

Pengklasifikasian Review Film Menggunakan *Information Gain* dan *Naïve Bayes Classifier*

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya*



Oleh :

Jemz Suzura Ginting

NIM : 09021381419123

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

2018

LEMBAR PENGESAHAN TUGAS AKHIR

Pengklasifikasian Review Film Menggunakan *Information Gain* dan
Naive Bayes Classifier

Oleh :

JEMZ SUZURA GINTING

NIM : 09021381419123

Palembang, November 2018

Pembimbing I

Novi Yusliani, MT
NIP 198211082012122001

Pembimbing II,

Yunita, M.Cs
NIP 198306062015042002

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.

NIP 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis, 15 November 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Jemz Suzura Ginting
NIM : 09021381419123
Judul : Pengklasifikasian Review Film Menggunakan *Information Gain* dan *Naive Bayes Classifier*


1. Ketua Penguji

Hardini Novianti, MT
NIP. 197911012014042002



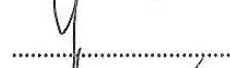
2. Pembimbing I

Novi Yusliani, MT
NIP. 198211082012122001



3. Pembimbing II

Yunita, M.Sc
NIP. 198306062015042002



4. Penguji I

Rusdi Efendi, M.Kom
NIP. 198201022015109101



5. Penguji II

Desty Rodiah, MT
NIP.



Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T
NIP. 197706012009121004



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Jemz Suzura Ginting
NIM : 09021381419123
Program Studi : Teknik Informatika Bilingual
Judul Skripsi : Pengklasifikasian Review Film
Menggunakan *Information Gain* dan
Naive Bayes Classifier
Hasil Pengecekan Software *iThenticate/Turnitin* : **11%**

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, November 2018



(Jemz Suzura Ginting)
NIM. 09021381419123

MOTTO DAN PERSEMBAHAN

“Jadilah Garam dan Terang Dunia.”

(Matius 5:13-16)

Kupersembahkan karya tulis ini kepada :

- Tuhan Yang Maha Esa
- Orang Tua
- Keluarga Besar
- Dosen Pembimbing
- Sahabat-sahabat
- Almamater
- Komunitas

CLASSIFICATION OF FILM REVIEW USING INFORMATION GAIN AND NAIVE BAYES CLASSIFIER

By:

Jemz Suzura Giniting
09021381419123

ABSTRACT

Currently, there have been many reviews of a film published online. One of the important things in a review is the opinion contained in it, that opinion can be in the form of positive or negative opinions. So that need for a system that can automatically produce a classification of positive or negative sentiments on a review. The method used in this research is Information Gain and Naive Bayes Classifier. The Information Gain method is performed for feature selection by selecting the most informative words and the Naive Bayes Classifier method is used to classify a film review. The test results of 10 Fold-Cross Validation in this research by using 180 training data and 20 test data obtained an average value of accuracy of 88% and the average of classification time of 0,0354 seconds.

Keywords: *Sentiment classification, Film review, Information Gain, Naive Bayes Classifier, 10 Fold-Cross Validation.*

PENGLASIFIKASIAN REVIEW FILM MENGGUNAKAN INFORMATION GAIN DAN NAIVE BAYES CLASSIFIER

Oleh:

**Jemz Suzura Giniting
09021381419123**

ABSTRAK

Saat ini telah banyak *review* terhadap suatu film dipublikasikan secara online. Salah satu hal yang penting dalam sebuah *review* adalah opini yang terkandung didalamnya, opini tersebut dapat berupa opini positif atau negatif. Sehingga diperlukannya suatu sistem yang secara otomatis dapat menghasilkan klasifikasi sentimen positif atau negatif terhadap suatu *review*. Metode yang digunakan pada penelitian ini adalah *Information Gain* dan *Naive Bayes Classifier*. Metode *Information Gain* dilakukan untuk seleksi fitur dengan cara memilih kata-kata yang paling informatif dan metode *Naive Bayes Classifier* digunakan untuk mengklasifikasikan suatu *review* film. Hasil pengujian *10 Fold-Cross Validation* pada penelitian ini dengan menggunakan 180 data latih dan 20 data uji didapatkan nilai rata-rata akurasi sebesar 88% dan waktu rata-rata klasifikasi sebesar 0,0354 detik.

Kata Kunci: Klasifikasi sentimen, *review* film, *Information Gain*, *Naive Bayes Classifier*, *10 Fold-Cross Validation* .

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunianya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika pada Fakultas Ilmu Komputer di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini terdapat banyak pihak yang telah memberikan bantuan dan dukungan baik dalam bentuk materil maupun moril. Untuk itu Penulis ingin menyampaikan rasa terima kasih yang tulus kepada :

1. Mama dan kakakku tercinta, Agustina Sitepu dan Defi Lianti Ginting, SE beserta abang sepupu, Billy Joestra Tarigan dan seluruh keluarga besarku yang selalu senantiasa mendoakan, menuntun, memberikan motivasi serta dukungan yang luar biasa kepada penulis;
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer;
3. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika;
4. Ibu Novi Yusliani, M.T selaku dosen pembimbing I dan Ibu Yunita, M.Sc. selaku dosen pembimbing II yang telah memberikan arahan, bimbingan, kritik dan saran kepada penulis untuk menyelesaikan tugas akhir;
5. Bapak Kanda Januar Miraswan, M.T selaku dosen yang telah membantu memberikan arahan, masukan dan saran dalam proses perancangan program pada tugas akhir penulis;

6. Bapak Rusdi Efendi, M.Kom selaku dosen penguji I dan Ibu Desty Rodiah, M.T selaku dosen penguji II yang telah memberikan masukan dan ilmu pengetahuan kepada penulis;
7. Bapak Osvari Arsalan, M.T selaku dosen pembimbing akademik;
8. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika dan staf Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu kepada penulis selama masa kegiatan perkuliahan;
9. Brehan Priangga Ginting, Erland Roy Sembiring, Agripa Bangun, dan Logos Sinuraya sahabat dari kecil hingga sekarang yang selalu memotivasi penulis untuk menjadi pribadi yang lebih baik lagi dan tempat berbagi cerita serta pengalaman yang tiada tara bagi penulis;
10. Anastaysa Pelawi, Okta Bangun, Putri Tarigan, Egy Sembiring dan Gaby Surbakti, sahabat yang senantiasa memberikan doa, dukungan serta motivasi yang tak terhingga bagi penulis saat menyelesaikan tugas akhir ini;
11. Kiki Ananda, sahabat yang selalu bersedia menemani disaat suka maupun duka bagi penulis;
12. Hakim Ali, Harry Dwi Wicaksono, Ali Ridho, dan Hanif Fadillah, sahabat seperjuangan yang telah memberikan bantuan, dukungan dan motivasi bagi penulis dalam penyelesaian tugas akhir.
13. Surrachmad Pratama, Rachmadi Atrima, Sugi Pramana, Aprilizon, Joeyriza Perdana Artoryan, Abdul Hamid Zulni, Theo Valdino, dan seluruh anggota

IF Bilingual 2014 yang telah memberikan banyak bantuan, memori dan bekerja sama dengan penulis;

14. Dela Tarigan A.Md., kakak yang telah membantu dalam memfasilitasi penyelesaian tugas akhir saya;
15. Pengurus PERMATA GBKP Palembang periode 2016-2018 yang telah memberikan bantuan, doa, dukungan dan motivasi yang tak terhingga bagi penulis saat menyelesaikan tugas akhir ini;
16. Himpunan Mahasiswa Teknik Informatika (HMIF) 2016, yang telah memberikan kepercayaan bagi penulis untuk berkontribusi selama 1 periode, terima kasih untuk segalanya dan maaf untuk kekeliruan selama menjabat;
17. Semua pihak lainnya yang tidak dapat disebutkan satu-persatu yang telah banyak berperan bagi penulis terutama dalam penyelesaian tugas akhir ini, terima kasih banyak atas semuanya.

Dalam proses penyelesaian tugas akhir ini penulis sadar bahwa masih terdapat banyak kekurangan dalam pengetahuan dan ilmu yang terbatas.

Penulis akhiri dengan segala hormat, semoga tugas akhir ini dapat memberikan manfaat yang baik bagi kita semua.

Palembang, November 2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
TANDA LULUS UJIAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	vii
DAFTAR ISI	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xviii

BAB 1 PENDAHULUAN

1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan	I-4
1.8 Kesimpulan	I-6

BAB II KAJIAN LITERATUR

2.1 Pendahuluan	II-1
2.2 Analisis Sentimen	II-1
2.3 <i>Preprocessing</i>	II-3
2.4 <i>Information Gain</i>	II-4

2.5 Naive Bayes Classifier	II-8
2.6 Penelitian Lain Yang Relevan.....	II-15
2.7 Kesimpulan.....	II-16

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1
3.3 Metode Pengumpulan Data	III-1
3.3.1 Jenis Data.....	III-1
3.3.2 Sumber Data	III-1
3.3.3 Teknik Pengumpulan Data	III-2
3.4 Tahapan Penelitian	III-2
3.4.1 Menentukan Ruang Lingkup Penelitian	III-2
3.4.2 Menemukan Dasar Teori yang Berkaitan dengan Penelitian .	III-2
3.4.3 Menetapkan Kriteria Pelatihan dan Pengujian Penelitian	III-3
3.4.4 Melakukan Pengembangan Perangkat Lunak dengan menggunakan <i>Rational Unified Process</i>	III-7
3.4.5 Melakukan Analisa Hasil Pengujian Penelitian	III-8
3.4.6 Membuat Kesimpulan Penelitian	III-9
3.5 Metode Pengembangan Perangkat Lunak	III-10
3.6 Penjadwalan Penelitian	III-14
3.7 Kesimpulan	III-19

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-1
4.2.2 Kebutuhan Sistem	IV-2
4.2.3 Analisis dan Desain	IV-4
4.2.3.1 Analisis Perangkat Lunak	IV-4

4.2.3.2	Desain Perangkat Lunak	IV-18
4.3	Fase Elaborasi	IV-30
4.3.1	Pemodelan Bisnis	IV-30
4.3.1.1	Perancangan Data	IV-31
4.3.1.2	Perancangan Antarmuka	IV-31
4.3.2	Kebutuhan Sistem	IV-34
4.3.3	Diagram <i>Sequence</i>	IV-35
4.4	Fase Konstruksi	IV-41
4.4.1	Kebutuhan Sistem	IV-41
4.4.2	Diagram Kelas	IV-41
4.4.3	Implementasi	IV-43
4.4.3.1	Implementasi Kelas	IV-43
4.4.3.2	Implementasi Antarmuka	IV-48
4.5	Fase Transisi	IV-50
4.5.1	Pemodelan Bisnis	IV-50
4.5.2	Kebutuhan Sistem	IV-50
4.5.3	Rencana Pengujian	IV-51
4.5.3.1	Rencana Pengujian <i>Use Case</i> Memasukkan Review	IV-51
4.5.3.2	Rencana Pengujian <i>Use Case</i> Melakukan Pelatihan Review	IV-51
4.5.3.3	Rencana Pengujian <i>Use Case</i> Melakukan Pengujian Review	IV-52
4.5.3.4	Rencana Pengujian <i>Use Case</i> Tampil Hasil Pelatihan	IV-52
4.5.3.5	Rencana Pengujian <i>Use Case</i> Tampil Hasil Pengujian	IV-52
4.5.4	Implementasi	IV-53
4.5.4.1	Pengujian <i>Use Case</i> Memasukkan Review	IV-53
4.5.4.2	Pengujian <i>Use Case</i> Melakukan Pelatihan Review	IV-55

4.5.4.3	Pengujian <i>Use Case</i> Melakukan Pengujian Review	IV-56
4.5.4.4	Pengujian <i>Use Case</i> Melakukan Tampil	
	Hasil Pelatihan	IV-57
4.5.4.5	Pengujian <i>Use Case</i> Melakukan Tampil	
	Hasil Pengujian	IV-58
4.6	Kesimpulan	IV-59

BAB V HASIL DAN ANALISIS PENELITIAN

5.1	Pendahuluan	V-1
5.2	Hasil Percobaan Penelitian	V-1
5.2.1	Hasil dan Analisis Pengujian Parameter <i>Threshold</i>	
	pada Seleksi Fitur	V-2
5.2.2	Hasil dan Analisis Pengujian Pengaruh <i>Information Gain</i>	
	dalam Pengklasifikasian <i>Review Film</i> menggunakan	
	<i>Naive Bayes Classifier</i>	V-4
5.3	Kesimpulan	V-8

BAB VI KESIMPULAN DAN SARAN

6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran	VI-2

DAFTAR PUSTAKA	xix
----------------------	-----

DAFTAR TABEL

	Halaman
Tabel II-1. Contoh Sederhana Perhitungan <i>Information Gain</i>	II-6
Tabel II-2. Contoh Sederhana perhitungan <i>Naïve Bayes Classifier</i>	II-11
Tabel III-1. Tabel Hasil Pengujian	III-8
Tabel III-2. Tabel Hasil Pengujian Keseluruhan Review Film	III-9
Tabel III-3. Tabel Penjadwalan Penelitian dalam Bentuk WBS	III-15
Tabel IV-1. Kebutuhan Fungsional	IV-3
Tabel IV-2. Kebutuhan Non Fungsional	IV-4
Tabel IV-3. Hasil <i>Tokenizing</i> Contoh Dokumen	IV-8
Tabel IV-4. Hasil <i>Filtering</i> Contoh Dokumen	IV-9
Tabel IV-5. Hasil Pembobotan Kata Contoh Dokumen	IV-8
Tabel IV-6. Contoh Perhitungan <i>Information Gain</i> (1)	IV-11
Tabel IV-7. Contoh Perhitungan <i>Information Gain</i> (2)	IV-12
Tabel IV-8. Contoh Perhitungan <i>Naive Bayes Classifiers</i>	IV-13
Tabel IV-9. Definisi Aktor <i>Use Case</i>	IV-19
Tabel IV-10. Definisi <i>Use Case</i>	IV-19
Tabel IV-11. Skenario Memasukkan <i>Review</i>	IV-20
Tabel IV-12. Skenario Melakukan Pelatihan <i>Review</i>	IV-21
Tabel IV-13. Skenario Melakukan Pengujian <i>Review</i>	IV-22
Tabel IV-14. Skenario Tampil Hasil Pelatihan	IV-23
Tabel IV-15. Skenario Tampil Hasil Pengujian	IV-24
Tabel IV-16. Implementasi Kelas	IV-43
Tabel IV-17. Rencana Pengujian <i>Use Case</i> Memasukkan <i>Review</i>	IV-51
Tabel IV-18. Rencana Pengujian <i>Use Case</i> Melakukan Pelatihan <i>Review</i> ...	IV-51
Tabel IV-19. Rencana Pengujian <i>Use Case</i> Melakukan Pengujian <i>Review</i> ..	IV-52
Tabel IV-20. Rencana Pengujian <i>Use Case</i> Tampil Hasil Pelatihan	IV-52
Tabel IV-21. Rencana Pengujian <i>Use Case</i> Tampil Hasil Pengujian	IV-53
Tabel IV-22. Pengujian <i>Use Case</i> Memasukkan <i>Review</i>	IV-54
Tabel IV-23. Pengujian <i>Use Case</i> Melakukan Pelatihan <i>Review</i>	IV-55

Tabel IV-24. Pengujian <i>Use Case</i> Melakukan Pengujian <i>Review</i>	IV-56
Tabel IV-25. Pengujian <i>Use Case</i> Tampil Hasil Pelatihan	IV-57
Tabel IV-26. Pengujian <i>Use Case</i> Tampil Hasil Pengujian.....	IV-58
Tabel V-1. Tabel Hasil Uji Pengaruh <i>Threshold</i>	V-3
Tabel V-2. Tabel Hasil Pengujian Pengaruh <i>Information Gain</i> terhadap <i>Naive Bayes Classifier</i>	V-4
Tabel V-3. Tabel Hasil Data Uji 1	V-5

DAFTAR GAMBAR

	Halaman
Gambar II-1 Tahapan Algoritma Klasifikasi Naive Bayes	II-14
Gambar III-1. Diagram Alur Proses Umum Perangkat Lunak	III-3
Gambar III-2. Diagram Alur Proses <i>Information Gain</i>	III-5
Gambar III-3. Diagram Alur Proses <i>Training</i> pada <i>Naive Bayes Classifier</i> ...	III-6
Gambar III-4. Diagram Alur Proses <i>Testing</i> pada <i>Naive Bayes Classifier</i>	III-7
Gambar III-5. Diagram Tahapan Penelitian.....	III-10
Gambar III-6. Arsitektur <i>Rational Unified Process</i>	III-12
Gambar IV-1. Contoh Proses Perhitungan <i>Naive Bayes Classifier</i>	IV-17
Gambar IV-2. Diagram <i>Use Case</i>	IV-18
Gambar IV-3. Kelas Analisis Memasukkan <i>Review</i>	IV-26
Gambar IV-4. Kelas Analisis Melakukan Pelatihan <i>Review</i>	IV-27
Gambar IV-5. Kelas Analisis Melakukan Pengujian <i>Review</i>	IV-28
Gambar IV-6. Kelas Analisis Tampil Hasil Pelatihan	IV-29
Gambar IV-7. Kelas Analisis Tampil Hasil Pengujian	IV-30
Gambar IV-8. Rancangan Antarmuka Menu Utama	IV-32
Gambar IV-9. Rancangan Antarmuka <i>Form</i> Pelatihan	IV-33
Gambar IV-10. Rancangan Antarmuka <i>Form</i> Pengujian	IV-34
Gambar IV-11. <i>Use Case</i> Memasukkan <i>Review</i>	IV-36
Gambar IV-12. Diagram <i>Sequence</i> Melakukan Pelatihan <i>Review</i>	IV-37
Gambar IV-13. Diagram <i>Sequence</i> Melakukan Pengujian <i>Review</i>	IV-38
Gambar IV-14. Diagram <i>Sequence</i> Tampil Hasil Pelatihan	IV-39
Gambar IV-15. Diagram <i>Sequence</i> Tampil Hasil Pengujian	IV-40
Gambar IV-16. Diagram Kelas Perangkat Lunak	IV-42
Gambar IV-17. Antarmuka Halaman Utama Perangkat Lunak	IV-48
Gambar IV-18. Antarmuka Halaman Pelatihan Perangkat Lunak	IV-49
Gambar IV-19. Antarmuka Halaman Pengujian Perangkat Lunak	IV-49
Gambar V-1. Contoh Data Uji	V-2
Gambar V-2. Analisa Kesalahan Prediksi Dokumen <i>Lima.txt</i>	V-7

DAFTAR LAMPIRAN

	Halaman
Model Probabilistik pada Uji 1	L1-1
Hasil Pengujian 10 <i>Fold Cross Validation</i>	L2-1
Kode Program	L3-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas latar belakang masalah, rumusan masalah, tujuan, manfaat penelitian, dan batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian.

Pendahuluan dimulai dengan penjelasan mengenai proses pengklasifikasian sentimen pada *review* film serta penelitian yang berkaitan dengan pengklasifikasian *review* yang menggunakan *Information Gain* dan *Naïve Bayes Classifier* yang menjadi latar belakang dari penelitian ini.

1.2 Latar Belakang

Saat ini telah banyak *review* atau opini terhadap suatu film dipublikasikan secara online. Hal ini dapat dimanfaatkan sebagai acuan untuk mengetahui kualitas film layak atau tidak untuk ditonton. Banyaknya informasi yang tersedia dalam suatu *review* film dapat memakan waktu yang lama jika membaca *review* tersebut secara keseluruhan. Salah satu hal yang penting dalam sebuah *review* adalah opini yang terkandung didalamnya, opini tersebut dapat berupa opini positif atau negatif. Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan mengelompokkan *review* menjadi opini positif atau negatif (Zhang *et al.*, 2011). Sehingga diperlukannya suatu sistem yang secara otomatis dapat menghasilkan klasifikasi sentimen positif atau negatif terhadap suatu *review*.

Naïve Bayes Classifier merupakan salah satu teknik klasifikasi yang berbasis probabilistik. Penelitian (Zhang *et al.*, 2011) membandingkan antara *Support Vector Machine* dan *Naïve Bayes Classifier* dalam pengklasifikasian sentimen pada *review restaurant*. Hasil yang didapatkan adalah *Naïve Bayes Classifier* mendapatkan akurasi yang lebih baik dibandingkan *Support Vector Machine*. Penelitian (Dey *et al.*, 2016) membandingkan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbour* dalam analisis sentimen pada data *review film* dan *review hotel* dalam bahasa Inggris. Akurasi yang didapatkan pada penelitian ini terkhusus pada data *review film* adalah *Naïve Bayes Classifier* mendapatkan akurasi sebesar 80% - 90% sedangkan *K-Nearest Neighbour* hanya sebesar 70% .

Namun, salah satu yang menjadi masalah pada klasifikasi teks adalah dimensi tinggi dari ruang fitur, bahkan jika semua fitur tersebut digunakan dapat mengurangi *accuracy* dari *classifier* (Chen *et al.*, 2009). Untuk menyelesaikan permasalahan ini, *feature selection* dapat membuat klasifikasi menjadi lebih efektif dan efisien dengan mengurangi jumlah data yang dianalisa, maupun mengidentifikasi fitur yang sesuai untuk dipertimbangkan dalam proses pelatihan (Moraes, Valiati, and Neto, 2012).

Penelitian (Chandani dkk, 2015) yaitu *Komparasi Algoritma Klasifikasi Machine Learning dan Feature Selection pada Analisis Sentimen Review Film*. Penelitian ini melakukan komparasi lima algoritma seleksi fitur yaitu *Information Gain*, *Chi-Square*, *Forward Selection* dan *Backward Elimination* yang diujikan dengan beberapa algoritma klasifikasi. Pada penelitian ini didapatkan hasil bahwa *Information Gain* mendapatkan nilai *accuracy* yang terbaik, dengan nilai *accuracy*

average adalah 84.57% pada seleksi fitur. Penelitian selanjutnya (Utami dan Wahono, 2015) berjudul *Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes*. Hasil penelitian ini menunjukkan bahwa integrasi metode *Information Gain* dan *Adaboost* mampu meningkatkan akurasi *Naïve Bayes* dari 70% menjadi 99,5%.

Berdasarkan uraian diatas, maka penelitian ini akan menggunakan *Information Gain* dan *Naïve Bayes Classifier* untuk membangun model klasifikasi sentimen dari sebuah *review* film berbahasa Indonesia secara otomatis.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah banyaknya jumlah pengguna yang menulis opini secara *online* terus meningkat dan belum adanya penggunaan metode *Information Gain* dan *Naïve Bayes Classifier* untuk pengklasifikasian *review* film berbahasa Indonesia.

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut:

1. Mengembangkan perangkat lunak untuk menentukan kategori sentimen pada *review* film berbahasa Indonesia.
2. Mengetahui tingkat akurasi dan waktu klasifikasi dalam pengklasifikasian sentimen *review* film menggunakan metode *Information Gain* dan *Naïve Bayes Classifier*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah :

1. Menghasilkan sistem yang secara otomatis dapat mengetahui kategori sentimen pada *review* film.
2. Bagi pembaca *review* film, dapat mempermudah mengetahui isi dari *review* film secara cepat.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Klasifikasi hanya dilakukan pada sentimen *review* film berbahasa Indonesia.
2. Klasifikasi yang dihasilkan berupa sentimen positif atau negatif.
3. Sistem tidak mengidentifikasi bahasa alay, *emoticons*, dan singkatan kata.
4. Data *review* film dimasukkan di dalam file dalam bentuk *.txt*

1.7 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini akan membahas tentang latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini membahas seluruh dasar-dasar teori yang digunakan dimulai dari informasi mengenai domain, dan tentang metode yang digunakan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai unit penelitian, tahapan yang dilaksanakan pada penelitian, tahapan proses secara umum, metode pengembangan perangkat lunak, tahap pengujian dan manajemen proyek penelitian.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas perancangan dan lingkungan implementasi, berupa analisis dari masalah yang dihadapi dalam penelitian serta perancangan perangkat lunak untuk pengklasifikasian review film menggunakan *Information Gain* dan *Naive Bayes Classifier* yang digunakan sebagai alat penelitian.

BAB V HASIL DAN ANALISIS PENELITIAN

Bab ini membahas implementasi dari hasil analisis dan perancangan yang telah dilakukan sebelumnya. Hasil analisis berupa kesimpulan yang dapat diperoleh dari penelitian. Melakukan pengujian perangkat lunak dan pengujian data penelitian.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi semua kesimpulan dari uraian-uraian yang telah dibahas sebelumnya, dan saran yang diharapkan dapat berguna untuk pengembangan penelitian lebih lanjut.

1.8 Kesimpulan

Pada bab ini telah dibahas latar belakang masalah penelitian ini dalam pengklasifikasian sentimen review film. Karena itu, penelitian ini akan mengimplementasikan metode *Information Gain* dan *Naive Bayes Classifier* untuk membangun sebuah perangkat lunak untuk menentukan kategori sentimen pada *review film* berbahasa Indonesia.

DAFTAR PUSTAKA

- Andilala. (2016). "Movie Review Sentimen Analisis dengan Metode Naive Bayes Base On Feature Selection." Pseudocode **III**.
- Chandani, V., et al. (2015). "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film." Journal of Intelligent Systems **1**.
- Chen, J., et al. (2009). "Feature selection for text classification with Naïve Bayes." Expert Systems with Applications: 5432-5435.
- Dey, L., et al. (2016). "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier." International Journal of Information Engineering and Electronic Business: 54-62.
- Efendi, R., et al. (2012). "Klasifikasikan Dokumen Berbahasa Indonesia Menggunakan Naive Bayes Classifier." Journal of Research in Computer Science and Applications **1**.
- Fachrurrozi, M. and N. Yusliani. (2015). "Analisis Sentimen Pengguna Jejaring Sosial Menggunakan Metode Support Vector Machine".
- Feldman, R. and J. Sanger (2007). The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data. United States of America by Cambridge University Press, New York.
- Hamzah, A. (2012). "Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokkan Teks Berita dan Abstract Akademis ".
- Kohavi, R. (2001). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." Appears in the International Joint Conference on Artificial Intelligence.
- Kruchten, P. (2000). The Rational Unified Process: An Introduction, Second Edition 2nd. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA ©2000

- Mitchell, T. M. (1997). *Machine Learning*, McGraw-Hill Science/Engineering / Math.
- Moraes, R., et al. (2012). "Document-level sentiment classification: An empirical comparison between SVM and ANN." *Expert Systems with Applications*.
- Naufal, A. R., et al. (2015). "Penerapan Bootstrapping untuk Ketidakseimbangan Kelas dan Weighted Information Gain untuk Feature Selection pada Algoritma Support Vector Machine untuk Prediksi Loyalitas Pelanggan." *Journal of Intelligent Systems* **1**.
- Pang, B. and L. Lee. (2008). "Opinion mining and sentiment analysis." *Foundations and Trends in Information Retrieval* **2**.
- Sunaryono. (2017). "Penelitian Komparasi Algoritma Klasifikasi dalam Menentukan Website Palsu." *Teknikom* **1**.
- Suri, D. J. and K. K. Purnamasari. (2017). "Perbandingan Seleksi Fitur Untuk Klasifikasi Sentimen SVM Pada Twitter."
- Utami, L. D. and R. S. Wahono. (2015). "Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes." *Journal of Intelligent Systems* **1**.
- Yang, Y. and J. O. Pedersen. (1997). "A Comparative Study on Feature Selection in Text Categorization."
- Zhang, Z., et al. (2011). "Sentiment classification of Internet restaurant reviews written in Cantonese." *Expert Systems with Applications*.