

**KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS
FILM DENGAN MENGGUNAKAN ALGORITMA ADD-ONE
SMOOTHING NAÏVE BAYES**

Skripsi



oleh

**FENDY CHRISTIAWAN
71110110**

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

**KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS
FILM DENGAN MENGGUNAKAN ALGORITMA ADD-ONE
SMOOTHING NAÏVE BAYES**

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

FENDY CHRISTIAWAN
71110110

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

PERNYATAAN KEASLIAN SKRIPSI


Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS FILM DENGAN MENGGUNAKAN ALGORITMA ADD-ONE SMOOTHING NAÏVE BAYES

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, 26 November 2015


FENDY CHRISTAWAN

71110110



HALAMAN PERSETUJUAN

Judul Skripsi : KLASIFIKASI GENRE FILM BERDASARKAN
SINOPSIS FILM DENGAN MENGGUNAKAN
ALGORITMA ADD-ONE SMOOTHING NAÏVE
BAYES

Nama Mahasiswa : FENDY CHRISTIAWAN

N I M : 71110110

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2015/2016

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 26 November 2015

Dosen Pembimbing I



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

Dosen Pembimbing II



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

HALAMAN PENGESAHAN

KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS FILM DENGAN MENGGUNAKAN ALGORITMA ADD-ONE SMOOTHING NAÏVE BAYES

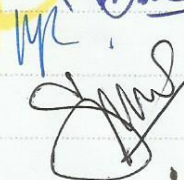
Oleh: FENDY CHRISTIAWAN / 71110110

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 10 Desember 2015

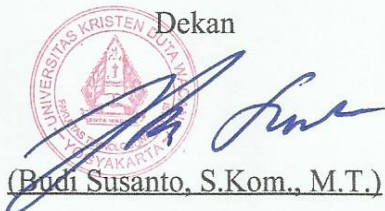
Yogyakarta, 20 Desember 2015
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., S.Kom., M.Cs.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. Rosa Delima, S.Kom., M.Kom.
4. Hendro Setiadi, M.Eng

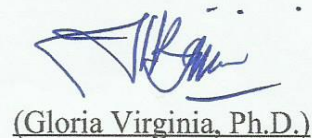


Dekan



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

Ketua Program Studi



(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerah, sehingga penulis dapat menyelesaikan dengan baik dan tepat waktu Tugas Akhir dengan judul Klasifikasi Genre Film berdasarkan Genre Film dengan menggunakan Algoritma *Add One Smoothing Naive Bayes*.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunaannya.

Dalam menyelesaikan pembuatan program dan laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak baik secara langsung maupun secara tidak langsung. Untuk itu, dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Bapak Antonius Rachmat C., S.Kom.,M.Cs selaku dosen pembimbing I, atas bimbingan, petunjuk, saran dan masukan yang diberikan selama pengerjaan Tugas Akhir ini sejak awal hingga akhir.
2. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku dosen pembimbing II, atas bimbingan, petunjuk, saran dan masukan yang diberikan selama pengerjaan Tugas Akhir ini sejak awal hingga akhir.
3. Dosen-dosen Universitas Kristen Duta Wacana yang telah memberikan bimbingan, pengajaran, dan dukungan kepada penulis selama ini.
4. Orang tua dan adik tercinta yang telah memberi dukungan, semangat, motivasi serta dorongan kepada penulis.
5. Evline Denita yang telah memberi semangat, motivasi, serta telah mendukung penulis dalam menyelesaikan Tugas Akhir ini.

6. Para sahabat terkasih “Cah Pap4t” (Hami, Arma, Krisna), Frengky, Hiro, Ko Vandrey, Ko Hendy, teman-teman pelayanan STAR One GBI Keluarga Allah dan sahabat lainnya yang telah memberikan semangat dan dukungan untuk penulis selama mengerjakan Tugas Akhir ini serta yang telah menjadi sahabat yang tulus kepada penulis selama menempuh perkuliahan di Universitas Kristen Duta Wacana.
7. Para teman seperjuangan yang mengambil Tugas Akhir di semester ini, khususnya Greg, Iput, Billy, Neshia, Irawan, dan juga teman-teman lainnya yang telah berjuang bersama, saling mendukung, dan saling mengingatkan selama pengerjaan Tugas Akhir ini.
8. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian. Sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf apabila ada kesalahan baik dalam penyusunan laporan maupun yang pernah penulis lakukan selama proses pembuatan program Tugas Akhir ini. Dan semoga Tugas Akhir ini dapat berguna bagi kita semua.

Yogyakarta, 26 November 2015

Penulis

INTISARI

Klasifikasi Genre Film Berdasarkan Sinopsis Film dengan menggunakan

Algoritma Add One Smoothing Naive Bayes

Imdb merupakan sebuah website yang berisi mengenai berbagai informasi mengenai film, pemain film, rating film dan berbagai hal lainnya yang berkaitan dengan film. Permasalahan yang timbul saat seseorang ingin menentukan dan sekedar membandingkan atau bahkan membuat sebuah sinopsis yang berkaitan dengan kategori tertentu, misalnya dalam genre *Action*, *Comedy*, *Horror*, dan *Romance* adalah apabila seseorang ingin membuat sebuah sinopsis yang tepat agar sesuai dengan keinginannya maka pada sistem informasi inilah bisa sebagai pembanding sinopsis agar sesuai dengan genre filmnya.

Melihat permasalahan tersebut, penulis membangun sebuah sistem yang dapat melakukan analisis terhadap data-data sinopsis film yang ada pada website imdb tersebut sebagai acuan dalam penentuan genre film. Sehingga nantinya sistem tersebut dapat menentukan dan membandingkan genre film dengan menerapkan Algoritma *Add One Smoothing Naive Bayes*, berdasarkan inputan *user* yang berupa sinopsis film.

Sistem klasifikasi genre film menggunakan *Add One Smoothing Naive Bayes* memberikan akurasi tertinggi untuk *feature selection* 40% yaitu sebesar 90% serta memberikan akurasi terendah untuk *feature selection* 10% yaitu sebesar 65%.

Kata kunci : klasifikasi, genre, film, sinopsis, *Add One Smoothing Naive Bayes*, akurasi, *feature selection*

DAFTAR ISI

HALAMAN SAMBUNG	ii
PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
UCAPAN TERIMA KASIH	vi
INTISARI	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR DAN GRAFIK	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Metode 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 <i>Text Mining</i>	7
2.2.2 Klasifikasi Dokumen	7
2.2.3 <i>Text Preprocessing</i>	7
2.2.4 <i>Text Transformation (Feature Generation)</i>	8
2.2.5 Pembobotan TF-IDF (<i>Terms Frequency-Inverse Document Frequency</i>)	8
2.2.6 <i>Feature Selection</i>	9
2.2.7 Algoritma <i>Add One Smoothing Naive Bayes</i>	9
2.3 Evaluasi Sistem	11
2.4 Contoh Kasus Algoritma <i>Add One Smoothing Naive Bayes</i>	12

2.4.1 <i>Tokenization</i>	13
2.4.2 <i>Stopword Removal</i>	13
2.4.3 Pembobotan TF-IDF.....	14
2.4.4 <i>Feature Selection</i>	15
2.4.5 Pengkategorian dengan <i>Add One Smoothing Naive Bayes</i>	15
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	17
3.1 Spesifikasi Kebutuhan Non Fungsional dan Fungsional	17
3.1.1 Spesifikasi Kebutuhan Non Fungsional	17
3.1.1.1 Spesifikasi Batasan Sistem	17
3.1.1.2 Spesifikasi Kebutuhan Perangkat Keras	17
3.1.1.3 Spesifikasi Kebutuhan Perangkat Lunak	18
3.1.1.4 Spesifikasi Sistem	18
3.1.2 Spesifikasi Kebutuhan Fungsional	20
3.1.2.1 Diagram Use Case	20
3.2 Perancangan Proses Sistem	21
3.2.1 <i>Flowchart</i>	21
3.2.1.1 <i>Flowchart</i> Sistem	21
3.2.1.2 <i>Flowchart Preprocessing Dokumen</i>	23
3.2.1.3 <i>Flowchart</i> Tokenisasi	25
3.2.1.4 <i>Flowchart</i> Pembobotan TF-IDF	26
3.2.1.5 <i>Flowchart</i> Normalisasi Pembobotan TF-IDF	28
3.2.1.6 <i>Flowchart</i> Klasifikasi dengan <i>Add One Smoothing</i> Naive Bayes	30
3.3 Perancangan Basis Data	33
3.3.1 Kamus Data	33
3.3.2 Diagram Skema Basis Data.....	34
3.4 Rancangan Antarmuka Sistem	35
3.5 Rancangan Pengujian dan Evaluasi Sistem	39
3.5.1 Rancangan Pengujian	39
3.5.2 Rancangan Evaluasi Sistem	40
3.6 Contoh Kasus Klasifikasi	40

BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	44
4.1 Implementasi Sistem	44
4.1.1 Antarmuka Sistem	44
4.1.2 Pengumpulan Dokumen.....	55
4.1.3 Pseudocode Sistem.....	55
4.1.3.1 Tahap <i>Preprocessing</i>	55
4.1.3.2 Tahap <i>Feature Selection</i>	57
4.1.3.3 Tahap Klasifikasi Dokumen	58
4.2 Evaluasi dan Analisis Sistem.....	59
4.2.1 <i>Feature Selection</i> 10%	60
4.2.2 <i>Feature Selection</i> 20%	62
4.2.3 <i>Feature Selection</i> 30%	64
4.2.4 <i>Feature Selection</i> 40%	66
4.2.5 Hasil dan Grafik Sistem.....	68
4.2.5.1 Grafik <i>Recall Feature Selection</i>	68
4.2.5.2 Grafik <i>Precision Feature Selection</i>	69
4.2.5.3 Grafik <i>Fall-Out Feature Selection</i>	70
4.2.5.4 Grafik <i>Accuracy Feature Selection</i>	70
4.2.6 Analisis Sistem.....	71
BAB 5 KESIMPULAN DAN SARAN	75
5.1 Kesimpulan	75
5.2 Saran	75
DAFTAR PUSTAKA	76

DAFTAR TABEL

Tabel 2.1	Hasil Pembobotan TF-IDF dari Contoh Kasus	14
Tabel 3.1	Kamus Data	33
Tabel 3.2	Hasil Tokenisasi Dokumen.....	41
Tabel 3.3	<i>Test Case</i>	41
Tabel 4.1	Hasil Pengujian dengan FS 10%	60
Tabel 4.2	Hasil <i>Confusion Matrix Feature Selection</i> 10%	61
Tabel 4.3	Hasil Pengujian dengan FS 20%	62
Tabel 4.4	Hasil <i>Confusion Matrix Feature Selection</i> 20%	63
Tabel 4.5	Hasil Pengujian dengan FS 30%	64
Tabel 4.6	Hasil <i>Confusion Matrix Feature Selection</i> 30%	65
Tabel 4.7	Hasil Pengujian dengan FS 40%	66
Tabel 4.8	Hasil <i>Confusion Matrix Feature Selection</i> 40%	67
Tabel 4.9	Hasil Analisis Sistem	68
Tabel 4.10	Perbandingan Token film dengan <i>feature selection</i>	72

DAFTAR GAMBAR DAN GRAFIK

Gambar 2.1	<i>Confusion Matrix</i> untuk Klasifikasi Biner	11
Gambar 2.2	Contoh Soal <i>Algoritma Add One Smoothing Naïve Bayes</i>	15
Gambar 3.1	Use Case Diagram	20
Gambar 3.2	<i>Flowchart</i> Sistem	22
Gambar 3.3	<i>Flowchart Preprocessing</i> Dokumen	23
Gambar 3.4	<i>Flowchart</i> Tokenisasi	25
Gambar 3.5	<i>Flowchart</i> Pembobotan TF-IDF	26
Gambar 3.6	<i>Flowchart</i> Normalisasi Pembobotan TF-IDF	28
Gambar 3.7	<i>Flowchart</i> Klasifikasi dengan <i>Add One Smoothing Naive Bayes</i>	30
Gambar 3.8	Skema Diagram Basis Data	34
Gambar 3.9	Rancangan Halaman <i>Login</i>	36
Gambar 3.10	Rancangan Halaman <i>Home</i>	36
Gambar 3.11	Rancangan Halaman <i>Home Admin</i>	37
Gambar 3.12	Rancangan Halaman Tambah Data Pelatihan	38
Gambar 3.13	Rancangan Halaman Pengujian	39
Gambar 4.1	Halaman <i>Login</i>	44
Gambar 4.2	Halaman <i>Home</i>	44
Gambar 4.3	Halaman <i>Home Admin</i>	45
Gambar 4.4	Halaman Tambah Data Pelatihan	46
Gambar 4.5	Halaman <i>Update Feature Selection</i>	47
Gambar 4.6	Halaman Pengujian	48
Gambar 4.7	Halaman <i>Case Folding</i>	49
Gambar 4.8	Halaman Penghilangan Tanda Baca dan Angka	50
Gambar 4.9	Halaman Tokenisasi	51
Gambar 4.10	Halaman Penghapusan <i>Stopwords</i>	52
Gambar 4.11	Halaman Penghitungan TF	53
Gambar 4.12	Halaman Hasil Klasifikasi	54
Listing 4.1	<i>Pseudocode</i> Tokenisasi dan Penghapusan <i>Stopword</i>	55

Listing 4.2	<i>Pseudocode</i> Perhitungan TF-IDF	56
Listing 4.3	<i>Pseudocode Feature Selection</i>	57
Listing 4.4	<i>Pseudocode</i> Klasifikasi Dokumen	58
Gambar 4.13	Contoh Pengujian dengan FS 30% dengan Hasil Benar	59
Gambar 4.13	Contoh Pengujian dengan FS 10% dengan Hasil Benar	60
Gambar 4.14	Contoh Pengujian dengan FS 10% dengan Hasil Salah	61
Grafik 4.1	Grafik <i>Recall Feature Selection</i>	68
Grafik 4.2	Grafik <i>Precision Feature Selection</i>	69
Grafik 4.3	Grafik <i>Fall-Out Feature Selection</i>	70
Grafik 4.4	Grafik <i>Accuracy Feature Selection</i>	71
Grafik 4.5	Grafik Perbandingan bobot TF-IDF	74

©UKDW

INTISARI

Klasifikasi Genre Film Berdasarkan Sinopsis Film dengan menggunakan

Algoritma Add One Smoothing Naive Bayes

Imdb merupakan sebuah website yang berisi mengenai berbagai informasi mengenai film, pemain film, rating film dan berbagai hal lainnya yang berkaitan dengan film. Permasalahan yang timbul saat seseorang ingin menentukan dan sekedar membandingkan atau bahkan membuat sebuah sinopsis yang berkaitan dengan kategori tertentu, misalnya dalam genre *Action*, *Comedy*, *Horror*, dan *Romance* adalah apabila seseorang ingin membuat sebuah sinopsis yang tepat agar sesuai dengan keinginannya maka pada sistem informasi inilah bisa sebagai pembanding sinopsis agar sesuai dengan genre filmnya.

Melihat permasalahan tersebut, penulis membangun sebuah sistem yang dapat melakukan analisis terhadap data-data sinopsis film yang ada pada website imdb tersebut sebagai acuan dalam penentuan genre film. Sehingga nantinya sistem tersebut dapat menentukan dan membandingkan genre film dengan menerapkan Algoritma *Add One Smoothing Naive Bayes*, berdasarkan inputan *user* yang berupa sinopsis film.

Sistem klasifikasi genre film menggunakan *Add One Smoothing Naive Bayes* memberikan akurasi tertinggi untuk *feature selection* 40% yaitu sebesar 90% serta memberikan akurasi terendah untuk *feature selection* 10% yaitu sebesar 65%.

Kata kunci : klasifikasi, genre, film, sinopsis, *Add One Smoothing Naive Bayes*, akurasi, *feature selection*

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi komputer yang pesat pada masa kini menjadi perhatian utama bagi manusia. Kemajuan teknologi komputer yang pesat ini menimbulkan bermacam-macam efek dalam berbagai bidang kehidupan manusia, baik efek negatif maupun efek positif. Salah satu efek positif dari kemajuan teknologi komputer adalah mempermudah pekerjaan manusia dalam melakukan analisis maupun perhitungan dalam suatu kasus tertentu. Dengan adanya program komputer yang menerapkan suatu algoritma tertentu, pekerjaan perhitungan maupun analisis yang sebelumnya sukar dan lama menjadi lebih praktis dan cepat dibandingkan secara manual pengerjaannya. Contohnya adalah pengelompokan dokumen yang dapat dikelompokkan secara efisien dan cepat dengan menggunakan beberapa metode daripada pengelompokan dokumen secara manual.

Pada penelitian yang pernah dilakukan yaitu mengenai klasifikasi dokumen menggunakan KNN, *K-Nearest Neighbour* (KNN) menggunakan klasifikasi ketetanggaan sebagai prediksi terhadap data baru. Tujuan dari algoritma KNN adalah mengklasifikasikan objek baru berdasarkan atribut dan data training (Irianto, J., 2012). Dalam penelitian ini saya juga akan melakukan klasifikasi namun menggunakan Algoritma *Add One Smoothing Naïve Bayes*, dalam kasus klasifikasi genre film berdasarkan *review*.

Pada kasus pengelompokan data dari sinopsis film ini yaitu berupa dokumen teks, ada banyak metode klasifikasi yang bisa diterapkan salah satunya adalah *Naïve Bayes*. Salah satu teknik yang umum digunakan dalam pemrosesan dokumen teks adalah dengan pembobotan kata-kata yang dianggap penting dalam dokumen, yaitu dengan menghitung *term-frequency* (tf). Akan tetapi, jumlah dokumen dimana sebuah token/kata muncul (df) juga harus dipertimbangkan untuk melihat seberapa penting suatu token dalam dokumen tersebut (*scarcity of tokens*). Dalam hal ini akan dihitung *inverse document frequency* untuk tiap token

(idf). Dari tf dan idf, kita akan mendapatkan *tf-idf weighting* (Robertson, 2005), yaitu suatu formula untuk menghitung bobot hubungan suatu token di dalam dokumen. Pada penelitian ini penulis menggunakan Algoritma *Add One Smoothing Naïve Bayes*, yang merupakan varian lain dari Algoritma *Naïve Bayes* dengan menekankan penggunaan *tf-idf weighting* untuk pembobotan token yang digunakan untuk klasifikasi genre film.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah, akan dirumuskan permasalahan yang memberikan hasil dari algoritma *Add One Smoothing Naïve Bayes*. Secara garis besar, masalah yang dapat dirumuskan sebagai berikut :

1. Seberapa tepat dan akuratkah Algoritma *Add One Smoothing Naïve Bayes* digunakan dalam kasus ini?
2. Ketika menggunakan *frequency-based feature selection*, apakah mempengaruhi ketepatan hasil pada kasus ini?

1.3 Batasan Masalah

Batasan-batasan masalah dalam pembuatan aplikasi ini adalah :

- a. Data pelatihan yang digunakan sebagai proses analisis adalah data sinopsis atau *review* film yang diambil dari website www.imdb.com
- b. Sampel data pelatihan yang digunakan sejumlah 100 sinopsis film. Yang meliputi 4 kategori film yaitu : *Action, Comedy, Horror, Romance* dengan masing-masing 25 sampel sinopsis film. Kategori yang dipilih merupakan kategori yang tunggal (bukan gabungan genre).
- c. Data yang akan diproses berupa sinopsis atau *review* film berbahasa Inggris dengan mengambil 2 paragraf dari sinopsis film tersebut. Hasilnya adalah genre atau kategori film yang menurut sistem paling tepat.

1.4. Tujuan Penelitian

Tujuan dari penelitian Tugas Akhir ini adalah menguji akurasi dari implementasi Algoritma *Add One Smoothing Naïve Bayes* untuk mengkategorikan genre film berdasarkan sinopsis ataupun *review* film.

1.5 Metode Penelitian

Metodologi yang digunakan dalam penelitian ini adalah :

1. Studi Pustaka

Studi Pustaka dilakukan dengan mempelajari teori-teori melalui buku, artikel, jurnal dan bahan lain yang mendukung dan berhubungan dengan *text mining* serta Algoritma *Add One Smoothing Naïve Bayes*, sekaligus metode-metode pendukung lainnya yang dibutuhkan.

2. Pengumpulan data

Tahap ini adalah pengumpulan data berupa sinopsis atau *review* film dari website www.imdb.com yang akan digunakan sebagai data pelatihan dan data pengujian.

3. Perancangan sistem

Tahap ini berisi perancangan basis data dan perancangan antarmuka untuk sistem yang akan dibangun. Untuk perancangan basis data akan menggunakan *tool SQL Server 2008 R2*, dan untuk perancangan desain antarmuka akan menggunakan *tool Pencil* sebagai perkiraan sistem aplikasi yang akan dibuat.

4. Pembangunan sistem

Tahap ini merupakan tahap pembuatan program sesuai dengan rancangan yang telah dibuat pada tahap perancangan sistem atau tahap sebelumnya. Pembuatan program ini sendiri menggunakan *tool Visual Basic 2010*. Pada tahap ini juga membuat program berdasarkan langkah-langkah yang sudah dijelaskan.

5. Implementasi dan testing

Pada tahap ini merupakan pengujian terhadap program, dengan memasukkan inputan berupa sinopsis film yang akan diuji, kemudian judul tersebut akan diklasifikasikan genre yang tepat berdasarkan sinopsis tersebut. Dalam pengujian ini dokumen yang akan diuji akan diproses dalam program yang telah diimplementasikan tahap-tahap mulai dari *tokenization*, *stopword removal*, pembobotan *tf-idf*, *feature selection* dan kemudian masuk pengkategorian dengan algoritma *Add One Smoothing Naïve Bayes*.

6. Analisis hasil percobaan dan evaluasi

Setelah dilakukan pengujian terhadap program, tahap selanjutnya adalah menganalisis keefektifan program dalam melakukan klasifikasi terhadap sinopsis-sinopsis film berdasarkan genre yang tepat.

1.6 Sistematika Penulisan

Penulisan laporan tugas akhir ini dibagi menjadi lima bab yaitu:

Bab 1, Pendahuluan, yang memberikan gambaran umum mengenai apa yang diteliti dalam tugas akhir ini. Pendahuluan berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

Bab 2, Tinjauan Pustaka, yang terdiri dari dua bagian utama, yakni tinjauan pustaka dan landasan teori. Tinjauan pustaka menguraikan berbagai teori mengenai klasifikasi dengan algoritma *Add One Smoothing Naïve Bayes* yang didapatkan dari berbagai sumber pustaka yang digunakan untuk penyusunan tugas akhir. Landasan teori memuat penjelasan tentang konsep dan prinsip utama yang diperlukan untuk memecahkan masalah dalam penelitian. Hanya penjelasan yang berhubungan dengan penelitian yang dilakukan yang akan dicantumkan di sini.

Bab 3, Analisis dan Perancangan Sistem, yang mencakup perancangan sistem yang akan dibuat, yakni mengenai kebutuhan *hardware* dan *software*,

spesifikasi sistem, arsitektur sistem, *use case diagram*, algoritma yang digunakan, skema *database*, dan rancangan antarmuka sistem.

Bab 4, Implementasi dan Analisis Sistem, yang memuat hasil implementasi dan pembahasan mengenai pengujian sistem yang dibuat berdasarkan bab 3, beserta hasil dari sistem yang dijalankan dan analisis dari sistem yang dibuat.

Bab 5, Kesimpulan dan Saran, berisi kesimpulan dari hasil penelitian yang didapatkan dan saran untuk memberikan analisis dan pengembangan yang lebih baik lagi pada penelitian ke depannya dalam topik yang serupa.

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan analisis sistem, maka diperoleh kesimpulan sebagai berikut :

1. Hasil dari penelitian ini membuktikan bahwa Algoritma *Add One Smoothing Naive Bayes* dapat diimplementasikan pada kasus ini.
2. Hasil dari sistem klasifikasi *Add One Smoothing Naive Bayes* memberikan akurasi tertinggi untuk *feature selection* 40% yaitu sebesar 90%. Hasil sistem klasifikasi *Add One Smoothing Naive Bayes* memberikan akurasi terendah untuk *feature selection* 10% yaitu sebesar 65%.
3. *Feature selection* mempengaruhi keakuratan sistem dalam memberikan prediksi genre film. Dalam kasus ini dan dalam rentang *feature selection* 10% hingga 40% menunjukkan peningkatan hasil akurasi, hanya dengan menggunakan *feature selection* 40% sistem sudah dapat menghasilkan nilai keakuratan yang sangat baik.
4. Dari hasil perbandingan genre film *comedy* dan *romance* ada hasil yang tidak sesuai atau tidak tepat. Dimana hasil ini dikarenakan adanya kemiripan token-token yang mengandung dari kedua genre film tersebut dan data dari sinopsis itu sendiri juga mempengaruhi.

5.2 Saran

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

1. Jumlah dokumen pelatihan dapat ditambahkan dengan data yang lebih spesifik sesuai dengan genre filmnya sehingga sistem semakin akurat dalam memberikan prediksi genre film yang tepat.
2. Dapat ditambahkan proses *stemming* dalam bahasa inggris untuk lebih meningkatkan akurasi sistem. Karena dengan proses *stemming* itu sendiri dari banyaknya varian kata akan menjadi kesatuan kata yaitu dengan diambil kata dasar dari kata tersebut.
3. Dapat ditambahkan *stopword list* agar menghasilkan token yang tidak bersifat umum.

DAFTAR PUSTAKA

- Darujati, C. & Gumelar Agustinus B. (2012). PEMANFAATAN TEKNIK SUPERVISED UNTUK KLASIFIKASI TEKS BAHASA INDONESIA. *JURNAL LINK Vol 16/No. 1/Februari 2012*.
- Feldman, Ronen. dan Sanger, James. (2007). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge : Cambridge University Press.
- Grossman, D.A. dan Frieder, O. (2004). *Information Retrieval Algorithms dan Heuristics, 2nd edition*. New York : Springer.
- Irianto, J. (2012). Klasifikasi Dokumen Menggunakan Metode K-Nearest Neighbour (KNN). *Jurnal Penelitian Komputer*, Volume 02, No.1, 2013
- Manning, C.D., Raghavan, P., Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge : Cambridge University Press.
- Nuraini, A., Susanto,B., & Proboyekti,U. (2007). IMPLEMENTASI NAIVE BAYES CLASSIFIER PADA PROGRAM BANTU PENENTUAN BUKU REFERENSI MATAKULIAH. *JURNAL INFORMATIKA, VOLUME 3 NOMOR 2, NOVEMBER 2007*.
- Soucy, P., & Mineau, G. W. (2010). A Simple Feature Selection Method for Text Classification.
- Weiss, S. M., Apte, C., Damerau, F. J., Johnson, D. E., Oles, F. J., Goetz, T., & Hampp, T. (2008). Maximizing text-mining performance. *IEEE Intelligent Systems & Their Applications*. doi:10.1109/5254.784086