

**IMPLEMENTASI ALGORITMA IB1  
TERHADAP KLASIFIKASI SENTIMEN PADA REVIEW  
FILM**

Skripsi



Oleh:

**OSWIN RAHADIYAN HARTONO**

**71110012**

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI  
INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
2015

# **IMPLEMENTASI ALGORITMA IB1 TERHADAP KLASIFIKASI SENTIMEN PADA REVIEW FILM**

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:

**OSWIN RAHADIYAN HARTONO**

**71110012**

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI  
INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA

2015

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **IMPLEMENTASI ALGORITMA IBI TERHADAP KLASIFIKASI SENTIMEN PADA REVIEW FILM**

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, 16 Juni 2015



**OSWIN RAHADIYAN HARTONO**

71110012

## HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI ALGORITMA IBI TERHADAP  
KLASIFIKASI SENTIMEN PADA REVIEW FILM

Nama Mahasiswa : OSWIN RAHADIYAN HARTONO

N I M : 71110012

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2014/2015

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 16 Juni 2015

Dosen Pembimbing I



Gloria Virginia, S.Kom., MAI, Ph.D.

Dosen Pembimbing II



Antonius Rachmat C., S.Kom., M.Cs.

## HALAMAN PENGESAHAN

### IMPLEMENTASI ALGORITMA IBI TERHADAP KLASIFIKASI SENTIMEN PADA REVIEW FILM

Oleh: OSWIN RAHADIYAN HARTONO / 71110012

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 Juni 2015


Yogyakarta, 17 Juni 2015  
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. Yuan Lukito, S.Kom., M.Cs.
4. Ignatia Dhian E K R, S.Kom, M.Eng



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

Ketua Program Studi  
  
(Gloria Virginia, Ph.D.)

## UCAPAN TERIMAKASIH

Puji dan syukur penulis haturkan kepada Tuhan Yesus Kristus yang telah menganugerahkan kasih dan penyertaan-Nya untuk penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma IB1 terhadap Klasifikasi Sentimen pada Review Film” dengan tepat waktu.

Dengan ini penulis menyadari bahwa penelitian ini banyak menerima saran dan masukan dari berbagai pihak. Oleh karena itu, pada kesempatan ini tidak lupa penulis ingin mengucapkan terima kasih kepada:

1. Antonius Rachmat C, S.Kom., M.Cs. sebagai dosen pembimbing dan koordinator Tugas Akhir dan kepada Gloria Virginia S.Kom., MAI, Ph.D. sebagai dosen pembimbing yang sudah banyak membimbing penulis dalam menyelesaikan penelitian ini.
2. Keluarga terkasih terutama ibu dari penulis, Lili Endangwati yang selalu memberikan dukungan, semangat dan motivasi kepada penulis dalam menyelesaikan penelitian ini.
3. Teman-teman seperjuangan penulis yang telah bersama-sama bekerja keras dan saling memotivasi dalam menyelesaikan penelitian masing-masing.
4. Semua pihak yang telah membantu penulis dalam menyelesaikan penelitian ini baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu per satu.

Akhir kata penulis ingin meminta maaf jika ada kesalahan dalam penyusunan laporan skripsi ini. Terima kasih.

Yogyakarta, 18 Mei 2015

Penulis

## ABSTRAKSI

### IMPLEMENTASI ALGORITMA IB1 TERHADAP KLASIFIKASI SENTIMEN PADA REVIEW FILM

*Review* dapat digunakan dalam dunia perfilman sebagai pertimbangan untuk menentukan apakah sebuah film akan ditonton atau tidak. Dengan semakin banyaknya film yang dirilis, maka semakin banyak pula *review* penonton maupun pengamat film yang beredar di situs-situs perfilman. Kebanyakan *review* tidak dikategorikan berdasarkan isinya secara eksplisit apakah tergolong dalam *review* yang positif atau negatif. Akibatnya, pembaca *review* harus menentukan sendiri apakah *review* yang dibacanya bersifat positif atau negatif.

Adanya klasifikasi sentimen untuk *review* sebuah film akan memberi kemudahan bagi pengguna dalam mencari jenis *review* yang diinginkan. Adapun penelitian ini penulis menggunakan algoritma IB1 dimana teks direpresentasikan dalam vektor dengan menggunakan metode TF-IDF untuk pembobotan, serta komponen WordNet untuk mengekspansi token dengan sinonim nya. Setiap dokumen uji akan dihitung jarak kemiripannya dengan setiap dokumen latih menggunakan *cosine similarity*. Pemilihan kelas dilakukan berdasarkan *similar instances* (atau jumlah  $k$  dalam  $k$ -NN) yang dipakai. Dokumen uji yang terklasifikasi dengan benar akan dimasukkan dalam kumpulan dokumen latih. Dalam penelitian ini penulis melakukan beberapa pengujian evaluasi untuk mengetahui efektivitas klasifikasi algoritma IB1 dengan atau tanpa WordNet.

Sistem klasifikasi yang dikembangkan dapat memberikan akurasi yang cukup tinggi. Rata-rata akurasi tanpa menggunakan WordNet sebesar 72,9% sedangkan rata-rata akurasi dengan menggunakan WordNet menghasilkan nilai sebesar 78,7%. Nilai rata-rata F-Measure yaitu 0,719 untuk klasifikasi tanpa WordNet dan 0.772 untuk klasifikasi menggunakan WordNet

Kata kunci : klasifikasi, review film, IB1, WordNet

## DAFTAR ISI

HALAMAN JUDUL.....	
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
ABSTRAKSI.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR LISTING.....	xv
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Metodologi Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	7
2.2.1 Text Mining.....	7
2.2.2 Klasifikasi.....	8
2.2.3 Klasifikasi Sentimen.....	8
2.2.4 Preprocessing Data.....	9
2.2.4.1 Tokenisasi.....	9
2.2.4.2 Komponen WordNet.....	10
2.2.4.3 Penghapusan Stopword.....	10



2.2.5 Pembobotan TF-IDF.....	10
2.2.6 Algoritma IB1 untuk Klasifikasi Teks.....	11
2.2.7 Tahap Evaluasi.....	13
<b>BAB 3 ANALISIS DAN PERANCANGAN SISTEM.....</b>	<b>16</b>
3.1 Spesifikasi Kebutuhan Perangkat Keras dan Perangkat Lunak.....	16
3.1.1 Spesifikasi Kebutuhan Perangkat Keras.....	16
3.1.2 Spesifikasi Kebutuhan Perangkat Lunak.....	16
3.1.3 Spesifikasi Sistem.....	17
3.2 Arsitektur Sistem.....	18
3.3 Flowchart.....	19
3.4 Kamus Data.....	27
3.5 Skema Diagram Basis Data.....	33
3.6 Rancangan Antarmuka.....	34
3.7 Rancangan Pengujian dan Evaluasi Sistem.....	35
3.7.1 Rancangan Pengujian.....	35
3.7.2 Rancangan Evaluasi.....	36
3.8 Contoh Kasus Klasifikasi.....	38
<b>BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....</b>	<b>48</b>
4.1 Implementasi Sistem.....	48
4.1.1 Antarmuka Sistem.....	48
4.1.2 Profil Korpus.....	54
4.1.3 Pseudocode Sistem.....	56
4.1.3.1 Tahap Preprocessing.....	56
4.1.3.2 Tahap Pembobotan TF-IDF.....	60
4.1.3.3 Tahap Klasifikasi.....	62
4.2 Evaluasi dan Analisis Sistem.....	65
4.2.1 Evaluasi Tanpa WordNet Dengan 5 Similar Instances.....	66
4.2.2 Evaluasi Tanpa WordNet Dengan 7 Similar Instances.....	68
4.2.3 Evaluasi Tanpa WordNet Dengan 9 Similar Instances.....	70
4.2.4 Evaluasi Menggunakan WordNet Dengan 5 Similar Instances...	72
4.2.5 Evaluasi Menggunakan WordNet Dengan 7 Similar Instances...	74

4.2.6 Evaluasi Menggunakan WordNet Dengan 9 Similar Instances...	76
4.3 Grafik dan Perbandingan Hasil Evaluasi.....	81
4.3.1 Grafik Perbandingan Akurasi WordNet dan Non WordNet.....	81
4.3.2 Grafik Perbandingan F-Measure WordNet dan Non WordNet....	82
4.3.3 Grafik Precision, Recall dan F-Measure Kelas Positif.....	84
4.3.4 Grafik Precision, Recall dan F-Measure Kelas Negatif.....	85
4.3.5 K-Fold Validation.....	86
BAB 5 KESIMPULAN DAN SARAN.....	88
5.1 Kesimpulan.....	88
5.2 Saran.....	88
DAFTAR PUSTAKA.....	90
LAMPIRAN.....	

© UKDW

## DAFTAR TABEL

Tabel 2.1	Confusion Matrix.....	15
Tabel 3.1	Tabel Dokumen_latih.....	27
Tabel 3.2	Tabel Preprocessing_latih.....	27
Tabel 3.3	Tabel Stopword.....	28
Tabel 3.4	Tabel Stopword_nama.....	28
Tabel 3.5	Tabel Token.....	29
Tabel 3.6	Tabel Dokumen_uji.....	29
Tabel 3.7	Tabel Preprocessing_uji.....	30
Tabel 3.8	Tabel Film.....	30
Tabel 3.9	Tabel TokenWordNet.....	31
Tabel 3.10	Tabel Klasifikasi.....	32
Tabel 3.11	Daftar Token.....	39
Tabel 3.12	Hasil Token Setelah Penghapusan Stopword.....	39
Tabel 3.13	Relasi WordNet.....	40
Tabel 3.14	Pembobotan TF-IDF Pelatihan.....	41
Tabel 3.15	Pembobotan TF-IDF Pelatihan dengan WordNet.....	41
Tabel 3.16	TF-IDF dengan Pengujian D5.....	42
Tabel 3.17	TF-IDF dengan Pengujian D5 menggunakan WordNet.....	44
Tabel 3.18	TF-IDF dengan Pengujian D6.....	45
Tabel 3.19	TF-IDF dengan Pengujian D6 menggunakan WordNet.....	46
Tabel 3.20	Confusion Matrix.....	47
Tabel 4.1	Daftar Film Referensi.....	54
Tabel 4.2	Confusion Matrix.....	65
Tabel 4.3	Tabel Hasil Klasifikasi Tanpa WordNet 5 <i>Instances</i> .....	66
Tabel 4.4	<i>Confusion Matrix 5 Instances</i> .....	67
Tabel 4.5	Tabel Hasil Klasifikasi Tanpa WordNet 7 <i>Instances</i> .....	68
Tabel 4.6	<i>Confusion Matrix 7 Instances</i> .....	69

Tabel 4.7	Tabel Hasil Klasifikasi Tanpa WordNet 9 <i>Instances</i> .....	70
Tabel 4.8	<i>Confusion Matrix</i> 9 <i>Instances</i> .....	71
Tabel 4.9	Tabel Hasil Klasifikasi WordNet 5 <i>Instances</i> .....	72
Tabel 4.10	<i>Confusion Matrix</i> WordNet 5 <i>Instances</i> .....	73
Tabel 4.11	Tabel Hasil Klasifikasi WordNet 7 <i>Instances</i> .....	74
Tabel 4.12	<i>Confusion Matrix</i> WordNet 7 <i>Instances</i> .....	75
Tabel 4.13	Tabel Hasil Klasifikasi WordNet 9 <i>Instances</i> .....	76
Tabel 4.14	<i>Confusion Matrix</i> WordNet 9 <i>Instances</i> .....	77
Tabel 4.15	Hasil Evaluasi Pengujian Sistem.....	78
Tabel 4.16	Hasil Evaluasi Pengujian Sistem dengan WordNet.....	79
Tabel 4.17	Hasil K-Fold Validation Tanpa WordNet.....	86
Tabel 4.18	Hasil K-Fold Validation Dengan WordNet.....	87

## DAFTAR GAMBAR

Gambar 2.1	Ilustrasi Tahapan <i>Text Mining</i> .....	8
Gambar 2.2	Algoritma IB1.....	12
Gambar 3.1	Arsitektur Klasifikasi Review Film.....	18
Gambar 3.2	<i>Flowchart</i> Sistem.....	19
Gambar 3.3	<i>Flowchart Preprocessing</i> .....	20
Gambar 3.4	<i>Flowchart</i> Tokenisasi.....	21
Gambar 3.5	<i>Flowchart</i> Penghapusan Stopword.....	22
Gambar 3.6	<i>Flowchart</i> Pembobotan TF-IDF.....	23
Gambar 3.7	<i>Flowchart</i> Pelatihan Dokumen.....	24
Gambar 3.8	<i>Flowchart</i> Pengujian Dokumen.....	25
Gambar 3.9	Skema Basis Data.....	33
Gambar 3.10	<i>Form Home</i> .....	34
Gambar 3.11	<i>Form Training</i> .....	34
Gambar 3.12	<i>Form Testing</i> .....	35
Gambar 4.1	Halaman Utama Sistem.....	48
Gambar 4.2	Halaman <i>Review</i> Film.....	49
Gambar 4.3	Halaman Tambah Dokumen Latih.....	50
Gambar 4.4	Halaman <i>Preprocessing</i> Dokumen Latih.....	50
Gambar 4.5	Halaman Tambah Dokumen Uji.....	51
Gambar 4.6	Halaman Pengujian.....	51
Gambar 4.7	Halaman Hasil Pengujian.....	52
Gambar 4.8	Halaman Thesaurus.....	53
Gambar 4.9	Tampilan <i>Database</i> WordNet.....	60
Gambar 4.10	Grafik Akurasi Sistem.....	81
Gambar 4.11	Grafik Regresi Linier Akurasi Sistem.....	82
Gambar 4.12	Grafik F-Measure.....	82
Gambar 4.13	Grafik Regresi Linier F-Measure.....	83

Gambar 4.14	Grafik Evaluasi Kelas Positif Tanpa WordNet.....	84
Gambar 4.15	Grafik Evaluasi Kelas Positif Dengan WordNet.....	84
Gambar 4.16	Grafik Evaluasi Kelas Negatif Tanpa WordNet.....	85
Gambar 4.17	Grafik Evaluasi Kelas Negatif Dengan WordNet.....	85

©UKDW

## DAFTAR LISTING

Listing 4.1	<i>Pseudocode Lowercase</i> .....	56
Listing 4.2	<i>Pseudocode Tokenisasi</i> .....	57
Listing 4.3	<i>Pseudocode Deteksi Token Negatif</i> .....	57
Listing 4.4	<i>Pseudocode Penghapusan Stopword</i> .....	58
Listing 4.5	<i>Pseudocode Implementasi WordNet</i> .....	59
Listing 4.6	<i>Pseudocode Pembobotan TF-IDF</i> .....	60
Listing 4.7	<i>Pseudocode Klasifikasi Dokumen</i> .....	62

©UKDW

## ABSTRAKSI

### IMPLEMENTASI ALGORITMA IB1 TERHADAP KLASIFIKASI SENTIMEN PADA REVIEW FILM

*Review* dapat digunakan dalam dunia perfilman sebagai pertimbangan untuk menentukan apakah sebuah film akan ditonton atau tidak. Dengan semakin banyaknya film yang dirilis, maka semakin banyak pula *review* penonton maupun pengamat film yang beredar di situs-situs perfilman. Kebanyakan *review* tidak dikategorikan berdasarkan isinya secara eksplisit apakah tergolong dalam *review* yang positif atau negatif. Akibatnya, pembaca *review* harus menentukan sendiri apakah *review* yang dibacanya bersifat positif atau negatif.

Adanya klasifikasi sentimen untuk *review* sebuah film akan memberi kemudahan bagi pengguna dalam mencari jenis *review* yang diinginkan. Adapun penelitian ini penulis menggunakan algoritma IB1 dimana teks direpresentasikan dalam vektor dengan menggunakan metode TF-IDF untuk pembobotan, serta komponen WordNet untuk mengekspansi token dengan sinonim nya. Setiap dokumen uji akan dihitung jarak kemiripannya dengan setiap dokumen latih menggunakan *cosine similarity*. Pemilihan kelas dilakukan berdasarkan *similar instances* (atau jumlah  $k$  dalam  $k$ -NN) yang dipakai. Dokumen uji yang terklasifikasi dengan benar akan dimasukkan dalam kumpulan dokumen latih. Dalam penelitian ini penulis melakukan beberapa pengujian evaluasi untuk mengetahui efektivitas klasifikasi algoritma IB1 dengan atau tanpa WordNet.

Sistem klasifikasi yang dikembangkan dapat memberikan akurasi yang cukup tinggi. Rata-rata akurasi tanpa menggunakan WordNet sebesar 72,9% sedangkan rata-rata akurasi dengan menggunakan WordNet menghasilkan nilai sebesar 78,7%. Nilai rata-rata F-Measure yaitu 0,719 untuk klasifikasi tanpa WordNet dan 0.772 untuk klasifikasi menggunakan WordNet

Kata kunci : klasifikasi, review film, IB1, WordNet



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Tinjauan atau *review* seseorang yang ditujukan kepada suatu objek atau produk sangat berpengaruh terhadap penilaian publik atas produk tersebut (Sahoo, 2013). *Review* dapat digunakan dalam dunia perfilman sebagai pertimbangan untuk menentukan apakah sebuah film akan ditonton atau tidak. Dengan semakin banyaknya film yang dirilis, maka semakin banyak pula *review* penonton maupun pengamat film yang beredar di situs-situs perfilman.

Masalahnya, kebanyakan *review* tidak dikategorikan berdasarkan isinya secara eksplisit apakah tergolong dalam *review* yang positif atau negatif. Akibatnya, pembaca *review* harus menentukan sendiri apakah *review* yang dibacanya bersifat positif atau negatif.

Adanya klasifikasi sentimen untuk *review* sebuah film akan memberi kemudahan bagi pengguna dalam mencari jenis *review* yang diinginkan. Adapun penelitian ini ditujukan untuk melakukan klasifikasi dengan menggunakan *review* film yang terdapat dalam situs [rottentomatoes.com](http://www.rottentomatoes.com) dengan menggunakan metode dalam teori *Text Mining*.

Rotten Tomatoes adalah salah satu situs perfilman yang memuat informasi film selain IMDB. Rotten Tomatoes mengelompokan *review* dalam *critic review* dan *audience review*. *Audience review* yaitu komentar dari pengguna yang diberi skor oleh pengguna tersebut sedangkan *critic review* adalah komentar dari *reviewer* yang telah terdaftar. *Critic review* dikelompokan dalam *fresh* atau *rotten review*. Keduanya digunakan dalam penentuan rating film tersebut (<http://www.rottentomatoes.com/about/>). Penentuan kelas masih dilakukan secara manual, kelas dari *review* harus sesuai karena berdampak pada akurasi rating ([http://www.rottentomatoes.com/help\\_desk/critics/](http://www.rottentomatoes.com/help_desk/critics/)). Dalam penelitian ini

dilakukan proses klasifikasi teks secara otomatis berdasarkan isi teks dimana tantangannya adalah dalam mengidentifikasi sentimennya.

Dalam klasifikasi teks terdapat bermacam-macam metode untuk mengklasifikasikan data teks seperti Decision Trees, Naive Bayes, Support Vector Machine, Instance-Based Learning, dan lain sebagainya. Penelitian ini menggunakan klasifikasi model Vector Space dengan metode Instance-Based Learning: IB1 yang menggunakan pembobotan tf-idf dalam pemrosesan dokumen teks. Instance Based Learning adalah metode klasifikasi yang sederhana namun mampu memberikan hasil yang baik (Aha, Kibler., & Albert, 1991). Studi yang dilakukan Vijayarani (2013), Cufoglu (2008) dan Pratomo (2008) menunjukkan bahwa algoritma IBL menghasilkan performa yang baik dalam mengklasifikasi teks kedalam kelas-kelas yang ditentukan.

## 1.2 Rumusan Masalah

Rumusan masalah yang dapat dituliskan dalam tugas akhir ini adalah:

1. Bagaimana akurasi dari klasifikasi yang dihasilkan oleh metode IB1?
2. Bagaimanakah pengaruh hasil klasifikasi dan evaluasi dengan menggunakan WordNet?

## 1.3 Batasan Masalah

Parameter-parameter pembatas dalam penelitian ini adalah sebagai berikut:

- a) Klasifikasi hanya dilakukan terhadap dokumen teks.
- b) Teks *review* film berbahasa Inggris dan diambil dari <http://rottentomatoes.com>
- c) *Review* film dibagi ke dalam 2 kategori yaitu *positive review* dan *negative review*.
- d) Stoplist bersumber dari [http://tonyb.sk/\\_my/ir/stop-words-collection-2014-02-24.zip](http://tonyb.sk/_my/ir/stop-words-collection-2014-02-24.zip)
- e) Stoplist nama bersumber dari <http://outpost9.com/files/WordLists.html>

## 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menerapkan algoritma IB1 secara efektif dalam sebuah sistem klasifikasi sentimen *review* film yang akan mengkategorikan *review* dalam dua kelas, yaitu *positive review* atau *negative review*. Penggunaan WordNet bertujuan untuk meningkatkan kualitas hasil klasifikasi teks.

## 1.5 Metodologi Penelitian

Penelitian mempunyai lima tahapan utama yaitu :

### 1. Studi Literatur

Tahapan awal dari penelitian dan perancangan sistem. Literatur-literatur yang dipelajari berupa buku teks, jurnal dan *e-book* yang membahas tentang *text mining*, *information retrieval*, *sentiment analysis*, *data preprocessing*, algoritma *classifier*, dan evaluasi hasil klasifikasi.

### 2. Pengumpulan Data

Data set yang dibutuhkan berupa dokumen *review* film tahun 2012-2015 yang diambil dari <http://www.rottentomatoes.com> sebanyak 120 dokumen. Data berasal dari genre dan film yang berbeda-beda antara lain *action*, *drama*, *comedy*, *science fiction* dan *suspense*. Untuk setiap film diambil 10 *review* yang terdiri dari 5 *positive review* dan 5 *negative review*. Sebanyak 100 dokumen akan digunakan sebagai dokumen latih dan 20 dokumen akan digunakan sebagai dokumen uji. *Review* diambil dari halaman web dan dimasukkan kedalam *database*.

### 3. Pembuatan Sistem

Setiap dokumen teks akan direpresentasikan dalam sebuah vektor. Elemen dari vektor tersebut adalah bobot setiap *term* yang muncul pada sebuah dokumen. Vektor ini akan menjadi objek algoritma IB1. Dalam proses klasifikasi, *output* dari proses klasifikasi sistem yang dibangun adalah penentuan kelas untuk tiap dokumen uji.

#### 4. Evaluasi

Evaluasi performa sistem dapat diukur dengan menggunakan metode evaluasi *Precision*, *Recall* dan *F-measure*. *Relevant document* yang digunakan untuk mengevaluasi sistem yaitu dokumen uji. Untuk mengetahui relevansi dokumen yang digunakan dapat dilakukan dengan mencocokkannya pada website rottentomates.com dimana review sudah memiliki kelas positif atau negatif. Tujuannya supaya meminimalisir kesalahan klasifikasi pada dokumen uji.

#### 5. Analisis

Analisis dilakukan terhadap hasil evaluasi performa sistem dengan memperhatikan dua hal yaitu : jumlah *instance* dari nilai kemiripan terdekat dan penggunaan WordNet.

### 1.6 Sistematika Penulisan

Penulisan laporan tugas akhir ini dibagi menjadi lima bab yaitu :

Bab 1 yaitu Pendahuluan, yang memberikan gambaran umum tentang penelitian yang dilakukan oleh penulis. Pendahuluan berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

Bab 2 yaitu Tinjauan Pustaka, yang tersusun atas dua bagian utama yaitu tinjauan pustaka dan landasan teori. Tinjauan pustaka menceritakan berbagai teori yang berkaitan dengan klasifikasi teks dan metode IB1 dari penelitian serupa yang dilakukan sebelumnya dan sumber pustaka untuk penyusunan tugas akhir.

Bab 3 yaitu Analisis dan Perancangan Sistem, yang berisi tentang perancangan sistem yang akan dibuat baik mengenai kebutuhan *hardware* dan *software*, spesifikasi sistem, arsitektur sistem, *use case diagram*, *pseudocode* algoritma yang digunakan, skema basis data dan rancangan antarmuka sistem.

Bab 4 yaitu Implementasi dan Analisis Sistem, dimana hasil implementasi, pembahasan tentang pengujian sistem yang dibuat dan analisis dari sistem akan diuraikan dalam bab ini.

Bab 5 yaitu Kesimpulan dan Saran yang berisi tentang kesimpulan dari hasil penelitian yang dilakukan dan saran untuk pengembangan yang lebih baik untuk penelitian berikutnya dalam topik yang serupa.

©UKDW

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Kesimpulan yang didapat dari hasil implementasi dan analisis sistem dalam Bab 4 antara lain :

1. Dalam analisis performa sistem, klasifikasi menggunakan WordNet memiliki nilai rata-rata akurasi sebesar 78,7% dan F-Measure 0,778 atau 77,8%. Sedangkan klasifikasi tanpa menggunakan Wordnet menghasilkan rata-rata akurasi sebesar 72,7% dan F-Measure 0,719 atau 71,9%. Akurasi klasifikasi IB1 pada kasus ini menghasilkan akurasi paling rendah sebesar 60% dan paling tinggi 95%. Hal ini menunjukkan metode *Instance Based Learning* mampu menghasilkan klasifikasi yang baik (diatas 50%) untuk studi kasus dalam penelitian ini. Untuk pengujian secara keseluruhan (1 sampai dengan 99 *instances*), klasifikasi dengan WordNet menunjukkan hasil yang lebih baik dari klasifikasi tanpa WordNet.
2. Sistem klasifikasi dengan menggunakan komponen *database* WordNet menunjukkan hasil yang optimal pada pemakaian 1 - 33 *instances* atau  $k \leq 30\%$  dari jumlah dokumen latih. Sedangkan untuk klasifikasi tanpa WordNet menunjukkan hasil optimal jika pemakaian 1-20 *instances* atau  $k \leq 20\%$  dari dokumen latih.

#### 5.2 Saran

Saran untuk studi penelitian dan pengembangan sistem di kemudian hari adalah:

1. Penggunaan API dalam pengumpulan data secara otomatis sehingga mempersingkat waktu pengumpulan data.
2. Penggunaan *stemming* saat tahap *preprocessing* untuk memperkecil dimensi vektor sehingga sistem memproses lebih cepat.

3. Penggunaan *hash map*, *store procedure* dan PL-SQL untuk mempercepat proses pembacaan data dan efisiensi waktu pemrosesan data.
4. Identifikasi sentimen berdasarkan konteks kalimatnya dengan menerapkan metode-metode yang ada dalam *Natural Language Processing*. Misalnya untuk deteksi frase seperti *life support*, *true love*, *never ever*, *don't even try* dan lain sebagainya.
5. Penerapan *feature selection* untuk penggunaan data dalam jumlah besar sehingga mempersingkat waktu pemrosesan data tanpa mengorbankan hasil klasifikasi karena fitur yang dipilih harus representatif.
6. Optimalisasi penggunaan WordNet, misalnya penggunaan tabel-tabel lain dalam *database* WordNet sehingga hasil klasifikasi lebih baik.
7. Penambahan kelas netral dalam klasifikasi sentimen *review* film.

© UKDW

## DAFTAR PUSTAKA

- Aha, D., Kibler., D., & Albert, M. (1991). Instance-Based Learning.
- Amiri, H., & Chua, T.-S. (2012). Sentiment Classification Using the Meaning of Words. *AAAI* .
- Berry, M. W., & Kogan, J. (2010). *Text Mining : Application and Theory*. New Jersey: John Wiley & Sons.
- Cufoglu, A. d. (2008). Classification Accuracy Performance of Naïve Bayesian (NB), Bayesian Networks (BN), Lazy Learning of Bayesian Rules(LBR) and Instance-Based Learner (IB1) – comparative study. Cairo: University of Westminster.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, 2nd Edition*. San Fransisco: Morgan Kauffman Publishers.
- Jivani, A. G. (2011). A Comparative Study of Stemming Algorithms.
- Liu, B. (n.d.). *Opinion Mining and Summarization*. Retrieved September 4, 2014, from <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>.
- Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval*. New York: Cambridge University Press.
- Sahoo, N. d. (2013). The Impact of Online Product Reviews on Product Returns and Net Sales.
- Tala, F. Z. (2008). A Study of Stemming Effects on Information.
- Vijayarani, S., & Muthulakshmi, M. (2013). Comparative Analysis of Bayes and Lazy Classification Algorithms. *ijarce* .
- Widjojo, E. (2013). Implementasi Rocchio's Classification Dalam Mengkategorikan Renungan Harian Kristen.
- Witten, I. H. (2010). *Text Mining*. Hamilton: University of Waikato.