMPULAN DAN SARAN	64
5.1. Kesimpulan.....	64
5.2. Saran	64
DAFTAR PUSTAKA	66
LAMPIRAN A	Lampiran A-1
LAMPIRAN B	Lampiran B-1
LAMPIRAN C	Lampiran C-1

DAFTAR TABEL

Tabel 2.1. Tabel Resume Penelitian.....	8
Tabel 2.2. Tabel Pembobotan TF-IDF	15
Tabel 2.3. Tabel Persamaan <i>Support Vector Machine</i>	16
Tabel 2.4. Tabel Kata Tidak Baku dan Kata Baku	20
Tabel 3.1. Tabel <i>Use Case</i> Memasukkan Kalimat Masukkan Hasil Evaluasi Dosen	28
Tabel 3.2. Tabel <i>Use Case</i> Meng- <i>import File</i> Hasil Evaluasi Dosen	29
Tabel 3.3. Tabel <i>Use Case</i> Pelatihan Data	29
Tabel 3.4. Tabel <i>Confusion Matrix</i>	39
Tabel 4.1. Tabel Hasil Pengujian Pertama dengan Mengubah <i>Range</i> di Setiap Kelas.....	54
Tabel 4.2. Tabel Hasil Pengujian Kedua dengan Dokumen Latih Berjumlah Sama di Setiap Kelasnya.....	57
Tabel 4.3. Tabel Hasil Pengujian Data Latih	58
Tabel 4.4. Tabel Hasil Pengujian Penerapan <i>K-fold Cross Validation</i>	59

DAFTAR GAMBAR

<i>Gambar 2.1. Flowchart Pembobotan dengan TF-IDF</i>	10
<i>Gambar 2.2. Contoh beberapa hyperlane</i>	13
<i>Gambar 2.3. Flowchart Klasifikasi dengan SVM dan Analisis</i>	14
<i>Gambar 2.4. Coincidence Matrix</i>	19
<i>Gambar 2.5. Flowchart Preprocessing</i>	19
<i>Gambar 2.6. Form Kuesioner Online Evaluasi Dosen UKDW</i>	21
<i>Gambar 3.1. Alur Kerja Sistem</i>	27
<i>Gambar 3.2. Use Case Diagram Sistem Evaluasi Dosen</i>	28
<i>Gambar 3.3. User Interface Halaman Awal Sistem</i>	30
<i>Gambar 3.4. User Interface Hasil Sentiment Analysis pada Fitur Input Data</i>	30
<i>Gambar 3.5. User Interface Halaman Login</i>	31
<i>Gambar 3.6. User Interface Halaman Import</i>	31
<i>Gambar 3.7. User Interface Halaman Import Bebas Error Messages</i>	32
<i>Gambar 3.8. User Interface Halaman Isi File pada Fitur Import File</i>	32
<i>Gambar 3.9. User Interface Halaman Hasil Sentiment Analysis pada Fitur Import File</i>	33
<i>Gambar 3.10. User Interface Halaman Latih Data</i>	33
<i>Gambar 3.11. User Interface Daftar Kalimat pada File Input</i>	33
<i>Gambar 3.12. User Interface Daftar Token</i>	34
<i>Gambar 3.13. User Interface Hasil Pembobotan TF-IDF</i>	34
<i>Gambar 3.14. User Interface Halaman Sukses Pelatihan Data</i>	35
<i>Gambar 3.15. User Interface Halaman Sukses Pelatihan Data</i>	35
<i>Gambar 3.16. ER-Diagram Database Sistem</i>	36
<i>Gambar 3.17. Struktur Input .csv Pada Fitur Import File Pada Proses Pelatihan Data</i>	38
<i>Gambar 3.18. Struktur Input .csv untuk Fitur Import File Pada Proses Pengujian Data</i>	38

<i>Gambar 3.19.</i> Pembagian Jumlah Data	40
<i>Gambar 4.1.</i> Halaman Awal <i>Input Data</i>	41
<i>Gambar 4.2.</i> Halaman Hasil <i>Sentiment Analysis</i> pada Modul <i>Testing</i>	42
<i>Gambar 4.3.</i> Halaman <i>Login Sistem</i>	42
<i>Gambar 4.4.</i> Halaman Awal Pelatihan Data	42
<i>Gambar 4.5.</i> Halaman Daftar Kalimat pada Modul Latih Data.....	43
<i>Gambar 4.6.</i> Halaman Hasil Tokenisasi	43
<i>Gambar 4.7.</i> Halaman Bobot	44
<i>Gambar 4.8.</i> Halaman Notifikasi Selesai Pelatihan Data	44
<i>Gambar 4.9.</i> Halaman Notifikasi Selesai Validasi Data.....	45
<i>Gambar 4.10.</i> Halaman Awal Modul <i>Import</i>	45
<i>Gambar 4.11.</i> Halaman Daftar Kalimat pada Modul <i>Import</i>	46
<i>Gambar 4.12.</i> Halaman Hasil <i>Sentiment Analysis</i> pada Modul <i>Import</i>	46
<i>Gambar 4.13.</i> Contoh isi tabel users	47
<i>Gambar 4.14.</i> Contoh isi tabel sinonim	47
<i>Gambar 4.15.</i> Contoh Isi Tabel stopwords	48
<i>Gambar 4.16.</i> Contoh Isi Tabel datalatih.....	48
<i>Gambar 4.17.</i> Contoh Isi Tabel tokenlatih	49
<i>Gambar 4.18.</i> Contoh Isi Tabel bobot	49
<i>Gambar 4.19.</i> Contoh Isi Tabel persamaan	50
<i>Gambar 4.20.</i> Contoh Isi Tabel validasi	50
<i>Gambar 4.21.</i> Contoh Isi Tabel datauji.....	51
<i>Gambar 4.22.</i> Contoh Isi Tabel tokenuji	51
<i>Gambar 4.23.</i> Pemanggilan Fungsi untuk Membuat Matriks.....	52
<i>Gambar 4.24.</i> Isi Fungsi <i>invert()</i>	53
<i>Gambar 4.25.</i> Isi Fungsi <i>multiply()</i>	54
<i>Gambar 4.26.</i> <i>Source code</i> fungsi <i>tokenize</i>	56
<i>Gambar 4.27.</i> Grafik Perbandingan Nilai Akurasi pada Penerapan Perubahan Bobot.....	59
<i>Gambar 4.28.</i> Grafik Perbandingan Nilai (P,P) pada Penerapan Perubahan Bobot	60

<i>Gambar 4.29.</i> Grafik Perbandingan Nilai (N,N) pada Penerapan Perubahan Bobot	60
<i>Gambar 4.30.</i> Grafik Perbandingan Nilai (Nt,Nt) pada Penerapan Perubahan Bobot	61
<i>Gambar 4.31.</i> Grafik Akurasi Hasil Pengujian	61
<i>Gambar 4.32.</i> Ilustrasi Penggunaan Tabel persamaan	62

© UKDW

DAFTAR LAMPIRAN

Lampiran A	A-1
Lampiran B.....	B-1
Lampiran C.....	C-1

© UKDW

INTISARI

PENERAPAN *SENTIMENT ANALYSIS* PADA HASIL EVALUASI DOSEN DENGAN MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Universitas Kristen Duta Wacana menerapkan sistem kuesioner *online* yang harus diisi mahasiswa untuk menuangkan opini mahasiswa mengenai kinerja dosennya. Opini yang dimasukkan dapat berupa opini positif, negatif, maupun netral. Sebagai tindak lanjut, opini mahasiswa ini harus direkap dan diklasifikasi apakah opini tersebut bersentimen negatif, positif, atau netral agar dapat dijadikan bahan penilaian atas kinerja dosen tersebut. Pengklasifikasian secara manual oleh manusia membutuhkan waktu yang lama dan tidak efektif. Akan lebih mudah apabila pengklasifikasian ini dapat dilakukan secara otomatis.

Pada penelitian ini, Penulis akan membuat sistem yang menerapkan *sentiment analysis* pada hasil evaluasi dosen dengan metode *support vector machine* (SVM). Penulis meneliti unjuk kerja dari metode SVM berdasarkan nilai akurasi. Untuk mendapatkan akurasi terbaik, Penulis juga meneliti faktor apa saja yang dapat mempengaruhi tingkat akurasi. Faktor yang akan diteliti berupa *range* yang tepat untuk membagi hasil persamaan di setiap kelas sentimen, bobot TF-IDF dari data uji, dan komposisi data latih.

Berdasarkan hasil penelitian Penulis, didapatkan akurasi tertinggi yaitu 67,83%. Melalui beberapa pengujian, juga didapatkan kesimpulan bahwa *range* yang terbaik pada kasus penelitian ini adalah < 0 untuk kelas negatif, 0 untuk kelas netral, dan > 0 untuk kelas positif. Bobot TF-IDF dari data uji tidak memberikan perubahan yang cukup signifikan terhadap nilai akurasi.

Selain itu, komposisi data latih juga merupakan faktor yang mempengaruhi nilai akurasi. Hal ini disebabkan karena tidak semua data dapat dikatakan data latih yang baik. Hanya data yang memenuhi beberapa kriteria sajalah yang dapat dikatakan sebagai data latih yang baik. Penentuan data latih yang baik dilakukan dengan metode *k-fold crossvalidation*.

Kata Kunci: [*Support Vector Machine, Sentiment analysis, Penyelesaian sistem persamaan linear, matrix inverse, k-fold crossvalidation, pembobotan TF-IDF*]

©UKDW

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Universitas yang baik dan terpercaya selalu memperhatikan perkembangan dan kondisi yang terjadi di universitas tersebut, salah satunya dengan memantau kinerja dari dosen – dosen. Dosen yang baik, bertanggung jawab dan berkompeten patut diberi apresiasi oleh pihak dari universitas, sebaliknya untuk dosen yang kurang berkompeten dapat memperbaiki kompetensinya agar menjadi lebih baik. Salah satu cara untuk mengetahui kompetensi dosen adalah dengan melakukan evaluasi dosen oleh mahasiswa, karena mahasiswa adalah sebagai pihak yang merasakan secara langsung kinerja dari dosen tersebut dan diharapkan dapat memberikan penilaian secara bijaksana dan objektif.

Evaluasi dosen dapat dilakukan dengan cara pemberian kuesioner *online* kepada mahasiswa. Hal ini sudah diterapkan oleh Universitas Kristen Duta Wacana Yogyakarta. Kuesioner ini menampung pendapat mahasiswa tentang dosen yang mengampu mata kuliah yang mereka ambil. Pendapat yang diberikan dapat mencakup segi kedisiplinan, pengajaran, dan lain – lain. Pendapat – pendapat ini kemudian diklasifikasikan ke dalam pendapat bersentimen positif, negatif, atau netral, sehingga dapat diketahui kompetensi dari dosen tersebut.

Sejauh ini, Universitas Kristen Duta Wacana melakukan pengklasifikasian pendapat mahasiswa secara manual. Berdasarkan pengalaman penulis sebagai *volunteer* yang merekap data tersebut, pengklasifikasian secara manual memang menghasilkan data yang akurat karena manusia dapat membedakan dengan tepat apakah kata atau kalimat tersebut bermakna positif, negatif, atau netral, namun hal ini dirasa kurang efisien. Pengklasifikasian secara manual membutuhkan banyak waktu dan tenaga. Alangkah lebih baiknya bila pengklasifikasian dilakukan secara

otomatis (dengan sistem), sehingga pihak perekapan data hanya perlu meng-*input* pendapat – pendapat mahasiswa dan sistem akan menghasilkan *output* berupa pendapat yang sudah diklasifikasikan beserta diagram jumlah pendapat positif, negatif, maupun netral.

Sistem ideal yang dimaksud menggunakan prinsip pengolahan teks yang disebut *sentiment analysis* atau *opinion mining*. Salah satu metode yang dapat digunakan dalam *opinion mining* atau *sentiment analysis* adalah *support vector machines* (SVM). Metode ini memiliki tingkat akurasi yang lebih tinggi dibandingkan metode lain (misalnya: Naive Bayes Classifier, K-Nearest Neighbor, dan lain – lain). Hal ini dibuktikan pada penelitian sebelumnya yang pernah dilakukan oleh Yudi Wibisono (Wibisono, 2005), Fatimah Wulandini dan Anto Satriyo Nugroho (Fatimah , 2009), serta Ni Wayan Saraswati (Saraswati, 2013).

1.2. Perumusan Masalah

Berdasarkan uraian di atas, garis besar rumusan masalah yang telah dihimpun adalah: Bagaimana implementasi dari metode *Support Vector Machine* (SVM) terhadap *sentiment analysis* evaluasi dosen dilihat dari tingkat akurasinya?

1.3. Batasan Masalah

- a. Sistem hanya akan memproses kata dalam bahasa Indonesia baku. Untuk masukan dalam bahasa yang tidak baku atau tidak dalam bahasa Indonesia yang sering muncul, sistem akan berusaha mengubah menjadi bahasa baku dengan tabel sinonim yang dibuat sendiri oleh Penulis (jumlah kata terbatas). Sedangkan jika masukan tersebut tidak sesuai syarat dan jarang muncul, akan diabaikan.
- b. Sistem tidak menggunakan tabel *wordnet* dikarenakan inputan data menggunakan bahasa Indonesia. Sebagai penggantinya, Penulis

menggunakan tabel khusus yang dibuat sendiri agar proses pencarian lebih cepat dan efektif.

- c. Sistem hanya akan mengolah data evaluasi dosen Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana semester gasal tahun ajaran 2014/2015.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk membuat sistem yang dapat melakukan *sentiment analysis* pada opini berbahasa Indonesia sebagai opini positif, negatif, maupun netral dengan metode SVM.

1.5. Metodologi Penelitian

Proses penelitian evaluasi dosen yang menerapkan *sentiment analysis* dengan metode SVM memiliki beberapa langkah dari pengumpulan data hingga analisis implementasinya yang dalam hal ini berupa evaluasi data yang dihasilkan sistem. Langkah – langkah yang harus dilakukan dalam pembuatan sistem antara lain pengumpulan data, *preprocessing*, menyiapkan data latih, pembobotan kata dengan TF-IDF, klasifikasi dengan SVM, menguji akurasi data latih terhadap data uji.

1.5.1. Pengumpulan Data

Sample data evaluasi dosen diambil dari hasil evaluasi dosen FTI UKDW Program Studi Teknik Informatika tahun ajaran 2014/2015 semester gasal. Data evaluasi berisi opini positif, negatif, dan netral dari mahasiswa kepada masing – masing dosen. Untuk selanjutnya data *input* ini akan disebut “*sample data*”. Secara keseluruhan, opini diisikan dalam bahasa Indonesia. Data ini akan dibagi menjadi 2 bagian, diantaranya 70% sebagai data latih dan 30% sebagai data uji. Total data yang akan digunakan sebanyak 437 buah yang terdiri dari opini negatif

233 buah, opini positif 147 buah, dan opini netral 56 buah. Dari 436 buah data, akan dibagi menjadi 307 data latih dan 130 data uji.

1.5.2. *Preprocessing*

Langkah *preprocessing* yang dipilih untuk sistem ini adalah:

1. *Tokenisasi*: Tokenisasi adalah proses memotong dokumen *sample data* menjadi token.
2. *Case folding*: proses penyeragaman bentuk huruf (*uppercase* atau *lowercase*) serta penghilangan tanda baca. Dalam hal ini hanya menerima alfabet dari a-z.
3. *Stopword removal*: proses ini dilakukan penyaringan kata hasil dari tokenisasi. Hasil dari *stopword removal* adalah kata – kata penting atau kata – kata yang sering digunakan untuk mempresentasikan dokumen, sedangkan kata – kata seperti “yang”, “di”, “ke”, dan lain sebagainya akan dibuang karena dianggap tidak merepresentasikan isi dokumen. *Stopword removal* juga bertujuan untuk mengurangi jumlah kata sehingga dapat memproses data lebih cepat.
4. *Filtering*: bertujuan hampir sama dengan *stopword removal* yaitu menyaring kata hasil tokenisasi, namun untuk kasus penelitian ini proses *filtering* akan menyaring kata dalam bahasa Indonesia saja yang terdapat di kamus KBBI dan sesuai dengan EYD (bahasa baku), sisanya diabaikan (Feizar, Indrianti, & Yudistira, 2014).

1.5.3. **Pembobotan**

Metode pembobotan yang digunakan dalam penelitian ini adalah *Term Frequency – Inverse Document Frequency* (TF-IDF). Pembobotan TF – IDF menunjukkan bobot setiap kata pada setiap dokumen.

1.5.4. Metode Klasifikasi

Proses klasifikasi menggunakan *support vector machines* dimulai dengan mengubah teks menjadi data vektor. Vektor dalam penelitian ini memiliki dua komponen yaitu `word_id` dan `bobot`. SVM mencoba untuk menemukan garis terbaik (*hyperplane*) yang membagi semesta data menjadi dua kelas, dan kemudian mengklasifikasikan dokumen uji berdasarkan di sisi mana dari garis tersebut mereka muncul.

Garis terbaik adalah garis yang dengan tegas membagi antar kelas dan memiliki margin terbesar terhadap contoh titik pelatihan terdekat di setiap kelas. Oleh karena itu, vektor yang dianggap penting adalah vektor yang menentukan margin tersebut. Vektor – vektor ini disebut *support vector*, dan merupakan kombinasi dari vektor – vektor yang memberikan keputusan fungsi (kelas atau bukan kelas) untuk pengklasifikasian SVM (Saraswati, 2011).

1.5.5. Validasi dan Evaluasi

Proses validasi sistem menggunakan data validasi yang berupa seluruh dokumen latih untuk menentukan *term* yang paling sering muncul di setiap kelas dan menggunakan data latih berjumlah masing – masing 10% di setiap *class* dari dokumen latih sebagai penentu nilai pembatas (*range*) antara kelas sentimen positif, negatif, dan netral. Proses evaluasi menggunakan tingkat akurasi. Parameter untuk menghitung akurasi didapatkan dari tabel *confusion matrix* 3 x 3. Penjelasan lebih rinci mengenai proses validasi dan perhitungan akurasi akan dijelaskan pada Bab 2 dan Bab 3.

1.6. Sistematika Penulisan

Bab 1 penyusunan skripsi diberikan uraian mengenai latar belakang dilakukannya penelitian mengenai *sentiment analysis* pada evaluasi dosen dan alasan dipilihnya metode *support vector machine* pada penelitian ini. Selanjutnya terdapat uraian mengenai permasalahan yang menjadi inti dari penelitian beserta

jangkauan dari penelitian yang sedang dilakukan guna memperjelas fokus dari permasalahan. Tujuan dari segi ilmiah maupun manfaat bagi pihak pengguna juga diuraikan dalam subbab tujuan penelitian. Pada akhir Bab 1, diuraikan metode penelitian yang merupakan langkah – langkah dilakukannya penelitian dan sistematika penulisan yang berisi penjelasan singkat dari isi setiap bab.

Bab 2 berisikan tinjauan pustaka dan landasan teori. Pada bagian tinjauan pustaka diuraikan berbagai teori yang didapatkan dari berbagai sumber pustaka yang digunakan untuk penyusunan Tugas Akhir misalnya rincian penelitian yang pernah dilakukan sebelumnya dan berkaitan dengan penelitian ini, sedangkan landasan teori memuat penjelasan tentang konsep dan prinsip utama yang diperlukan untuk memecahkan masalah riset. Selanjutnya pada Bab 3 mencakup analisis teori–teori yang digunakan dan bagaimana menterjemahkannya ke dalam suatu sistem yang hendak dibuat. Pada bab ini memuat materi dan alat yang dipakai di dalam riset, variabel yang digunakan dan data yang akan dikumpulkan, serta uraian cara perancangan atau simulasi yang akan dilakukan.

Bab Implementasi dan Analisis Sistem memuat hasil riset dan pembahasan dari riset yang sifatnya terpadu. Hasil riset ditunjukkan dalam bentuk tabel, grafik, dan gambar, sedangkan pembahasan berisi hasil yang diperoleh berupa penjelasan secara kualitatif, kuantitatif, atau statistis. Pada Bab Kesimpulan dan Saran berisi pernyataan singkat dan tepat yang dijabarkan dari hasil analisis kegiatan riset dalam penyusunan skripsi. Selain itu Bab 5 juga berisi metode atau langkah dalam riset yang belum dilakukan namun dirasa akan memperbaiki kinerja sistem jika langkah atau metode tersebut dilaksanakan pada riset mendatang.

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil analisis penelitian yang telah dibahas pada Bab 4, maka dapat ditarik kesimpulan sebagai berikut:

Sistem sudah dapat melakukan *sentiment analysis* dengan metode SVM terhadap hasil evaluasi dosen FTI UKDW Program Studi Teknik Informatika tahun ajaran 2014/2015 semester gasal. Dengan metode SVM, sistem dapat melakukan *sentiment analysis* dengan menggunakan 3 kelas sentiment (positif, negatif, netral). Akurasi tertinggi SVM pada sistem ini yaitu 67,83%. Akurasi tertinggi dicapai sistem yang menggunakan dataset 6 sebagai data latih. Semua hasil ini dicapai pada sistem yang tidak menerapkan perubahan bobot pada dokumen uji dan menggunakan *range* > 0 untuk kelas sentimen positif, < 0 untuk kelas sentimen negatif, dan 0 untuk kelas sentimen netral.

5.2. Saran

Sistem pengklasifikasian ini sangat memungkinkan untuk dilakukan pengembangan lebih lanjut sesuai kebutuhan yang terus bertambah, sehingga dapat meningkatkan akurasi sistem. Saran yang diajukan Penulis dalam pengembangan sistem kedepannya adalah sebagai berikut:

- a. Memperkaya variasi *term* yang ada di tabel sinonim dan menambah maupun mengurangi isi dari tabel *stopword* agar tidak terjadi terhapusnya kata yang merupakan ciri khas dari suatu *class* sentimen.
- b. Menggunakan metode yang lebih baik untuk sistem penyelesaian persamaan linear agar terbentuk *hyperplane* yang lebih akurat. Akan lebih baik bila

menggunakan algoritma yang dapat membentuk matriks *inverse* pada semua matriks (tidak hanya matriks persegi) agar sistem dapat mengatasi kasus jumlah dokumen latih lebih banyak dari jumlah *term* unik pada dokumen latih.

- c. Pada penelitian ini, proses *sentiment analysis* dilakukan hanya berdasarkan pembobotan yang diukur dari angka kemunculan dan mengabaikan makna kata. Akan lebih baik bila pada penelitian selanjutnya diterapkan juga kedekatan suatu kata dengan kata yang lain (*semantic*). Misalnya mengkolaborasi tabel *sinonim* dengan tabel *wordnet*.
- d. Berdasarkan hasil analisis, didapatkan hasil bahwa *range* pembatas antar kelas sentiment bisa saja berubah. Akan lebih baik bila sistem memiliki fitur yang memperbolehkan pengguna mengatur sendiri *range* tersebut.

© UTKD

DAFTAR PUSTAKA

- Adikara, P. P. (2012, November 17). *Kamus Kata Dasar dan Stopword List Bahasa Indonesia*. Retrieved from <http://hikaruyuuki.lecture.ub.ac.id/http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/>
- Barber, I. (2010, Januari 1). *Bayesian Opinion Mining*. Retrieved from [phpir.com: http://phpir.com/bayesian-opinion-mining](http://phpir.com/http://phpir.com/bayesian-opinion-mining)
- Castagnetto, J. M. (2010, Oktober 5). *Package Information: Math_Matrix*. Retrieved from [pear.php.net: https://pear.php.net/package/Math_Matrix/download](https://pear.php.net/https://pear.php.net/package/Math_Matrix/download)
- Croft, W. B., Metzler, D., & Strohman, T. (2015). *Search Engines Information Retrieval in Practice*. Pearson Education, Inc.
- Fatimah . (2009). Text Classification Using Support Vector Machine for Webmining Based Spatio Temporal Analysis of the Spread of Tropical Disease. *International Conference on Rural Information and Communication Technology 2009*, (hal. 189-192). Jakarta.
- Feizar, F. H., Indrianti, & Yudistira, N. (2014). *Analisis Sentimen Opini Film Berbahasa Indonesia Berbasis Kamus Menggunakan Metode Neighbor-Weighted K-Nearest Neighbor*. Universitas Brawijaya, Teknik Informatika. Malang: Universitas Brawijaya.
- Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines*. Kluwer Academic.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press.

- Novantirani, A., Sabariah, M. K., & Effendy, V. (n.d). *Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine*. Universitas Telkom, Teknik Informatika. Bandung: Universitas Telkom.
- Putranti, N. D., & Winarko, E. (2014, January 15). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS*, 8, 91-100.
- Rochadiani, T. H., & Santosa, R. G. (2014, Oktober 10). *Penyelesaian Sistem Persamaan Linear (SPL)*. Yogyakarta, Yogyakarta, Indonesia.
- Saraswati, N. W. (2011). *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis*. Universitas UDAYANA, Teknik Elektro. Denpasar: Universitas UDAYANA.
- Saraswati, N. W. (2013). Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis. *Seminar Nasional Sistem Informasi Indonesia* (hal. 585 - 591). Denpasar: Sesindo.
- Wibisono, Y. (2005, Agustus 20). *Klasifikasi Berita Berbahasa Indonesia Menggunakan Naive Bayes Classifier*. Dipetik Oktober 07, 2015, dari academia.edu:
http://www.academia.edu/9702863/Klasifikasi_Berita_Berbahasa_Indonesia_menggunakan_Naive_Bayes_Classifier
- Wicaksono, A. F. (2011, Januari 20). *My Live Journals*. Dipetik September 15, 2015, dari Apa itu Sentiment Analysis / Opinion Mining?: <http://alfanfarizki.blogspot.co.id/2011/01/apa-itu-sentiment-analysis-opinion.html>