



TESIS - KS142501

**PEMANFAATAN FILTER DALAM *OBJECT-BASED*  
*OPINION MINING* PADA *REVIEW* PRODUK  
PARIWISATA**

BONDA SISEPHAPUTRA

NRP. 5113202012

DOSEN PEMBIMBING

Dr. Ir. Aris Tjahyanto., M.Kom

PROGRAM MAGISTER

BIDANG KEAHLIAN SISTEM INFORMASI

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2016



THESIS - KS142501

# **THE UTILIZATION OF FILTER ON OBJECT-BASED OPINION MINING IN TOURISM PRODUCT REVIEWS**

BONDA SISSEPHAPUTRA

NRP. 5113202012

SUPERVISOR

Dr. Ir. Aris Tjahyanto., M.Kom

MAGISTER PROGRAM

MAJOR IN INFORMATION SYSTEM

DEPARTMENT OF INFORMATICS ENGINEERING

FACULTY OF INFORMATION TECHNOLOGY

SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY

SURABAYA

2016

## **LEMBAR PENGESAHAN**

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar

**Magister Komputer (M.Kom)**

di

**Institut Teknologi Sepuluh Nopember**

oleh:

**Bonda Sisepaputra**

**NRP. 5113202012**

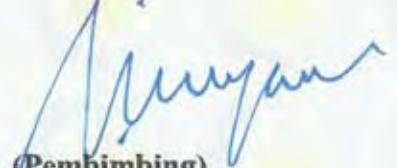
**Tanggal Ujian : 13 Juli 2016**

**Periode Wisuda : September 2016**

Disetujui oleh:

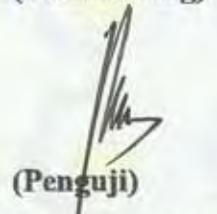
**Dr. Ir. Aris Tjahyanto., M.Kom**

**NIP. 196503101991021001**

  
(Pembimbing)

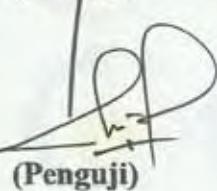
**Dr. Eng. Febriliyan Samopa, S.Kom, M.Kom**

**NIP. 197302191998021001**

  
(Penguji)

**Mahendrawathi ER., S.T., M.Sc., Ph.D**

**NIP. 197610112006042001**

  
(Penguji)

**Direktur Program Pascasarjana,**



**Prof.Ir. Djamar Manfaat, M.Sc, Ph.D**

**NIP. 56012021987011001**

## **PEMANFAATAN FILTER DALAM *OBJECT-BASED OPINION MINING* PADA REVIEW PRODUK PARIWISATA**

Nama Mahasiswa : Bonda Sisepaputra  
NRP : 5113202012  
Pembimbing : Dr. Ir. Aris Tjahyanto., M.Kom

### **ABSTRAK**

Dalam menentukan kualitas dari suatu produk pariwisata dapat dilihat dari banyak aspek karena masing-masing memiliki keunikan masing-masing, sehingga untuk mendapatkan gambaran yang lebih jelas maka harus dievaluasi masing-masing aspek yang ada. Penggalian informasi yang cocok adalah dengan *object-based opinion mining*. Penerapan aturan NLP untuk menentukan orientasi semantik dari objek yang berhasil diekstrak terbukti efektif. Akan tetapi, kinerja pada tugas ekstraksi objek cukup buruk dalam domain pariwisata.

Pada penelitian ini peneliti menggunakan filter dalam proses ekstraksi objek yang terdapat pada data *review* hotel dan restoran. Penggunaan data filter pada *object-based opinion mining* berhasil memperoleh hasil ekstraksi objek yang lebih baik, karena dapat memfilter objek-objek yang bukan merupakan objek yang berpengaruh terhadap kualitas suatu hotel atau resoran.

Penerapan filter dalam proses ekstraksi objek jelas dapat meningkatkan presisi yaitu pada pendekatan *frequent object* yang semula 45.7% menjadi 64.49% pada *review* hotel dan dari 44.82% menjadi 64.61% pada *review* restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan *frequent object* dan *infrequent object* dimana terjadi peningkatan hasil presisi yang semula 22.33% menjadi 63.02% pada *review* hotel dan dari 21.6% menjadi 65.4% pada *review* restoran. Pada proses pengukuran performa klasifikasi sentimen objek menggunakan semua objek yang telah berhasil diekstrak, penggunaan filter mendapatkan hasil yang dominan secara keseluruhan dibanding dengan tanpa filter. Pada pendekatan *frequent object* mendapatkan 56.85% akurasi, 60.91% presisi, dan 79.93% recall pada *review* hotel, serta mendapatkan hasil 58.85% akurasi, 63.26% presisi, dan 84.14% recall pada *review* restoran. Pada proses klasifikasi sentimen kalimat penggunaan objek terekstraksi sebagai atribut klasifikasi untuk menentukan orientasi kalimat mendapatkan hasil yang lebih baik dibanding dengan hanya menggunakan kalimat *review* secara langsung sebagai atribut klasifikasi, walaupun penggunaan objek dari *object-based opinion mining* masih kurang dibandingkan dengan penggunaan aspek corpus sebagai atribut.

**Kata Kunci:** *object-based opinion mining*, *opinion mining*, analisis sentimen, pariwisata, hotel, restoran

## **THE UTILIZATION OF FILTER ON OBJECT-BASED OPINION MINING IN TOURISM PRODUCT REVIEWS**

By : Bonda Sisepaputra  
Student Identity Number : 5113202012  
Supervisor : Dr. Ir. Aris Tjahyanto., M.Kom

### **ABSTRACT**

The quality of a tourism product can be seen from many objects because each product is a unique individual. Therefore to get a clearer picture about a product, each object has to be evaluated. The information about each object can be extracted by using object-based opinion mining. Based on the previous research, the application of Natural Language Processing (NLP) rules on object-based opinion mining to determine the orientation of the semantic objects has showed good result. However, the performance of the extraction objects task quite bad in the tourism domain. In this study, researchers apply a filter on the objects' extraction process of the hotel and restaurant review data. The utilization of data filter in object-based opinion mining has succeeded in obtaining better objects' extraction result, because the utilization of filter will eliminate each object which is unrelated with the object of the hotel or restaurant domain.

The application of filter in the objects' extraction process can improve the precision of frequent object approach from 45.7% to 64.49% on of the hotel review and from 44.82% to 64.61% on the restaurant review. In the objects' extraction process which is using frequent and infrequent approach, the precision is increased from 22.33% to 63.02% on the hotel review and from 21.6% to 65.4% on the restaurant review. In the performance measurement process, with the usage of all objects that have been successfully extracted, the use of filter got more dominant result compared to no filter on object sentiment classification process. The utilization of filter on frequent object approach got 56.85% accuracy, 60.91% precision, and 79.93% recall on the hotel review, and got 58.85% accuracy, 63.26% precision, and 84.14% recall on the restaurant review. While in the process of sentences sentiment classification, the usage of extracted object as a classification attribute to determine the orientation of the sentence got better results than only using the sentence review directly as an classification attribute, although the results object of object-based opinion mining is worse than the usage of the object of the corpus data as an attribute.

**Keywords:** *object-based opinion mining, opinion mining, Sentiment Analysis, tourism, hotel, restaurant*

## DAFTAR ISI

DAFTAR ISI .....	xi
DAFTAR GAMBAR .....	xv
DAFTAR TABEL .....	xvii
BAB 1 PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan Penelitian.....	5
1.4. Ruang Lingkup Penelitian .....	5
1.5. Kontribusi Penelitian .....	7
1.6. Sistematika Penulisan Dokumen.....	7
BAB 2 LANDASAN TEORI DAN KAJIAN PUSTAKA .....	9
2.1. Penelitian Terkait .....	9
2.2. <i>Aspect-Based Opinion Mining</i> .....	11
2.3. Praproses Data.....	13
2.4. Ekstraksi Aspek.....	14
2.5. Penentuan Orientasi Aspek .....	16
2.5.1. Aturan Orientasi Kata .....	18
2.5.2. Aturan Orientasi Aspek.....	19
2.6. Penyusunan Rangkuman Hasil Analisis .....	22
2.7. Klasifikasi Teks.....	22
2.7.1. <i>Naive Bayes</i> .....	22
2.7.2. <i>Support Vector Machines (SVM)</i> .....	23
2.7.3. Random Forests .....	23
2.8. Pengukuran Performa .....	24

2.8.1.	Akurasi .....	24
2.8.2.	<i>Precision</i> .....	24
2.8.3.	<i>Recall</i> .....	25
2.8.4.	<i>F-Measure</i> .....	25
	BAB 3 METODOLOGI PENELITIAN .....	27
3.1.	Penyiapan Data.....	28
3.2.	Praproses Data.....	30
3.3.	<i>Object-based Opinion Mining</i> .....	32
3.3.1.	Ekstraksi Objek .....	33
3.3.2.	Penentuan Orientasi Objek .....	35
3.4.	Klasifikasi Orientasi Kalimat Review .....	35
3.5.	Uji Coba dan Analisis Hasil Penelitian .....	36
3.5.1.	Skenario Uji Coba.....	36
3.5.2.	Analisis Hasil Penelitian .....	39
3.6.	Penyusunan Kesimpulan dan Saran Pengembangan Penelitian.....	40
	BAB 4 UJI COBA DAN ANALISIS HASIL .....	41
4.1.	Data Uji Coba.....	41
4.2.	Lingkungan Uji Coba .....	41
4.3.	Skenario Uji Coba .....	42
4.3.1.	Uji Coba Ekstraksi Objek.....	42
4.3.2.	Uji Coba Klasifikasi Sentimen Objek.....	43
4.3.3.	Uji Coba Deteksi Kalimat Opini .....	43
4.3.4.	Uji Coba Klasifikasi Sentimen Kalimat Review .....	43
4.4.	Pelaksanaan Uji Coba .....	44
4.4.1.	Pembuatan Database dan Program Prototype .....	44
4.4.2.	Persiapan Data Uji Coba .....	45

4.4.3.	Praproses Data Uji Coba .....	47
4.4.4.	Proses Uji Coba Ekstraksi Objek .....	49
4.4.5.	Proses Uji Coba Klasifikasi Sentimen Objek.....	51
4.4.6.	Proses Uji Coba Deteksi Kalimat Opini .....	53
4.4.7.	Proses Uji Coba Klasifikasi Sentimen Kalimat Review .....	53
4.5.	Hasil Uji Coba.....	55
4.5.1.	Uji Coba Ekstraksi Objek .....	55
4.5.2.	Uji Coba Klasifikasi Sentimen Objek.....	56
4.5.3.	Uji Coba Deteksi Kalimat Opini .....	58
4.5.4.	Uji Coba Klasifikasi Sentimen Kalimat Review .....	59
4.6.	Analisis Hasil Uji Coba .....	61
4.6.1.	Analisis Eksatraksi Objek .....	61
4.6.2.	Analisis Klasifikasi Sentimen Objek .....	62
4.6.3.	Analisis Deteksi Kalimat Opini.....	63
4.6.4.	Analisis Klasifikasi Sentimen Kalimat Review .....	64
	BAB 5 KESIMPULAN DAN SARAN .....	67
5.1.	Kesimpulan .....	67
5.2.	Saran .....	69
	DAFTAR PUSTAKA .....	71
	LAMPIRAN A .....	A-1
	LAMPIRAN B.....	B-1
	LAMPIRAN C.....	C-1
	LAMPIRAN D .....	D-1
	LAMPIRAN E.....	E-1

## DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> .....	24
Tabel 3.1 Detail Data Corpora.....	31
Tabel 3.2 Detail Data Corpora.....	37
Tabel 3.3 Detail Objek yang Ditemukan dalam Corpora.....	37
Tabel 4.1 Spesifikasi Lingkungan Uji Coba – Perangkat Keras .....	42
Tabel 4.2 Spesifikasi Lingkungan Uji Coba - Perangkat Lunak .....	42
Tabel 4.3 Contoh Hasil Tokenizing .....	47
Tabel 4.4 Contoh Hasil POS Tagging.....	48
Tabel 4.5 Contoh Hasil Chunking .....	48
Tabel 4.6 Contoh Hasil Stopping .....	49
Tabel 4.7 Contoh Hasil Stemming.....	49
Tabel 4.8 Contoh perubahan format objek.....	54
Tabel 4.9 Hasil Uji Ekstraksi Objek Hotel.....	55
Tabel 4.10 Hasil Uji Ekstraksi Objek Restoran.....	55
Tabel 4.11 Waktu Pengujian Ekstraksi Objek.....	56
Tabel 4.12 Hasil Uji Klasifikasi Sentimen Objek Hotel (Objek yang Benar) .....	56
Tabel 4.13 Hasil Uji Klasifikasi Sentimen Objek Restoran (Objek yang Benar) .	57
Tabel 4.14 Hasil Uji Klasifikasi Sentimen Objek Hotel (Semua Objek).....	57
Tabel 4.15 Hasil Uji Klasifikasi Sentimen Objek Restoran (Semua Objek) .....	57
Tabel 4.16 Waktu Pengujian Klasifikasi Sentimen Objek.....	58
Tabel 4.17 Hasil Uji Deteksi Kalimat Opini Hotel.....	58
Tabel 4.18 Hasil Uji Deteksi Kalimat Opini Restoran .....	58
Tabel 4.19 Waktu Pengujian Deteksi Kalimat Opini .....	59
Tabel 4.20 Hasil Uji Klasifikasi dengan Atribut Data Kalimat Review .....	60
Tabel 4.21 Hasil Uji Klasifikasi dengan Atribut Data Object Data Corpus .....	60
Tabel 4.22 Hasil Uji Klasifikasi dengan Atribut Data Frequent Object .....	60
Tabel 4.23 Hasil Uji Klasifikasi dengan Atribut Data Frequent object dan Infrequent Object .....	60

Tabel 4.24 Hasil Uji Klasifikasi dengan Atribut Data Frequent Object yang Terfilter .....	61
Tabel 4.25 Hasil Uji Klasifikasi dengan Atribut Data Frequent Object dan Infrequent Object yang Terfilter .....	61
Tabel 4.26 Hasil Uji Klasifikasi dengan Atribut Data Filter Object .....	61

## **DAFTAR GAMBAR**

Gambar 3.1 Metodologi Penelitian.....	27
Gambar 3.2 Gambaran Umum Alur Data Penelitian.....	28
Gambar 3.3 Contoh Data Uji (Marrese-Taylor, 2014) .....	28
Gambar 3.4 Tahapan Praproses .....	30
Gambar 3.5 Contoh hasil <i>POS Tagging</i> .....	31
Gambar 3.6 Tahapan <i>Object-based Opinion Mining</i> .....	33
Gambar 4.1 Rancangan Database .....	45
Gambar 4.2 Proses Ekstraksi Objek .....	50
Gambar 4.3 Proses Klasifikasi Sentimen Objek.....	52

# **BAB 1**

## **PENDAHULUAN**

Pada bab awal ini dijelaskan mengenai gambaran penelitian dari latar belakang penelitian, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, hingga kontribusi penelitian.

### **1.1. Latar Belakang**

Penggunaan media internet beberapa tahun terakhir sangat meningkat tajam seiring dengan berkembangnya teknologi yang mendukungnya. Website sudah menjadi media penyedia dan penyimpanan informasi yang sangat besar. Para pengguna secara online menambahkan data dan informasi secara online setiap hari, salah satunya adalah ulasan tentang suatu produk atau jasa (Bucur, 2015). Orang akan cenderung membaca ulasan dari orang lain terlebih dahulu sebelum menentukan pilihan dalam membeli barang atau menggunakan jasa (Marrese-Taylor, 2013). Dengan demikian perusahaan atau individu semakin menggunakan konten dalam media ini untuk menggali informasi, sehingga dapat membuat mereka mengambil keputusan dengan lebih baik (Park & Kim, 2009; Zhu & Zhang, 2010). Akan tetapi dengan banyaknya informasi yang ada membuat kesulitan untuk membaca keseluruhan ulasan tersebut. Kehadiran sebuah sistem yang mampu mengekstrak informasi secara otomatis sangatlah dibutuhkan untuk memberikan gambaran dari keseluruhan ulasan yang ada. Deteksi dan ekstraksi opini dari ulasan suatu produk atau jasa sudah menjadi salah satu topik yang diteliti beberapa tahun belakangan. Didalam literatur *opinion mining* juga disebut sebagai analisis sentimen, yaitu mempelajari pendapat dan perasaan seseorang yang dituangkan dalam text menggunakan mesin komputasi. Tantangan dari penelitian ini adalah bagaimana mengekstrak pengetahuan dari data yang tidak terstruktur (Bucur, 2014).

Informasi tentang pariwisata menjadi salah satu yang diakses oleh masyarakat. Dengan tersedianya informasi ini maka masyarakat tidak lagi kesulitan dalam mencari informasi tempat tujuan mereka menghabiskan masa liburan.

Banyak website yang menyediakan informasi ulasan tentang produk pariwisata dan menyediakan fitur untuk para pengunjung website dapat memberikan umpan balik yang berupa ulasan atau opini mereka terhadap objek tersebut. Masalah yang terjadi adalah dengan banyaknya *feedback* yang ada akan sulit untuk seseorang membaca semua ulasan dan mendapatkan gambaran secara keseluruhan. Setiap opini seseorang juga mengandung kelengkapan opini dan perspektif yang berbeda-beda sehingga akan sulit dievaluasi. Sebuah sistem yang dapat merangkum ulasan dari data yang ada dapat menawarkan perspektif keseluruhan, akan menghemat banyak waktu dan mempermudah proses keputusan bagi para pencari informasi.

Hal yang nantinya akan menjadi tantangan dalam penggalian informasi terhadap data *review* adalah banyak kalimat yang menyebutkan opini dari obyek yang sedang dikaji atau juga dari setiap fitur dan komponen yang ada didalamnya. Di sisi lain, banyak juga kalimat yang tidak mengandung opini, dan juga menyebutkan benda-benda yang tidak sesuai dengan atribut atau komponen dari obyek yang diulas (Marrese-Taylor, 2014).

Dalam menentukan kualitas dari suatu produk pariwisata dapat dilihat dari banyak aspek karena masing-masing tempat memiliki keunikan masing-masing, sehingga untuk mendapatkan gambaran yang lebih jelas maka harus dievaluasi masing-masing aspek yang ada. Dengan demikian maka yang paling cocok untuk penggalian informasi ini adalah dengan *aspect-based opinion mining*. Dalam metode ini analisis dilakukan secara mendetail sampai dengan analisis pendapat untuk masing-masing aspek yang ada di dalam suatu opini, sehingga nantinya akan didapat gambaran yang lebih jelas untuk masing-masing aspek suatu produk pariwisata. Dalam suatu produk atau jasa memiliki kelebihan dan kekurangan masing-masing, sehingga sangat penting bagi seseorang untuk mengetahui secara detail tentang gambarang masing-masing aspek yang ditawarkan. Di dalam banyak literatur cenderung mengklasifikasikan produk ke dalam 2 jenis yaitu fisik dan jasa yang tidak berwujut (Marrese-Taylor, 2013). Didalam ulasan benda fisik maka opini yang ada cenderung mengarah ke fitur atau fungsi dari produk tersebut. Akan tetapi untuk objek yang akan diteliti yaitu produk pariwisata adalah lebih mengarah ke produk jasa yang tidak berwujud, dimana membuat ulasan yang akan mungkin

diberikan oleh orang lain memungkinkan tidak hanya mengenai objek secara langsung saja melainkan objek-objek pendukungnya.

Marrese-Taylor (2014) melakukan penelitian dimana mereka menerapkan analisis *aspect-based opinion mining* dalam bidang pariwisata khususnya pengolahan data opini terhadap layanan hotel dan restoran. Hasil penelitian tersebut menghasilkan sebuah *prototype* pengolahan data opini dan memiliki hasil *output* analisis secara visual yang dapat dengan mudah dipahami banyak pihak.

Penerapan aturan NLP untuk menentukan orientasi semantik dari aspek yang berhasil diekstrak terbukti sangat efektif, mencapai presisi dan recall rata-rata 90%. Sebaliknya, Hasil dari penelitian tersebut menunjukkan bahwa kinerja pada tugas ekstraksi aspek cukup buruk dalam domain pariwisata. Algoritma ini hanya mampu mengekstrak hampir 30% dari total ekspresi eksplisit untuk hotel dan hampir 40% untuk restoran. Hal tersebut disebabkan oleh persentase yang tinggi dari ekspresi diekstraksi yang tidak sesuai dengan ekspresi aspek nyata untuk kedua domain tersebut. Di sisi lain, klasifikasi sentimen menunjukkan hasil yang cukup baik, tetapi dalam kasus ini sebagian besar kesimpulan yang mungkin sulit untuk dibuktikan karena tugas ini hanya dievaluasi dari ekspresi aspek yang diekstrak. Karena ekspresi ini merupakan yang paling sederhana, menentukan orientasi sentimen pada aspek tersebut mungkin lebih mudah. Akibatnya, presisi dan recall bisa menurun ketika semua ekspresi aspek dianggap. Penyebab buruknya hasil ekstraksi aspek adalah pemberi opini menyebutkan benda-benda yang tidak sesuai dengan atribut atau komponen dari produk, hal tersebut juga berakibat diperolehnya presisi yang rendah dalam ekstraksi aspek eksplisit pada kedua kasus. Misalnya, dalam kasus hotel, pengguna umumnya merujuk pada benda-benda seperti waktu, hari dan kota. Meskipun relevan dengan cerita tentang hotel tetapi bukan termasuk aspek yang diharapkan.

Dikarenakan hasil yang diperoleh oleh Marrese-Taylor (2014) dalam hal ekstraksi fitur masih kurang, maka pada penelitian kali ini peneliti akan mencoba untuk menggunakan filter untuk mendapatkan hasil yang lebih baik. Fokus dari penelitian ini adalah untuk menerapkan filter dengan menggunakan data filter yang disusun berdasarkan faktor dan komponen yang ada dan berhubungan dengan domain hotel dan restoran, sehingga dapat memperbaiki hasil ekstraksi aspek

dengan mengeliminasi objek yang dianggap tidak mewakili faktor-faktor yang berhubungan dengan kualitas suatu produk pariwisata.

Dalam mempelajari penelitian yang dilakukan Marrese-Taylor (2014) dimana menjadi referensi utama dalam penelitian ini, penulis mendapatkan kesulitan dalam memahami dan menerapkan konsep aspek yang dikemukakan oleh Marrese-Taylor dalam tahap ekstraksi aspek. Untuk membuatnya lebih simple dan tidak menyalahi konsep yang ada, akhirnya dalam penelitian ini penulis menyebut komponen atau atribut yang merupakan bagian dari domain hotel dan restoran yang menjadi target opini akan disebut sebagai objek. Objek diekstrak dari kata benda dan frase kata benda yang ada dalam kalimat review yang dianggap mempengaruhi kualitas suatu layanan produk pariwisata. Sehingga metode *opinion mining* dalam penelitian ini akan disebut sebagai analisis *object-based opinion mining* dengan acuan berdasarkan analisis *aspect-based opinion mining* yang dilakukan oleh Marrese-Taylor (2014).

Selain itu, pada penelitian ini penulis mencoba untuk menguji apakah hasil dari *object-based opinion mining* dapat digunakan untuk menentukan orientasi sentimen secara umum. Selain membutuhkan informasi objek-objek yang dikomentari, tentu saja pembaca *review* produk pariwisata perlu untuk memahami kesimpulan dari setiap *review* yang ada. Oleh sebab itu, peneliti mencoba untuk mengklasifikasikan *review* data ke dalam beberapa kelas: positif, netral, atau negatif. Sehingga dapat ditarik kesimpulan apakah obyek wisata yang layak direkomendasikan atau tidak. Pada penelitian ini akan dilakukan pemrosesan data hasil pengolahan *object-based opinion mining* dengan melakukan proses klasifikasi menggunakan machine learning. Peneliti akan membandingkan hasil prediksi dengan menggunakan tiga teknik klasifikasi yang berbeda, yaitu Naïve Bayes, support-vector machines (SVM), dan Random Forest (RF).

## 1.2. Rumusan Masalah

Banyaknya *review* yang ada akan sulit bagi seseorang untuk membaca semua ulasan dan mendapatkan kesimpulan secara menyeluruh. Berdasarkan penelitian yang dilakukan Marrese Taylor (2014), mereka menggunakan teknik *aspect-based opinion mining* untuk mengekstrak aspek dari data *review* dan dapat memprediksi

orientasi sentimen setiap aspek, dimana dalam penelitian ini proses analisis itu mengalami beberapa penyesuaian pada bagian ekstraksi aspek sehingga disebut *object-based opinion mining*. Hasil ekstraksi aspek yang diperoleh pada penelitian sebelumnya masih kurang maksimal. Berdasarkan hal tersebut, maka rumusan masalah yang mendasari penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan filterisasi terhadap objek yang berhasil diekstrak oleh *object-based opinion mining* pada data *review* layanan hotel dan restoran sehingga dapat meningkatkan hasil dari *object-based opinion mining*.
2. Apakah hasil *object-based opinion mining* dapat digunakan untuk mengklasifikasikan orientasi kalimat *review* secara global.

### **1.3. Tujuan Penelitian**

Berdasarkan latar belakang dan rumusan masalah yang telah dijelaskan sebelumnya maka penelitian ini bertujuan untuk:

1. Menerapkan filterisasi objek pada *object-based opinion mining* pada data *review* layanan hotel dan restoran sehingga mendapatkan peningkatan hasil pada proses ekstraksi objek pada *object-based opinion mining*. Data filter yang digunakan disusun berdasarkan faktor dan komponen yang berhubungan dan mempengaruhi kualitas suatu layanan pada domain hotel dan restoran, sehingga dapat memperbaiki hasil ekstraksi dengan mengeliminasi objek yang dianggap tidak mewakili faktor-faktor yang tidak memiliki hubungan.
2. Menganalisis menggunakan tiga teknik klasifikasi, yaitu Naïve Bayes, support-vector machines (SVM), dan Random Forest (RF) untuk mengetahui performa dari hasil *object-based opinion mining* jika digunakan untuk mengklasifikasikan orientasi kalimat *review* secara global.

### **1.4. Ruang Lingkup Penelitian**

Penelitian ini memiliki beberapa batasan antara lain sebagai berikut:

- a. Penelitian ini menggunakan data *review* dan data uji yang ada pada penelitian Marrese-Taylor (2014).

- b. Data *review* yang digunakan berupa teks berbahasa Inggris.
- c. Penelitian merupakan modifikasi dari penelitian yang dilakukan oleh Marrese-Taylor (2014) dengan menambahkan filterisasi pada ekstraksi aspek menggunakan data filter. Selain itu terdapat perubahan pada tahap ekstraksi aspek, sehingga dalam penelitian ini proses analisis tidak disebut sebagai *aspect-based opinion mining* melainkan disebut sebagai *object-based opinion mining*. Perubahan tersebut terjadi karena kurangnya penjelasan secara mendetail dari tahapan ekstraksi aspek pada referensi yang membuat peneliti merasakan kesulitan untuk mengimplementasikannya. Perubahan yang dilakukan adalah tidak adanya proses *frequent itemset mining* dan *special linguistic rules filtering* pada proses ekstraksi aspek. Untuk menghindari kesalahan penyebutan istilah, maka dalam penelitian ini istilah yang digunakan untuk menyebutkan komponen atau atribut domain yang menjadi target opini yaitu aspek akan diganti dengan istilah objek. Dalam Kamus Besar Bahasa Indonesia (KBBI) objek memiliki arti benda, hal, dan sebagainya yang dijadikan sasaran untuk diteliti, diperhatikan, dan sebagainya. Berdasarkan hal tersebut penggunaan istilah objek dianggap sesuai, karena dalam penelitian ini objek diekstrak berdasarkan kata benda dan frase kata benda yang terdapat pada data review.
- d. Belum adanya konsep dan landasan ilmiah yang pasti dalam menyusun dan menentukan data filter yang sesuai, sehingga proses penyusunan data filter yang digunakan dalam penelitian ini dilakukan secara manual oleh peneliti, yaitu berdasarkan data ontologi yang didapat dari internet dan juga berdasarkan data aspek yang ada pada data corpus. Oleh karena itu apabila ada peneliti lain mencoba melakukan penelitian serupa maka kemungkinan besar data filter yang dihasilkan akan berbeda.
- e. Penelitian menerapkan analisis *object-based opinion mining* dalam bidang pariwisata khususnya pengolahan data opini terhadap layanan hotel dan restoran.

- f. Visualisasi hasil *object-based opinion mining* yang berupa bar chart sesuai dengan referensi secara detail tidak akan dibahas dalam penelitian ini.

### **1.5. Kontribusi Penelitian**

Hasil penelitian ini diharapkan dapat memberikan kontribusi baik secara teori maupun secara praktik. Kontribusi secara teori diperoleh dari penerapan filterisasi pada proses ekstraksi objek pada *object-based opinion mining* dengan menggunakan data filter. Filter diharapkan dapat memfilter objek-objek yang bukan merupakan komponen atau atribut yang dapat berpengaruh pada kualitas layanan dari hotel atau resoran, sehingga proses ekstraksi akan mendapatkan hasil yang lebih baik.

Dengan diperolehnya hasil *object-based opinion mining* yang lebih baik, maka informasi yang berhasil diekstrak diharapkan dapat bermanfaat bagi masyarakat umum untuk mendapatkan gambaran secara menyeluruh terhadap kualitas dari suatu produk pariwisata khususnya hotel dan restoran. Dengan demikian informasi tersebut dapat menjadi bahan pertimbangan dalam pengambilan keputusan.

### **1.6. Sistematika Penulisan Dokumen**

Sistematika penulisan dokumen laporan penelitian tesis ini dibagi menjadi lima bab yakni sebagai berikut:

## **BAB I PENDAHULUAN**

Pada bab ini dijelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, kontribusi penelitian, dan sistematika penulisan dalam pemaparan penelitian.

## **BAB II LANDASAN TEORI DAN KAJIAN PUSTAKA**

Pada bab ini dijelaskan mengenai kajian pustaka dari berbagai penelitian yang memiliki kaitan dengan penelitian ini. Kajian pustaka ini bertujuan untuk memperkuat dasar dan alasan dilakukannya penelitian ini. Selain kajian pustaka,

pada bab ini juga dijelaskan mengenai teori-teori terkait yang bersumber dari buku, jurnal, ataupun artikel yang berfungsi sebagai dasar dalam melakukan penelitian agar dapat memahami konsep atau teori penyelesaian permasalahan yang ada. Pada bab ini terdapat uraian mengenai penggalian teks yang mencangkup penjelasan mengenai *Aspect-based opinion mining*, ekstraksi aspek, prediksi sentimen dan pengukuran performa klasifikasi yang digunakan dalam penelitian ini.

### **BAB III METODOLOGI PENELITIAN**

Pada bab ini dijelaskan mengenai langkah-langkah penelitian beserta metode yang digunakan. Langkah-langkah penelitian dijelaskan dalam sebuah diagram alur yang sistematis dan akan dijelaskan tahap demi tahap.

### **BAB IV UJI COBA DAN ANALISIS HASIL**

Pada bab ini akan dilakukan uji coba terhadap metode yang telah dirancang sebelumnya. Uji coba ini dilakukan berdasarkan skenario uji coba yang telah dirancang sebelumnya. Selain itu pada bab ini juga dijelaskan mengenai analisis hasil uji coba.

### **BAB V KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari penelitian ini dan juga saran bagi penelitian mendatang yang berasal dari kekurangan ataupun temuan dari penelitian ini.

## **BAB 2**

### **LANDASAN TEORI DAN KAJIAN PUSTAKA**

Pada bab ini dijelaskan mengenai teori-teori yang mendasari penelitian dan kajian pustaka mengenai penelitian-penelitian yang terkait. Teori yang dijelaskan antara lain mengenai *aspect-based opinion mining*, praproses data, ekstraksi aspek, prediksi sentimen, penyusunan rangkuman hasil analisis.

#### **2.1. Penelitian Terkait**

Sebagaimana dinyatakan Liu (2007), dalam literatur *opinion mining* menawarkan dua pendekatan utama, dan berbasis non-aspek dan berbasis aspek. Pendekatan berbasis aspek sangat populer dan banyak penulis telah mengembangkan perspektif dan model mereka sendiri. Beberapa contoh diantaranya adalah (Archak, Ghose, & Ipeirotis, 2007; Decker & Trusov, 2010; Ku, Liang, & Chen, 2006; Lu, Zhai, & Sundaresan, 2009; Popescu & Etzioni, 2005; Titov & McDonald, 2008; Zhao & Li, 2009; Zhuang et al., 2006).

Penelitian terkait yang menjadi dasar dari penelitian ini adalah penelitian yang dilakukan oleh Marrese-Taylor pada tahun 2013 dan 2014, dimana penelitian tersebut merupakan pengembangan dari penelitian yang dilakukan oleh Liu (2007). Karena pendekatan Liu difokuskan pada ulasan produk fisik dan tidak bisa langsung diterapkan pada domain pariwisata, yang menyajikan fitur yang tidak dianggap oleh model. Dalam penelitian yang dilakukan Marrese-Taylor (2014), peneliti memperluas teknik *opinion mining* berdasarkan aspek yang telah dikembangkan oleh Bing Liu untuk menerapkannya ke domain pariwisata. Melalui studi menggunakan *review* online produk pariwisata, Marrese-Taylor (2014) mengusulkan penggunaan aturan berbasis NLP (*Natural Language Processing*) baru dan lebih kompleks untuk tugas subjektif dan klasifikasi sentimen pada aspek.

Marrese-Taylor sendiri melakukan dua kali penelitian dimana yang pertama pada tahun 2013 menawarkan sebuah pendekatan untuk mempertimbangkan alternatif baru untuk menemukan preferensi konsumen tentang produk pariwisata, khususnya hotel dan restoran, menggunakan opini yang tersedia di web sebagai

ulasan. Setelah itu Marrese-Taylor (2014) mengembangkan lagi hasil penelitian sebelumnya membuat fungsi visualisasi, perangkuman dan mengusulkan ekstensi dari penelitian sebelumnya untuk membantu pengguna mencerna ketersediaan *review* dengan cara yang mudah. Marrese-Taylor juga melakukan pengembangan arsitektur dan *tool* dalam *aspect-based opinion mining*, yang kemudian digunakan untuk membuat *prototype* dan menganalisis opini dari TripAdvisor. Adapun kontribusi yang dihasilkan penelitian Marrese-Taylor (2014) adalah berhasil memberikan kontribusi dalam hal analisis *aspect-based opinion mining* pada bidang pariwisata, lebih tepatnya pengolahan data opini tentang layanan hotel dan restoran dimana belum pernah dilakukan oleh peneliti lain. Penelitian tersebut juga mengembangkan bentuk rangkuman dan visualisasi yang baru dengan tujuan lebih mudah untuk dipahami semua pihak.

Dikarenakan hasil yang diperoleh oleh Marrese-Taylor (2014) dalam hal ekstraksi fitur masih kurang, maka pada penelitian kali ini peneliti akan mencoba untuk menggunakan data filter untuk mendapatkan hasil yang lebih baik. Dengan penggunaan filter diharapkan dapat memfilter aspek-aspek yang bukan merupakan aspek yang ada pada domain hotel atau resoran sehingga ekstraksi aspek akan mendapatkan hasil yang lebih baik.

Dalam mempelajari penelitian yang dilakukan Marrese-Taylor (2014) dimana menjadi referensi utama dalam penelitian ini, penulis mendapatkan kesulitan dalam memahami dan menerapkan konsep aspek yang dikemukakan oleh Marrese-Taylor dalam tahap ekstraksi aspek. Untuk membuatnya lebih simple dan tidak menyalahi konsep yang ada, akhirnya dalam penelitian ini penulis menyebut komponen atau atribut yang merupakan bagian dari domain hotel dan restoran yang menjadi target opini akan disebut sebagai objek. Objek diekstrak dari kata benda dan frase kata benda yang ada dalam kalimat review yang dianggap mempengaruhi kualitas suatu layanan produk pariwisata. Sehingga metode *opinion mining* dalam penelitian ini akan disebut sebagai analisis *object-based opinion mining* dengan acuan berdasarkan analisis *aspect-based opinion mining* yang dilakukan oleh Marrese-Taylor (2014).

Selain itu terdapat perubahan pada tahap ekstraksi aspek, sehingga dalam penelitian ini proses analisis tidak disebut sebagai *aspect-based opinion mining*

*melainkan* disebut sebagai *object-based opinion mining*. Perubahan tersebut terjadi karena kurangnya penjelasan secara mendetail dari tahapan ekstraksi aspek pada referensi yang membuat peneliti merasakan kesulitan untuk mengimplementasikannya. Perubahan yang dilakukan adalah tidak adanya proses *frequent itemset mining* dan *special linguistic rules filtering* pada proses ekstraksi aspek. Untuk menghindari kesalahan penyebutan istilah, maka dalam penelitian ini istilah yang digunakan untuk menyebutkan komponen atau atribut domain yang menjadi target opini yaitu aspek akan diganti dengan istilah objek. Dalam Kamus Besar Bahasa Indonesia (KBBI) objek memiliki arti benda, hal, dan sebagainya yang dijadikan sasaran untuk diteliti, diperhatikan, dan sebagainya. Berdasarkan hal tersebut penggunaan istilah objek dianggap sesuai, karena dalam penelitian ini objek diekstrak berdasarkan kata benda dan frase kata benda yang terdapat pada data review.

## **2.2. Aspect-Based Opinion Mining**

*Opinion mining* adalah salah satu konsep baru dalam web *intelligence* dan mencakup masalah ekstraksi dan analisis data pendapat. Salah satu pendekatan utama dalam *opinion mining* yang ada di literatur adalah *aspect-based opinion mining*. Perspektif ini dikembangkan oleh Liu (2007), dimana fokus dari proses penggalian datanya adalah mendapatkan fitur dari obyek yang dievaluasi dalam data pendapat.

*Aspect-based opinion mining* membagi teks ke dalam aspek-aspek, dapat disebut juga dengan fitur atau subtopik dalam literatur, yang biasanya sesuai dengan topik yang dianggap penting atau perwakilan dari teks yang sedang dianalisis (Liu, 2007). Liu mengusulkan bahwa data *review/opini* terbagi dalam 5 bagian, yaitu terdiri dari:

- a. Suatu entitas: Usulan untuk menunjukkan tujuan opini atau dengan kata lain apa yang sedang dievaluasi oleh opini. Entitas dapat berisi satu set komponen dan atribut, serta masing-masing komponen entitas dapat memiliki subkomponen dan atributnya sendiri. Akhirnya, suatu entitas dapat diuraikan ke *tree* atau hirarki sub atribut dan sub komponen.

- b. Aspek: Karena sulit untuk belajar entitas pada tingkat hirarki yang sewenang-wenang, hirarki ini disederhanakan untuk satu atau dua tingkat, yang menunjukkan sebagai aspek setiap komponen atau atribut dari entitas. Dengan cara ini, akar dari hirarki atau *tree* menjadi entitas sendiri, masing-masing daun merupakan aspek dan link yang berhubungan.
- c. Orientasi Sentimen, mengingat opini mengungkapkan sentimen positif atau negatif tentang apa yang mereka mengevaluasi.
- d. Penulis Opini, yang sesuai dengan pengguna (seseorang, suatu perusahaan, dll) yang memberikan pendapat.
- e. Waktu: Waktu dan tanggal ketika opini itu diberikan. Pendapat tersebut kemudian dianggap sebagai positif atau negatif tampilan, sikap, emosi atau penilaian tentang suatu entitas atau aspek entitas yang dari pemegang pendapat dalam waktu tertentu.

Ide dari *opinion mining* ini sebenarnya sangatlah sederhana yaitu pendekatan yang dilakukan berfokus untuk menentukan aspek apa yang dievaluasi dalam pendapat dan apa yang user katakan tentang aspek ini. Akan tetapi, meskipun idenya sederhana, banyak masalah yang berhubungan dengan pengolahan bahasa bermunculan.

Dengan cara ini, opini yang dipandang positif atau negatif, sikap, emosi atau penilaian tentang suatu entitas atau aspek entitas dari pemegang opini dalam waktu tertentu diharapkan dapat diketahui. Konsep berikut juga diperkenalkan:

- a. Ekspresi entitas: Sesuai dengan kata yang sebenarnya atau frase yang ditulis oleh pengguna untuk menunjukkan atau menandakan suatu entitas. Akibatnya, entitas kemudian menjadi generalisasi dari setiap ungkapan entitas yang digunakan dalam dokumen yang dianalisis, atau realisasi tertentu dari suatu ekspresi entitas. Menurut Liu (2007) konsep ini disebut nama entitas.
- b. Ekspresi aspek: Sebagai sebuah ekspresi entitas, ini adalah kata atau frasa sebenarnya yang ditulis oleh pengguna untuk menandakan atau menunjukkan sebuah aspek. Dengan demikian, aspek juga konsep umum yang terdiri dari setiap ekspresi aspek. Itu dinamakan aspek oleh Bing Liu.

Kim et al. (2011) memberikan *review* yang baik dari sejarah dan perngembangan berbasis aspek. Para penulis menunjukkan bahwa proses ini umumnya terdiri dari tiga langkah yang berbeda, yang juga dipertimbangkan oleh Liu yaitu ekstraksi aspek, penentuan orientasi aspek, dan penyusunan hasil kesimpulan.

### 2.3. Praproses Data

Praproses data dilakukan sebelum melakukan klasifikasi. Pada tahap ini data masukan akan diolah sedemikian rupa sehingga memenuhi standar data yang diperlukan dalam penelitian. Praproses ini dilakukan agar tingkat akurasi hasil klasifikasi meningkat (Haddi, 2013). Hasil klasifikasi yang datanya melalui tahap praproses terlebih dahulu memiliki tingkat akurasi lebih tinggi dibandingkan dengan klasifikasi tanpa melakukan praproses data terlebih dahulu (Naradhipa, 2012). Kinerja pengklasifikasi tidak mampu menghasilkan hasil klasifikasi yang memuaskan karena tidak dilakukan praproses terlebih dahulu (Ma'ady, 2014).

Dalam Hu dan Liu (2004), Hu dan Liu menyajikan teknik berbasis di NLP dan statistik. Beberapa tahap praproses yang dilakukan yakni sebagai berikut:

- a. *Tokenizing* yakni proses penguraian deskripsi yang semula berupa kalimat menjadi kata-kata atau dengan kata lain melakukan pemotongan string masukan yakni teks, berdasarkan tiap kata yang menyusunnya.
- b. *part-of-speech* (POS) tagging berfungsi untuk pemberian label kelas kata pada suatu kata dalam kalimat.
- c. *Chunking* yaitu proses *syntax tree* parsing yang digunakan untuk menemukan kata benda dan frase kata benda atau NP.
- d. *Stopping* berdasarkan *stoplist* yang berisi *stopwords* yang telah ditentukan sebelumnya. *Stopwords* merupakan kosakata yang bukan merupakan kata unik atau ciri pada suatu dokumen. Kosakata yang dimaksudkan adalah kata penghubung dan kata keterangan yang bukan merupakan kata unik misalnya “sebuah”, “oleh”, “pada”, dan sebagainya. (Dragut, 2009).
- e. *Stemming* pada kata-kata yang tersisa pada dokumen teks untuk mendapatkan kata dasar. *Stemming* merupakan proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi bentuk kata dasarnya

(Tala, 2003). Proses ini akan menghilangkan awalan, akhiran, sisipan, dan *fixes* (kombinasi dari awalan dan akhiran). Proses ini memiliki banyak variasi karena imbuhan tiap bahasa berbeda-beda.

#### 2.4. Ekstraksi Aspek

Tahap ini bertujuan untuk menemukan dan mengekstrak topik penting dalam teks yang kemudian akan digunakan untuk meringkas. Setelah data *review* selesai dilakukan tahap praproses data, kemudian dengan menggunakan *frequent itemset mining*, kata benda yang paling sering dan NP yang diambil. Set diekstrak dari kata benda dan NP kemudian disaring menggunakan aturan linguistik khusus. Aturan-aturan ini memastikan bahwa persyaratan dalam aspek-aspek yang terdiri dari lebih dari satu kata yang mungkin mewakili benda nyata bersama-sama dan juga menghilangkan aspek yang berlebihan. Mereka juga mengambil *non-frequent aspects* menggunakan pendekatan dengan mencari kata benda atau NP yang muncul dekat dengan kata-kata pendapat dengan frekuensi tinggi. Pendekatan ini tidak mengambil kata sifat atau jenis lain dari aspek non objek.

Seperti didefinisikan oleh Liu (2007), aspek tidak langsung muncul dalam teks tetapi ada dalam cara aspek berekspresi. Dengan demikian, ketika mencoba untuk menerapkan model opini Liu untuk mengambil pendapat dari data nyata, konsepnya jadi agak membingungkan atau tidak jelas. Juga tidak jelas bagaimana mengelola aspek yang muncul lebih dari sekali dalam dokumen. Setelah melihat masalah ini, model untuk membangun tupel opini dari dokumen dogmatis telah dikembangkan di oleh Marrese-Taylor (2014).

Untuk lebih sederhananya, gunakan satu set dokumen beropini  $D_i = d_{i1}; d_{i2}; \dots; d_{in}$  dengan satu entitas,  $e_i$ . Hal ini bisa diasumsikan realistik karena opini yang biasanya tersedia dalam bentuk *review* produk di web. Kemudian, setiap dokumen opini akan berinteraksi dengan *review* atau pendapat yang diberikan oleh pemegang  $H_k$  dalam waktu  $T_k$ . Jadi  $S_{ik}$  adalah himpunan semua kalimat di  $d_{ik}$ , dengan  $S_{ik} = S_{ij1}; S_{ij2} \dots; S_{ijn}$ . Pendapat tentang  $e_i$  dalam  $d_{ik}$  akan diekspresikan pada entitas itu sendiri dan pada subset  $A_{ik}$  dari aspek tersebut. Demikian pula, setiap aspek  $A_{ik}$  akan muncul pada  $d_{ik}$  sebagai seperangkat aspek ekspresi  $AE_{ijk}$ , bagian dari  $AE_{ij}$ . Entitas  $e_i$  akan muncul sebagai bagian dari ekspresientitas yang berbeda  $EE_{ik} \subseteq$

EE<sub>i</sub>. Dengan demikian, set  $EX_{D_i}$  didefinisikan sebagai himpunan semua ekspresi aspek dari semua aspek dan semua ekspresi entitas muncul dalam  $D_i$ . Sebuah kalimat berhubungan dengan salah satu ekspresi aspek atau ekspresi entitas hanya jika muncul dalam kalimat itu. Berikutnya, orientasi sentimen perlu ditentukan untuk setiap pasangan hanya jika setiap aspek ekspresi atau ekspresi entitas muncul. Setelah menentukan orientasi sentimen,  $h_k$  dan  $t_k$  sesuai dokumen,  $d_{ik}$  harus ditambahkan dalam rangka membangun setiap tuple opini.

Di sisi lain, usulan Liu menunjukkan bahwa tampaknya cukup beralasan jika kata benda yang sering digunakan dalam ulasan produk biasanya asli dan merupakan ekspresi aspek penting karena ketika orang mengomentari aspek yang berbeda dari sebuah produk, kosa kata yang mereka gunakan biasanya menyatu. Namun demikian, dua alasan utama yang menjelaskan fakta bahwa banyak yang ekspresi yang berbeda bisa menunjukkan konsep yang sama, terutama dalam domain pariwisata :

- a. Prinsip ekonomi dalam bahasa (Vicentini 2003) menunjukkan bahwa mereka mencoba untuk mengatakan banyak hal menggunakan beberapa kata saja. Sebagai contoh, kata "Hotel ini memiliki wi-fi yang baik." Sesuai dengan *lexicalization*, di mana ekspresi yang asli, "Hotel ini memiliki akses internet yang baik melalui wi-fi.", Disingkat sesuai dengan prinsip ekonomi.
- b. Setiap bahasa menyediakan sistem yang mengatur konsepnya sendiri, juga mengejar penyederhanaan. Untuk alasan itu, banyak kata-kata dalam bahasa Inggris (seperti juga dalam semua bahasa lainnya) hanya merupakan hiponim dari hipernim yang ditentukan. Sebuah hiponim adalah kata atau frase yang mana bagian semantik termasuk dalam kata lain, yang menjadi hipernimnya. Contohnya, merah, merah terang, merah tua, dan merah total semua adalah hiponim merah (hiperninya), yang pada gilirannya adalah sebuah hiponim warna (Fromkin, Rodman, & Hyams, 2010).

Dalam prakteknya (Marrese-Taylor, 2014), menemukan aspek yang dievaluasi dalam satu set dokumen opini adalah tugas yang sangat kompleks. Bahkan, mendeteksi ekspresi aspek dari satu set dokumen dengan opini harus menjadi tugas yang sama sekali berbeda dari mendefinisikan atau mencari aspek nyata di dalamnya, karena jumlah kemungkinan ekspresi dimensi muncul dalam teks

benar-benar besar. Mengenai ini, mereka telah mengatakan bahwa dalam ulasan produk pariwisata beberapa ekspresi sebenarnya digunakan.

Masalah lain yang ditemukan dalam proposal Liu (2007) terkait dengan konsep kalimat dan jarak kata, yang meskipun banyak digunakan, tidak didefinisikan dengan jelas. Meskipun dengan analisis linguistik yang lebih dalam, di sini kita akan mendefinisikan kalimat sebagai set perintah, termasuk kata-kata dan tanda baca. Salah satu yang muncul dalam dua posisi yang berbeda harus dipertimbangkan dua kali, dengan memposisikan di mana mereka muncul dengan berbeda. Dengan kata lain, kalimat  $S$  akan sesuai dengan seperangkat unik tupel (perintah  $n$  ; posisi). Perbedaan antara dua komponen yang berdekatan harus sama dengan 1. Dengan demikian, konsep jarak kata antara dua unsur kalimat akan sesuai dengan perbedaan posisi dari dua token di  $S$ .

$$\text{Jarak Kata } (t_a, t_b) = \text{posisi } (t_a) - \text{posisi } (t_b) \mid t_a, t_b \in S \quad (2.1)$$

Jarak Kata  $(t_a; t_b)$  adalah metrik pada set  $S$  karena memenuhi kondisi non negatif, identitas yang tak dapat dibedakan, simetri dan ketidaksamaan segitiga. Perhatikan bahwa jarak minimal antara 2 elemen dalam  $S$  adalah 1, dan itu terjadi antara unsur-unsur yang berdekatan. Maksimum jarak sesuai dengan  $|S| + 1$ .

Meskipun ada definisi dan pembentukan ini, dalam pekerjaan ini kami fokus pada tugas menentukan orientasi sentimen di tingkat aspek, jadi di sini kita hanya menerapkan teknik yang dikembangkan oleh Hu dan Liu (2004) untuk mengekstrak aspek yang sering muncul. Dengan kata lain, dalam upaya untuk membuat analisis yang tersisa lebih sederhana, kita akan mempertimbangkan ekspresi aspek menjadi hanya kata benda atau set kata benda yang kita sebut aspek ekspresi eksplisit. Kami tidak akan mengambil ekspresi aspek implisit.

## 2.5. Penentuan Orientasi Aspek

Tahap berikutnya adalah penentuan orientasi aspek, untuk menentukan orientasi sentimen pada setiap aspek. Ding, Liu dan Yu menawarkan pendekatan leksikon dan aturan berbasis di Ding, Liu, dan Yu (2008). Metode ini bergantung pada kamus kata sentimen yang berisi daftar kata-kata positif dan negatif (disebut

opini kata) yang digunakan untuk mencocokkan istilah dalam teks dogmatis. Juga, karena kata-kata khusus lainnya juga bisa mengubah orientasi, aturan linguistik khusus diusulkan. Antara lain, aturan ini menganggap kata-kata negatif “tidak” dan juga beberapa pola negasi umum. Namun, meskipun bagaimana sederhana aturan-aturan ini mungkin muncul, penting untuk menangani mereka dengan hati-hati, karena tidak semua kejadian aturan tersebut atau penampakan kata akan selalu memiliki arti yang sama. Dalam konteks ini, aturan yang dikembangkan oleh Ding, Liu dan Yu mencakup fungsi skor agregasi untuk menentukan orientasi aspek dalam kalimat menggabungkan beberapa kata-kata opini.

Secara sederhana aturan untuk menentukan orientasi sentimen pendapat adalah akan dianggap bahwa pengguna memberikan pendapat atas aspek jika dalam kalimat terdapat ekspresi aspek yang muncul, perasaan positif atau negatif diungkapkan pada pendapat. Berdasarkan Marrese-Taylor (2013) Beberapa aturan yang menjadi pertimbangan nantinya adalah sebagai berikut:

- a. Kasus yang paling umum adalah ketika perasaan sebagian besar dinyatakan atau ditunjukkan oleh kata-kata sentimen khusus, kelompok kata atau referensi komparatif. Dalam hal ini, orientasi pendapat adalah langsung dan diberikan oleh orientasi kata-kata di dekatnya.
- b. Rekomendasi, terima kasih atau ucapan terima kasih dan selamat dianggap pendapat positif. ekspresi aspek muncul dalam kalimat-kalimat ini harus ditandai dengan orientasi positif.

Ketika kalimat menyatakan tidak adanya aspek (misalnya menggunakan kata-kata ini tidak ada atau klausa setara lainnya,) ekspresi aspek yang sesuai harus ditandai dengan orientasi negatif. Hal ini diusulkan karena dalam kasus ini, pengguna menyatakan bahwa entitas tidak memiliki komponen atau atribut (aspek) yang harus memiliki. Sebaliknya, ketika kalimat menyatakan kehadiran aspek, ekspresi aspek yang sesuai harus ditandai dengan orientasi positif.

Mengambil karya Ding et al. (2008) sebagai inspirasi, Marrese-Taylor (2014) menggunakan seperangkat aturan untuk menentukan orientasi kalimat telah dikembangkan, yang selalu mempertimbangkan kata-kata opini sebagai dasar.

### 2.5.1. Aturan Orientasi Kata

Pertama, kita perlu menentukan orientasi setiap kata pada sebuah kalimat. Marrese-Taylor (2014) menggunakan algoritma 1 yang mengandung seperangkat aturan linguistik, yang dijelaskan di bawah ini:

#### Algorithm 1. Word orientation

```
1: if word is in opinion_words then
2:   mark(word)
3:   orientation ← Apply Opinion Word Rule(marked_word)
4: else
5:   if word is in neutral_words then
6:     mark(word)
7:     orientation ← 0
8:   end if
9: end if
10: if word is near a too_word then
11:   orientation ← Apply Too Rules(orientation)
12: end if
13: if word is near a negation_word then
14:   orientation ← Apply Negation Rules(orientation)
15: end if
16: return orientation
```

- a. **Aturan kata:** Opini positif akan memiliki nilai 1, yang menunjukkan orientasi positif yang dinormalisasi, sedangkan yang negatif akan memiliki skor -1. Setiap kata benda dan kata sifat dalam setiap kalimat yang bukan opini akan memiliki skor 0 dan akan disebut kata netral .
- b. **Aturan negasi:** Sebuah kata negasi atau frase biasanya membalikkan opini yang diungkapkan dalam sebuah kalimat Akibatnya, kata opini atau kata-kata netral yang terpengaruh oleh negasi perlu diperlakukan secara khusus. Tiga aturan harus diterapkan: Negasi Negatif → Positif, Negasi Positif → Negatif dan Negasi Netral → Negatif. Kata negasi dan frase meliputi: “tidak”, “ tak”, “tidak pernah”, “enggak”, “tidak mau”, “tidak bisa”, “jangan”, “tidak akan mau”, “tidak akan pernah”, “tidak seharusnya”, “gak”.
- c. **Aturan terlalu:** kalimat dimana kata “terlalu” atau “berlebihan” muncul, juga ditangani secara khusus. Kapan sebuah kata pendapat atau sebuah kata netral muncul di dekat salah satu syarat yang disebutkan, dilambangkan kata terlalu, orientasi akan selalu menjadi negatif (skor=-1).

### 2.5.2. Aturan Orientasi Aspek

Memiliki aturan yang membantu dalam menentukan setiap orientasi kata di sebuah kalimat, menjelaskan bagaimana semua orientasi harus dikombinasikan untuk menentukan orientasi final dari kalimat pada aspek tertentu. Pada penelitiannya, Marrese-Taylor (2014) hanya menganggap kata-kata yang ditandai sebagai kata opini atau kata-kata netral, yang kita sebut kata-kata yang ditandai, karena kata itulah yang akan memberikan orientasi untuk setiap kalimat. Untuk proses lebih jelasnya dijelaskan pada algoritma 2 di bawah ini:

**Algorithm 2.** Opinion orientation

```
1: if but_word is in sentence then
2:   orientation ← Opinion
   Orientation(aspect,marked_words,but_clause)
3:   if orientation ≠ 0 then
4:     return orientation
5:   else
6:     orientation ← Opinion
   Orientation(aspect,marked_words,not but_clause)
7:   if orientation ≠ 0 then
8:     return -1 x orientation
9:   else
10:    return 0
11:  end if
12: end if
13: else
14:   for all aspect_position in aspect do
15:     for all aspect_word in aspect_position do
16:       for all word in marked_words do
17:         suborientation += Word Orientation(word) / WD(aspect
word,word)
18:       end for
19:       orientation += suborientation
20:     end for
21:     final_orientation += orientation
22:   end for
23:   if final_orientation > 0 then
24:     return 1
25:   else
26:     if final_orientation < 0 then
27:       return -1
28:     else
29:       return 0
30:     end if
31:   end if
```

32: **end if**

- a. **Aspect words aggregation rule:** s menjadi kalimat yang berisi kumpulan aspek ekspresi  $A=\{a_1, \dots, a_m\}$ , masing-masing dari mereka muncul hanya satu kali dalam s. Juga, sedangkan  $AW_i$  merupakan himpunan kata-kata aspek tersebut terdiri  $a_i$ , di mana  $AW_i=\{aw_{i1}, aw_{i2}, \dots, aw_{in}\}$ . Setiap  $aw_{ij}$  akan disebut kata aspek dan itu akan sesuai dengan aspek ekspresi  $a_i$ . Jika skor untuk setiap kata pendapat dan kata netral dalam s diketahui, skor untuk setiap  $aw_{ij}$  di s dihasilkan oleh fungsi agregasi berikut:

$$score(aw_{ij}, s) = \sum_{ow_j \in s} \frac{score(ow_j)}{WD(ow_j, aw_{ij})} \quad (2.2)$$

Dimana  $ow_j$  adalah kata opini atau kata netral di s,  $WD(ow_j, aw_{ij})$  adalah jarak kata antara kata aspek  $aw_{ij}$  dan kata opini  $ow_j$  di s. Baris 17 mengimplementasikan rumus ini dalam Algoritma 2. (Marrese-Taylor, 2014) mengambil fungsi ini dari Ding et al. (2008); Namun, proposisi mereka tidak memiliki penjelasan tentang bagaimana fungsi harus diterapkan untuk ekspresi aspek yang terdiri dari lebih dari satu kata (yang kita sebut *compound*). Kita telah melihat bahwa dalam produk pariwisata ulasan beberapa ekspresi aspek sebenarnya dalam *compound*. Misalnya, dalam kalimat "*The hotel had a poor view of the beautiful lake.*" Ekspresi aspek yang harus diambil oleh algoritma Liu adalah *lake view*. Namun, usulan Liu tidak menjelaskan bagaimana orientasi pada aspek ini harus diperoleh dalam kalimat. Untuk mempertimbangkan kasus ini, (Marrese-Taylor, 2014) mengusulkan bahwa formula tidak boleh digunakan untuk setiap ekspresi aspek melainkan untuk setiap kata di setiap ekspresi. Orientasi ini dikumpulkan sesuai dengan aturan berikutnya.

- b. **Aspect aggregation rule:** Untuk setiap ekspresi aspek compound  $a_i$  di s, orientasi akan dihitung mempertimbangkan nilai dari semua kata-kata yang membentuk itu,  $aw_{ij} \in AW_i$ , menurut persamaan berikut, yang dilaksanakan sejalan 19 Algoritma 2.

$$score(a_i, s) = \sum_{aw_{ij} \in AW_i} score(aw_{ij}, s) \quad (2.3)$$

- c. **Position aggregation rule:** Kita juga telah melihat bahwa dalam ulasan produk pariwisata ekspresi aspek bisa muncul lebih dari sekali dalam sebuah kalimat. Kasus ini tidak tercakup oleh proposal Liu, tetapi di sini kita memerlukan metode untuk menutupi kasus ini. Mengandaikan bahwa  $a_i$  muncul  $t$  kali dalam  $s$  dan mengetahui skor masing-masing aspek ekspresi penampilan  $a_{ik}$ ,  $k \in \{1, 2, \dots, t\}$ , (Marrese-Taylor, 2014) mengusulkan bahwa skor akhir dari  $a_i$ , atau  $fscore(a_i, S)$ , harus dihitung dengan hanya menambahkan nilai-nilai skor dari semua penampilan  $a_i$  di  $s$ , menurut persamaan berikut.

$$fscore(a_i, s) = \sum_{k=1}^t score(a_i^k, s) \quad (2.4)$$

rumus muncul di line 21 Algoritma 2. Perhatikan bahwa ketika  $a_i$  hanya muncul satu kali dalam  $s$ ,  $fscore(a_i, S) = score(a_i, S)$ . Akhirnya, baris 23-31 menunjukkan bagaimana orientasi dihitung sesuai dengan  $fscore$  dari setiap ekspresi aspek. Jika  $fscore(a_i, S)$  positif, opini ini dianggap positif pada  $a_i$  (baris 23 dan 24) dan jika negatif, pendapat dianggap negatif pada  $a_i$  (baris 26 dan 27). Jika tidak ada kasus-kasus ini terjadi, kalimat yang netral (garis 29) dipertimbangkan.

- d. **But clauses rules:** (Marrese-Taylor, 2014) menggunakan persis aturan yang sama diusulkan dalam Ding et al. (2008). Aturan ini menyatakan bahwa ketika sebuah kata  $a_i$  (termasuk kata tetapi atau sinonim apapun) muncul dalam kalimat  $s$ ,  $s$  harus dipecah menjadi dua segmen, satu sebelum dan satu setelah  $a_i$ . Jika orientasi setiap kata aspek  $a_{ij}$  muncul di segmen kalimat setelah  $a_i$  adalah nol, orientasi kemudian harus ditentukan dengan menggunakan segmen sebelum  $a_i$ , tapi menugaskan hasil yang berlawanan. (Marrese-Taylor, 2014) menyadari bahwa ambiguitas sedikit ada sejak dalam beberapa kasus  $a_{ij}$  bisa tampil di luar segmen dipertimbangkan. Di sini, kita hanya mengusulkan bahwa  $a_{ij}$  harus ditambahkan pada posisi akhir dari segmen yang sesuai untuk menghindari masalah konsistensi. Baris 1-12 pada Algoritma 2 menerapkan aturan ini.

## **2.6. Penyusunan Rangkuman Hasil Analisis**

Langkah terakhir dari *aspect-based opinion mining* adalah penyusunan rangkuman hasil analisis, untuk mempresentasikan hasil olahan dengan cara yang sederhana. Dalam konteks ini, output analisis yang disajikan adalah sumber informasi yang baik untuk menghasilkan ringkasan kuantitatif. Secara khusus, Liu mendefinisikan ringkasan tersebut berdasarkan aspek opini (Hu & Liu, 2004, 2006), yang terdiri dari bar chart yang menunjukkan jumlah pendapat positif dan negatif tentang setiap aspek dari satu entitas. Liu, Hu, dan Cheng (2005), juga mengusulkan bahwa *bar chart* dapat digunakan untuk membandingkan satu set produk yang dipilih, menunjukkan himpunan semua aspek produk yang dipilih pada grafik. Dalam hal ini, setiap batang atas atau di bawah sumbu x dapat ditampilkan di dua skala: (1) jumlah sebenarnya dari opini positif atau negatif dinormalisasi dengan jumlah maksimal opini yang objektif tentang fitur dari produk apapun dan (2) persen pendapat positif atau negatif, yang menunjukkan perbandingan dalam hal persentase dari ulasan positif dan negatif.

## **2.7. Klasifikasi Teks**

Klasifikasi dengan menggunakan pendekatan *machine learning* dengan metode *supervised learning* banyak digunakan dalam penelitian-penelitian klasifikasi sentimen. Dalam penelitian ini akan digunakan tiga teknik klasifikasi, yaitu Naïve Bayes, *support-vector machines* (SVM), dan *Random Forests* (RF) untuk mengetahui apakah hasil *object-based opinion mining* dapat digunakan untuk mengklasifikasikan orientasi *review* secara global dengan baik. Pada penelitian ini peneliti tidak akan membahas banyak tentang masing-masing teknik klasifikasi ini karena proses klasifikasi ini hanya akan menjadi penelitian tambahan saja untuk mengolah output dari *object-based opinion mining* untuk mengetahui performa dari hasil tersebut dalam klasifikasi orientasi sentimen secara umum.

### **2.7.1. Naïve Bayes**

*Naïve Bayes classifier* adalah classifier probabilistik terkenal serta sangat praktis, dan telah digunakan di banyak aplikasi. Ini mengasumsikan bahwa semua atribut dari contoh yang independen satu sama lain yang diberikan konteks kelas,

yaitu, asumsi independen. Beberapa studi menunjukkan bahwa Naive Bayes melakukan dengan sangat baik di banyak domain (Domingos dan Pazzani, 1997) terlepas dari asumsi independen yang salah.

### **2.7.2. Support Vector Machines (SVM)**

*Support vector machine* (SVM) merupakan sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi linear dalam sebuah ruang fitur berdimensi tinggi. SVM dapat dinyatakan sebagai gabungan klasifikasi *Nearest Neighbors* dan *regressi linier* (James, 2014). Tujuan dari SVM sendiri adalah untuk membuat sebuah batas yang disebut *hyperplane* (fungsi pemisah) terbaik diantara fungsi yang tidak terbatas jumlahnya yang menyebabkan tiap partisi menjadi homogen di tiap sisi. *Hyperlane* terbaik antara dua kelas dapat ditemukan dengan mengukur margin *hyperlane* tersebut dan mencapai titik maksimalnya. Adapun data yang berada pada bidang pembatas disebut *support vector*.

### **2.7.3. Random Forests**

Random forest merupakan pengembangan dari Decision Tree dengan menggunakan beberapa Decision Tree, dimana setiap Decision Tree telah dilakukan training menggunakan sampel individu dan setiap atribut dipecah pada tree yang dipilih antara atribut subset yang bersifat acak. Dan pada proses klasifikasi, individunya didasarkan pada vote dari suara terbanyak pada kumpulan populasi tree.

Random Forest yang dihasilkan memiliki banyak tree, dan setiap tree ditanam dengan cara yang sama. Tree dengan variabel x akan ditanam sejauh mungkin dengan tree dengan variabel y. Dan dalam perkembangannya, sejalan dengan bertambahnya data set, maka tree pun ikut berkembang. Penempatan tree yang saling berjauhan membuat apabila terdapat tree disekitar tree x berarti pohon tersebut merupakan perkembangan dari tree x (Rong Jia et al, 2009). Beberapa fungsi learning yang dihasilkan random forest digunakan strategi ensemble "bagging" untuk mengatasi masalah overfitting apabila dihadapkan data set yang kecil. Pada makalah ini Ensemble digunakan untuk melakukan resampled data

dengan mengklasifikasi ulang data outlier sehingga merubah struktur data set yang asli.

## 2.8. Pengukuran Performa

Dalam klasifikasi, terdapat beberapa cara pengukuran performa klasifikasi. Akan tetapi cara yang paling sering digunakan adalah dengan perhitungan *precision*, *recall*, dan akurasi (Jian Sheu, 2008). Nilai pengukuran kinerja tersebut didapat melalui perbandingan dari setiap nilai yang terdapat pada *confusion matrix*.

**Tabel 2.1 Confusion Matrix**

	<b>Predict Positive</b>	<b>Predict Negative</b>
<b>Actual Positive</b>	Tp	Fn
<b>Actual Negative</b>	Fp	Tn

Keterangan:

- TP (*True Positive*) : Jumlah klasifikasi yang benar dari data positif  
FP (*False Positive*) : Jumlah klasifikasi yang salah dari data negatif  
FN (*False Negative*) : Jumlah klasifikasi yang salah dari data positif  
TN (*True Negative*) : Jumlah klasifikasi yang benar dari data negatif

### 2.8.1. Akurasi

Akurasi adalah nilai perbandingan antara nilai data yang diklasifikasikan secara benar dengan seluruh data. Rumus perhitungannya adalah sebagai berikut.

$$\text{Akurasi} = \frac{Tp + Tn}{Tp + Fp + Fn + Tn} \quad (2.5)$$

### 2.8.2. Precision

*Precision* atau presisi adalah perbandingan antara jumlah data pada suatu kelas yang diklasifikasi secara benar dengan seluruh data pada kelas yang sama. Rumus perhitungannya adalah sebagai berikut.

$$\text{Presisi} = \frac{Tp}{Tp + Fp} \quad (2.6)$$

### 2.8.3. Recall

*Recall* adalah perbandingan antara jumlah data pada suatu kelas yang diklasifikasi secara benar dengan seluruh data yang diklasifikasi pada kelas yang sama.

$$Recall = \frac{Tp}{Tp+Fn} \quad (2.7)$$

### 2.8.4. F-Measure

*F-measure* merupakan parameter tunggal ukuran keberhasilan *retrieval* yang menggabungkan *recall* dan *precision*. (Rijsbergen, 1979).

$$F - measure = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (2.8)$$

Keterangan:

$\beta$  = parameter kepentingan relative aspek *precision* dan *recall*

$P$  = nilai *precision*

$R$  = nilai *recall*

Jika nilai  $\beta > 1$ , maka akan memberikan bobot kepentingan *recall* lebih tinggi daripada *precision*. Jika nilai  $\beta = 2$  maka akan bobot *recall* dua kali lebih besar daripada *precision*. Jika nilai  $\beta = 0.5$  maka bobot *precision* dua kali lebih besar daripada *recall*. Tetapi jika *recall* dan *precision* memiliki bobot yang sama maka  $\beta = 1$ , dan parameter *F-measure* dituliskan pada persamaan 2.9.

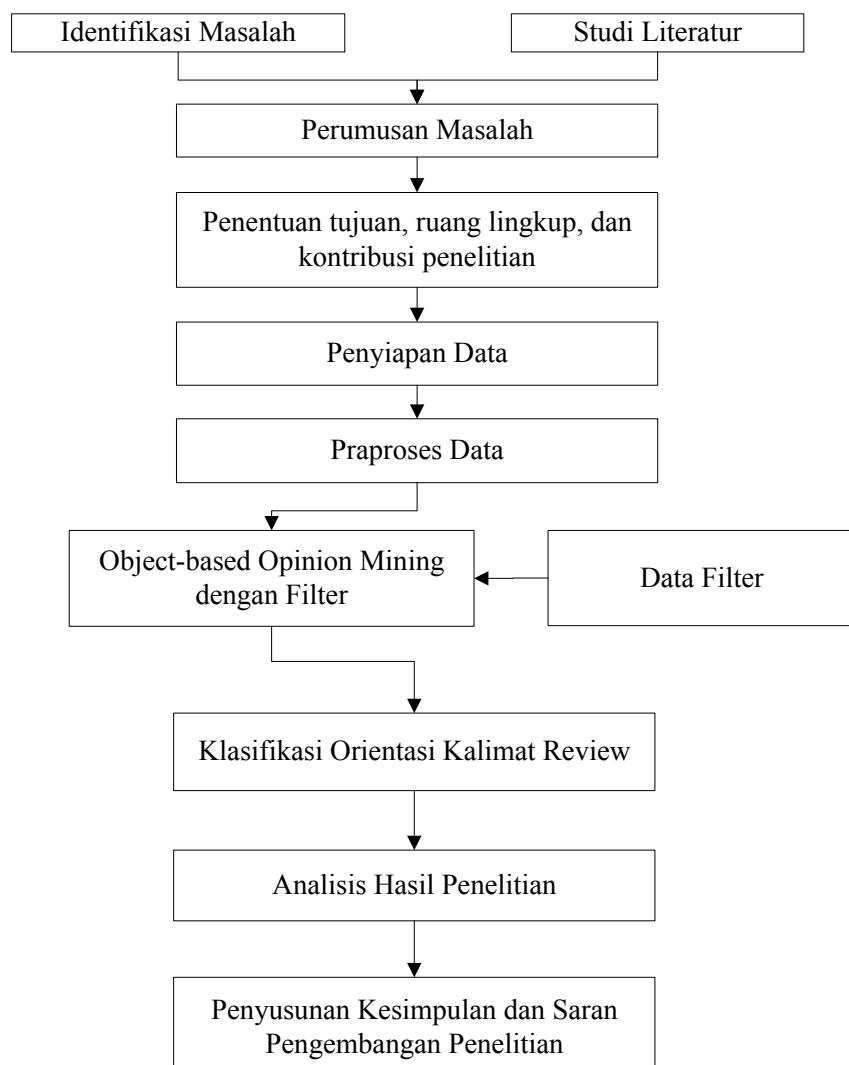
$$F - measure = \frac{2PR}{P+R} \quad (2.9)$$

Suatu sistem klasifikasi dinyatakan efektif jika hasil perhitungan menunjukkan ketepatan (*precision*) yang tinggi sekalipun *recall*-nya rendah.

## BAB 3

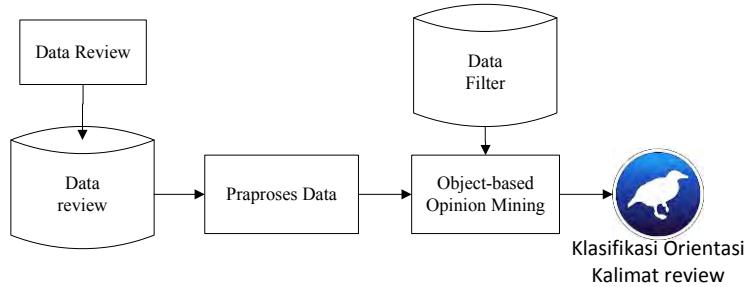
### METODOLOGI PENELITIAN

Dalam bab ini diuraikan tahap-tahap yang akan dilakukan pada penelitian ini. Secara umum, tahapan penelitian ini dapat dilihat pada gambar 3.1. Tahapan identifikasi masalah, perumusan masalah, dan penentuan tujuan, ruang lingkup, dan kontribusi penelitian telah dijelaskan pada bab 1. Sedangkan tahap studi literatur yang mencangkup landasan teori dan kajian pustaka telah dijelaskan pada bab 2. Pada bab 3 ini dijelaskan mengenai tahap praproses data, ekstraksi objek, penentuan orientasi opini, visualisasi hasil, analisis hasil penelitian, dan penyusunan kesimpulan dan saran pengembangan penelitian.



**Gambar 3.1 Metodologi Penelitian**

Sedangkan gambaran umum mengenai sistem yang dibangun dan digunakan pada penelitian ini dapat dilihat pada gambar 3.2 berikut ini. Setiap bagiannya akan dijelaskan lebih jauh pada bagian berikutnya.



**Gambar 3.2 Gambaran Umum Alur Data Penelitian**

### 3.1. Penyiapan Data

Terdapat beberapa data yang harus disiapkan dalam penelitian ini. Data-data tersebut merupakan data utama dan data pendukung dalam penggerjaan penelitian ini. Adapun data-data yang harus disiapkan adalah sebagai berikut:

#### a. Data *Review* dan Data Uji

Untuk mendapatkan hasil yang secara langsung dapat dibandingkan, maka pada penelitian ini digunakan data *review* dan data uji yang sama dengan yang digunakan pada penelitian Marrese-Taylor (2014). Data uji merupakan data *review* yang telah dilakukan analisis secara manual sehingga dapat digunakan untuk mengukur tingkat keberhasilan dari penelitian yang dilakukan. Contoh dari data uji dapat dilihat pada gambar 3.2.

line
1 /c1]/s1/ place[+], comfort[+][u], location[+][u] ### good place to stay at the end of a long flight in that it is very comfortable, with many facilities, and in the town , by the shore . 2 /c1]/s2/ ### however, puerto montt is not the best of places to explore. 3 /c1]/s3/ ### a better place is puerto varas which is just as near to the airport and has far more attractions. 4 /c1]/s4/ ### could not fault this aparthotel. 5 /c2]/s1/ hotel[+] ### my fiance and i spent three nights here in march 2012 and it's a sweet, quaint hotel. 6 /c2]/s2/ reservation[-] ### with that said, i called in early february 2012 to make a reservation and it got lost/misplaced.

**Gambar 3.3 Contoh Data Uji (Marrese-Taylor, 2014)**

### b. Opinion Lexicon

Data ini merupakan kumpulan dari kata-kata sentimen, dimana kata sentimen dikelompokkan ke dalam 2 kelompok yaitu kata positif dan kata negatif. Dalam penelitian ini digunakan opinion lexicon berbahasa Inggris sesuai dengan data review yang akan diolah. Data ini diambil dari website [www.cs.uic.edu](http://www.cs.uic.edu), dimana data ini terdiri dari 2.006 kata sentiment positif dan 4.783 kata sentiment negative.

### c. Data Filter

Dalam penelitian ini menggunakan data filter restoran dan hotel. Data ini digunakan untuk memfilter objek-objek diluar dari domain yang diteliti. Data filter yang digunakan dalam penelitian ini berasal dari berberapa sumber dan disusun ulang oleh peneliti sesuai dengan kebutuhan penelitian. Data filter yang digunakan terdiri dari 396 komponen yang berhubungan dengan hotel dan 260 komponen yang berhubungan dengan restoran.

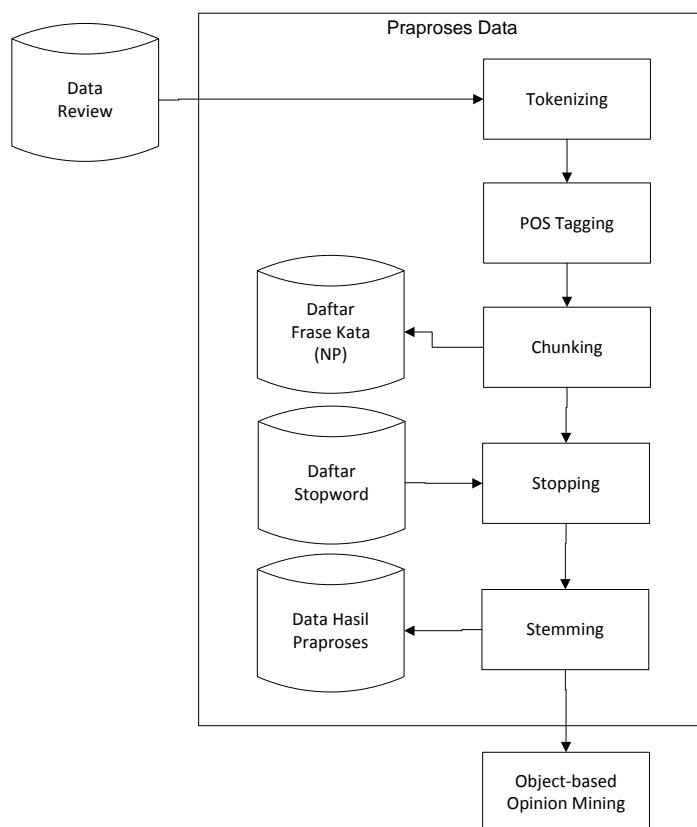
Data filter disusun berdasarkan data ontologi yang diperoleh di internet dimana data berupa file dengan format owl. Untuk membaca data ontologi tersebut maka akan digunakan Protégé. Protégé adalah platform gratis, open-source yang menyediakan komunitas pengguna dengan seperangkat alat untuk membangun model domain dan aplikasi berbasis pengetahuan dengan ontologi. Dari data ontologi tersebut penulis menyusun data filter sesuai dengan kebutuhan analisis dan dimasukkan ke dalam database. Hal ini dilakukan agar dapat mempermudah dalam membantu proses ekstraksi objek. Data ontologi hotel diperoleh dari [www.semanticweb.org](http://www.semanticweb.org) dan restoran diperoleh dari [gaia.fdi.ucm.es](http://gaia.fdi.ucm.es). Data ontologi tersebut lalu dipilih dan hanya dipakai yang dianggap sesuai dengan penelitian yang akan dilakukan. Selain itu peneliti juga menambahkan data filter berdasarkan dari data corpus untuk memaksimalkan hasil.

Belum adanya konsep dan landasan ilmiah yang pasti dalam menyusun dan menentukan data filter yang sesuai, sehingga proses penyusunan data filter yang digunakan dalam penelitian ini dilakukan secara

manual oleh peneliti, yaitu berdasarkan data ontologi yang didapat dari internet dan juga berdasarkan data aspek yang ada pada data corpus. Oleh karena itu apabila ada peneliti lain mencoba melakukan penelitian serupa maka kemungkinan besar data filter yang dihasilkan akan berbeda.

### 3.2. Praproses Data

Pertama kali yang dilakukan pada penelitian adalah menyimpan data *review* yang digunakan untuk penelitian perlu melalui tahap praproses sebelum data tersebut diolah lebih lanjut. Hal ini perlu dilakukan agar hasil klasifikasi memiliki tingkat akurasi yang lebih baik. Untuk semua tahapan praproses data ini peneliti akan menggunakan *NLTK libraries*, dimana untuk menggunakan *libraries* ini peneliti menggunakan Python sebagai bahasa pemrograman. Tahap praproses data pada penelitian ini terdiri dari 5 tahap, yaitu tokenizing, POS Tagging, Chunking, Stopping, dan Stemming. Adapun gambaran dari proses tersebut dapat dilihat pada gambar 3.4.



**Gambar 3.4 Tahapan Praproses**

Pertama-tama kalimat *review* yang sudah dipecah menjadi kalimat akan dipecah lagi menjadi perkata menggunakan teknik tokenizing. Kemudian akan dilakukan proses POS-Tagging untuk pemberian label kelas pada setiap kata. Part-of-speech tagging sangat penting dilakukan untuk dapat melakukan tahapan selanjutnya. Sesuai dengan yang dilakukan oleh Marrese-Taylor (2013 dan 2014), peneliti akan menggunakan NLTK libraries untuk mengurai setiap kalimat *review* untuk menghasilkan part-of-speech tag untuk setiap kata (apakah kata adalah kata benda, kata kerja, kata sifat, dll). Proses ini juga mengidentifikasi kata benda sederhana dan kata kerja. Gambar 3.5 menunjukkan contoh hasil dari kalimat dengan tag POS.

```
[('they', 'PRP'), ('gave', 'VBD'), ('us', 'PRP'), ('a', 'DT'), ('two', 'CD'), ('single', 'JJ'), ('beds', 'NNS'), (',', ','), ('instead', 'RB'), ('of', 'IN'), ('a', 'DT'), ('double', 'JJ'), ('bed', 'NN'), (',', ','), ('but', 'CC'), ('by', 'IN'), ('that', 'DT'), ('stage', 'NN'), (',', ','), ('we', 'PRP'), ('were', 'VBD'), ('too', 'RB'), ('tired', 'VBN'), ('to', 'TO'), ('care', 'VB'), (',', ','), ('and', 'CC'), ('we', 'PRP'), ('thought', 'VBD'), ('that', 'IN'), ('for', 'IN'), ('one', 'CD'), ('night', 'NN'), ('it', 'PRP'), ('was', 'VBD'), ("n't", 'RB'), ('worth', 'JJ'), ('complaining', 'VBG'), ('about', 'IN'), ('.', '.')]
```

**Gambar 3.5 Contoh hasil POS Tagging**

Proses selanjutnya adalah chunking, dimana digunakan untuk menemukan kata benda dan frase kata benda atau NP. Proses chunking dalam NLTK memiliki banyak rule yang bisa digunakan, tapi dalam penelitian ini peneliti hanya akan menggunakan dua rule saja. Rule-rule tersebut bisa dilihat pada tabel 3.1.

**Tabel 3.1 Detail Data Corpora**

Rule	Penjelasan	Contoh NP	Hasil Chunking
{<JJ>+<NN>}	kata sifat diikuti dengan kata benda	hot water	(hot,JJ),(water,NN)
{<NN>+<NN>}	kata benda diikuti dengan kata benda	tennis court	(tennis,NN), (court,NN)

Hasil NP yang akan dianggap sebagai objek adalah NP yang terdiri dari maksimal tiga suku kata, karena berdasarkan Hu dan Liu (2004), mereka percaya

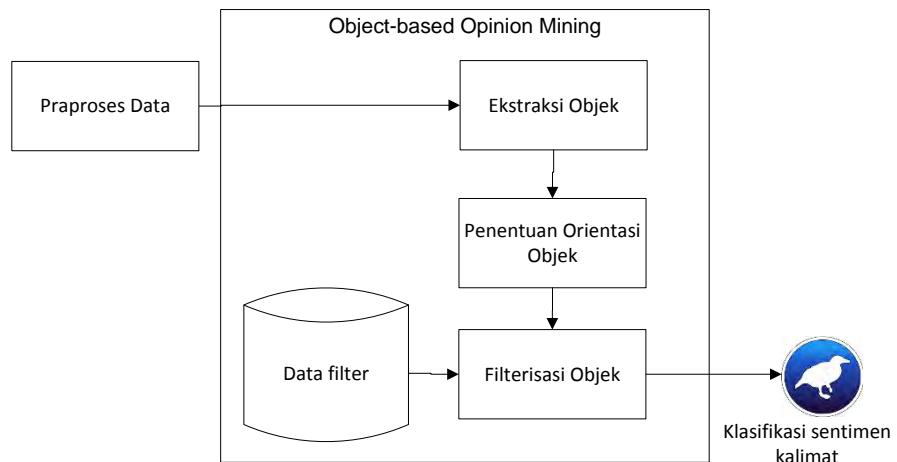
bahwa fitur produk tidak mengandung lebih dari tiga kata. Setelah proses chunking selesai maka, hasilnya akan disimpan kedalam database untuk mempermudah dalam proses selanjutnya.

Tahap yang dilakukan selanjutnya adalah melakukan proses stopping, dimana pada tahap ini bertujuan untuk menandai stopword yang ada pada data hasil POS Tagging, sehingga kata yang merupakan stopword akan diabaikan dalam proses penentuan orientasi kata. Kemudian, tahap berikutnya adalah proses stemming untuk menghilangkan awalan, akhiran, sisipan, dan *suffixes*. Pada penelitian ini peneliti menggunakan *Porter Stemmer* yang tersedia di NLTK. Setelah semua tahapan praprosesing data selesai, maka hasilnya akan disimpan dalam database.

### **3.3. *Object-based Opinion Mining***

Bagian utama dari penelitian ini adalah tahap *object-based opinion mining*, dimana data *review* yang sudah melalui praproses data akan diolah. Tujuan utama dari proses ini adalah untuk menemukan objek apa saja yang dikomentari di dalam *review* dan menyimpulkan apakah komentar tersebut positif atau negatif. Dalam *object-based opinion mining* dibagi menjadi dua proses analisis, yaitu proses ekstraksi objek dan proses penentuan orientasi objek. Sehingga setelah proses analisis selesai dapat dilakukan penyusunan rangkuman hasil analisis yang diharapkan menghasilkan informasi dapat dibaca dengan mudah oleh semua pihak.

Dalam penelitian ini penulis tidak akan berfokus kepada proses penyusunan rangkuman hasil analisis, melainkan lebih berfokus untuk memodifikasi *object-based opinion mining* dengan memanfaatkan filter dalam proses ekstraksi objek. Untuk proses visualisasi hasil *object-based opinion mining* dapat dilihat pada penelitian yang telah dilakukan oleh Marrese-Taylor (2014). Adapun gambaran dari proses *object-based opinion mining* dapat dilihat pada gambar 3.6.



**Gambar 3.6 Tahapan *Object-based Opinion Mining***

### 3.3.1. Ekstraksi Objek

Tahap ini bertujuan untuk menemukan dan mengekstrak topik penting dalam teks *review* yang kemudian akan digunakan untuk membagi-bagi opini kedalam sub-sub berdasarkan objek yang ditemukan. Pertama, pendapat yang diberikan oleh pengguna bisa atau tidak bisa secara eksplisit merujuk pada objek dan sebuah pendapat bisa menggunakan seluruh rangkaian ekspresi yang berbeda (Marrese-Taylor, 2013). Menimbang bahwa objek adalah atribut atau komponen dari suatu entitas, nama objek berupa kata benda atau frase kata benda.

Algoritma yang digunakan Marrese-Taylor (2014) untuk ekstraksi ekspresi objek, berdasarkan seringnya kata benda dan NP muncul dalam ulasan, mendapatkan kinerja yang buruk dalam domain pariwisata. Hasil ini menunjukkan bahwa pada kenyataannya beberapa ekspresi berbeda dapat digunakan untuk menunjukkan atribut atau komponen yang sama pada produk pariwisata. Oleh karena itu, tidak hanya kata-kata yang paling sering perlu dipertimbangkan ketika mengekstrak ekspresi objek untuk mencapai recall yang lebih baik untuk tugas ini. Demikian juga, fakta bahwa pengguna cenderung untuk menceritakan kisah tentang produk pariwisata menyebabkan presisi yang dihasilkan kurang bagus dalam tugas penggalian ekspresi objek karena dalam ulasan banyak objek yang bukan merupakan komponen atau atribut dari produk yang disebutkan.

Dikarenakan hasil yang diperoleh oleh Marrese-Taylor (2014) dalam hal ekstraksi fitur masih kurang memuaskan, maka pada penelitian kali ini peneliti akan

mencoba untuk membandingkannya dengan menggunakan data filter untuk mengetahui mana yang akan mendapatkan hasil yang lebih baik. Untuk membuktikan apakah pemanfaatan data filter akan mendapatkan hasil yang lebih baik, maka dalam penelitian ini penulis akan mencoba untuk melakukan lima pendekatan berbeda dalam melakukan proses ekstraksi objek. Pendekatan-pendekatan tersebut yaitu sebagai berikut:

- a. **Frequent Object:** pada pendekatan ini peneliti hanya akan menggunakan *frequent object* saja yang akan diidentifikasi sebagai kandidat objek dan di analisis orientasi sentimennya.
- b. **Frequent Object and Infrequent Object:** pada pendekatan ini peneliti akan menggunakan *frequent object* dan *infrequent object* yang akan diidentifikasi sebagai kandidat objek dan di analisis orientasi sentimennya.
- c. **Frequent Object with Filter:** pendekatan ini hampir sama dengan pendekatan *frequent object*, akan tetapi setelah *frequent object* diketahui maka akan dilakukan filterisasi dengan menggunakan data filter. *Frequent object* yang tidak terdapat dalam daftar filter akan digugurkan dari daftar kandidat objek. Proses filterisasi akan dilakukan saat orientasi objek sudah diketahui.
- d. **Frequent Object and Infrequent Object with Filter:** pendekatan ini hampir sama dengan pendekatan *frequent object and infrequent object*, akan tetapi setelah *frequent object* dan *infrequent object* diketahui maka akan dilakukan filterisasi dengan menggunakan data filter. *Frequent object* dan *infrequent object* yang tidak terdapat dalam daftar filter akan digugurkan dari daftar kandidat objek. Proses filterisasi akan dilakukan saat orientasi objek sudah diketahui.
- e. **Filter Object:** pada pendekatan ini proses ekstraksi objek dilakukan dengan cara langsung mencocokkan tiap kata dalam kalimat *review* dengan data filter.

Pada penelitian ini yang disebut sebagai *frequent object* adalah objek yang kemunculannya dalam seluruh *review* minimal 1% dari jumlah kalimat *review*. *frequent object* merupakan objek yang paling menarik dan sering dikomentari oleh orang lain. Jika kata benda atau frase kata benda muncul kurang dari 1% maka akan dianggap sebagai *infrequent object*.

### **3.3.2. Penentuan Orientasi Objek**

Tahap ini berfungsi untuk menentukan orientasi sentimen pada setiap objek. Proses pada tahap ini bergantung pada kamus kata sentimen yang berisi daftar kata-kata positif dan negatif yang digunakan untuk mencocokkan istilah dalam teks.

Proses penentuan orientasi sentimen pada objek juga perlu penjelasan lebih dalam. Liu (2007) membuat asumsi bahwa setiap kalimat mengungkapkan pendapat dari pemegang pendapat, yaitu hanya informasi yang terkandung dalam kalimat di mana objek muncul akan dipertimbangkan untuk menentukan apa yang pengguna ceritakan tentang objek ini. Kita akan menganggap bahwa pengguna memberikan pendapat atas objek jika dalam kalimat di mana objek muncul (secara implisit atau eksplisit) perasaan positif atau negatif. Perasaan ini sebagian besar dinyatakan atau ditunjukkan oleh kata-kata khusus, kelompok kata, referensi komparatif atau bahkan tanda baca. Namun demikian, lebih umum, deskripsi, penilaian atau rekomendasi akan di sini dianggap sebagai pendapat.

Pada tahapan ini peneliti akan menggunakan cara yang sama dengan apa yang telah dilakukan oleh Marrese-Taylor (2014) menggunakan seperangkat aturan untuk menentukan orientasi objek. Aturan tersebut dibagi menjadi dua tahapan yaitu aturan orientasi kata dan aturan orientasi objek. Aturan orientasi kata bertujuan untuk menentukan orientasi setiap kata pada sebuah kalimat. Sedangkan aturan orientasi objek membantu dalam menentukan setiap orientasi kata di sebuah kalimat, menjelaskan bagaimana semua orientasi harus dikombinasikan untuk menentukan orientasi final dari kalimat pada objek tertentu. Untuk penjelasan secara detail mengenai masing-masing aturan sudah dijelaskan pada bab 2.

### **3.4. Klasifikasi Orientasi Kalimat Review**

Setelah seluruh proses *object-based opinion mining* selesai, maka peneliti akan mencoba melakukan penelitian lanjutan untuk menganalisis apakah hasil objek dan orientasinya yang berhasil diekstrak dapat digunakan untuk melakukan klasifikasi untuk menentukan orientasi sentimen *review* secara global tiap kalimat. Untuk dapat melakukan analisis ini maka peneliti akan mencoba melakukan 7 skenario uji coba yang berbeda untuk masing-masing data *review* hotel dan restoran. Peneliti akan menggunakan WEKA sebagai tool untuk melakukan proses

klasifikasi. Dari proses klasifikasi ini nantinya akan menghasilkan kesimpulan, apakah hasil *object-based opinion mining* yang telah dihasilkan sebelumnya dapat digunakan untuk menentukan orientasi sentimen data *review* secara global.

### **3.5. Uji Coba dan Analisis Hasil Penelitian**

Pada bagian ini dijelaskan mengenai skenario uji coba dan analisisnya. Skenario uji coba ini merupakan rencana uji coba sehingga analisis dari uji coba yang dilakukan dapat menjawab rumusan masalah dan tujuan yang telah ditetapkan sebelumnya. Penerapan proses analisis akan menggunakan bahasa pemrograman Python untuk praproses data dan PHP untuk analisis *object-based opinion mining*. Sedangkan untuk penyimpanan data menggunakan database MySQL. Untuk proses terakhir yaitu proses klasifikasi orientasi *review* menggunakan WEKA sebagai tool.

#### **3.5.1. Skenario Uji Coba**

Untuk menjawab rumusan masalah dan tujuan yang ditentukan pada awal penelitian yang telah dijelaskan pada bab 1, maka dirancang skenario uji coba. Uji coba dilakukan juga untuk melakukan evaluasi terhadap performa algoritma dalam menganalisis data opini.

Untuk mendapatkan hasil yang secara langsung dapat dibandingkan, maka pada penelitian ini digunakan data *review* dan data uji yang sama dengan yang digunakan pada penelitian Marrese-Taylor (2014). Data uji merupakan data *review* yang telah dilakukan analisis secara manual sehingga dapat digunakan untuk mengukur tingkat keberhasilan dari penelitian yang dilakukan.

Marrese-Taylor (2014) mengunduh semua ulasan dari hotel dan restoran dalam bahasa Inggris tentang Lake District di TripAdvisor. Mereka memperoleh total 1.435 ulasan dan disimpan dalam dua file CSV yang berbeda. Dalam rangka untuk menghasilkan corpora untuk mengevaluasi kinerja dari algoritma, *review* dipilih secara acak 100 ulasan restoran dan hotel. Kemudian, setiap *review* dibagi menjadi kalimat. Akhirnya, setiap kalimat itu secara manual ditandai. Kalimat yang tampak ambigu atau benar-benar sulit untuk tag dibahas dengan annotator manusia kedua, seorang ahli dalam linguistik. Setelah kesepakatan tercapai, kalimat itu ditandai sesuai dengan kesepakatan itu. Detail data *corpora* yang akan digunakan

dapat dilihat pada tabel 3.2 dan untuk detail objek yang berhasil ditemukan dalam *corpora* dapat dilihat pada tabel 3.3.

**Tabel 3.2 Detail Data Corpora**

Corpus	Hotels	Restaurants
Reviews	100	100
Total sentences	789	470
Opinion sentences	609	368
Opinion sentences/sentences	77.19%	78.3%

**Tabel 3.3 Detail Objek yang Ditemukan dalam Corpora**

Object type	Hotels corpus		Restaurants corpus	
	Number	Percentage	Number	Percentage
Explicit	229	73.87	161	67.93
Explicit and implicit	30	9.68	26	10.97
Implicit	51	16.45	50	21.1
Total	310	100	237	100

Uji coba pada penelitian ini dibagi menjadi dua tahap. Uji coba tahap pertama bertujuan untuk menguji performa dari proses analisis *object-based opinion mining* terhadap masing-masing jenis pendekatan yang berbeda seperti yang telah dijelaskan pada bab 3.3.1. Terdapat tiga skenario uji coba pada tahap pertama ini. Skenario pertama ditujukan untuk menguji proses *object extraction*. Skenario kedua berfungsi untuk menguji *opinion sentence detection*. Dan skenario ketiga berguna untuk menguji peran *object sentiment classification*.

Setelah pengujian terhadap performa *object-based opinion mining* selesai, tahap uji coba selanjutnya adalah untuk menguji hasil objek beserta orientasinya apakah dapat digunakan untuk menentukan orientasi sentiment *review* secara global dengan mengklasifikasikannya ke dalam tiga kelas, positif, netral, dan negatif. Pada tahap uji coba ini semua hasil pendekatan dalam *object-based opinion mining* akan dilakukan proses klasifikasi dengan menggunakan WEKA. Dalam uji coba ini

peneliti akan membandingkan performa dari tiga metode klasifikasi yang berbeda untuk memprediksi orientasi sentimen kalimat *review*. Metode yang digunakan adalah Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Semua jenis metode ini sudah tersedia dalam WEKA.

Dalam uji coba tahap kedua, agar data hasil *object-based opinion mining* dapat diolah didalam WEKA maka data objek harus diekstrak kedalam format CSV terlebih dahulu. Selain itu berdasarkan data corpus peneliti terlebih dahulu melakukan manual tag terhadap data *review* dan menentukan tiap-tiap kalimat yang ada dalam data corpus kedalam tiga kelas, yaitu positif, netral, dan negatif. Nantinya akan ada tujuh set data yang akan dilakukan uji coba untuk masing-masing skenario dan masing-masing produk. Tujuh set data tersebut adalah sebagai berikut:

- a. **Review Sentence:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah kalimat *review*.
- b. **Data Corpus Object:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah aspek dan orientasi yang ada pada data corpus dimana aspek ini merupakan hasil manual tag dari penelitian Marrese-Taylor (2014).
- c. **Frequent Object:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah *frequent object* lengkap dengan orientasinya yang merupakan hasil dari proses *object-based opinion mining*.
- d. **Frequent Object and Infrequent Object:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah *frequent object* dan *infrequent object* lengkap dengan orientasinya yang merupakan hasil dari proses *object-based opinion mining*.
- e. **Frequent Object with Filter:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah *frequent object* lengkap dengan orientasinya yang merupakan hasil dari proses *object-based opinion mining* yang telah terfilter oleh data filter.
- f. **Frequent Object and Infrequent Object with Filter:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah *frequent object* dan *infrequent object* lengkap dengan orientasinya yang merupakan hasil dari proses *object-based opinion mining* yang telah terfilter oleh data filter.

- g. **Filter Object:** data yang akan digunakan dalam proses klasifikasi sebagai atribut adalah objek yang muncul dalam *review* yang merupakan object yang ada dalam data filter beserta orientasinya yang berhasil diperoleh dari proses *object-based opinion mining*.

Pada tahap klasifikasi untuk memprediksi orientasi sentiment *review* secara global menggunakan WEKA, pengukuran performa dilakukan menggunakan *k-fold cross-validation* yakni menggunakan 10 *folds*. Dimana semua data yang terkumpul akan dibagi ke dalam sepuluh kelompok (10 *folds*) atau dapat disebut sebagai 10-*cross-validation* (Refaeilzadeh, Tang, & Liu, 2009). Secara bergantian sembilan kelompok akan dijadikan data latih dan satu kelompok akan dijadikan data uji. Pembagian data latih dan uji dilakukan secara acak oleh pengklasifikasi.

### 3.5.2. Analisis Hasil Penelitian

Berdasarkan hasil pada percobaan yang telah dirancang sesuai skenario uji coba, diakukan analisis untuk mengetahui performa dari proses *object-based opinion mining* dan proses klasifikasi orientasi kalimat *review*. Semua hasil pengujian yang telah dilakukan tersebut akan saling dibandingkan untuk mengetahui pendekatan manakah yang menghasilkan hasil yang lebih baik.

Karena dalam penelitian ini dalam *object-based opinion mining* tidak menggunakan data training maka untuk evaluasi, maka dilakukan analisis secara manual untuk mendapatkan data uji. Untuk setiap kalimat dalam tinjauan, jika hal itu menunjukkan pendapat pengguna, semua fitur yang pengkaji telah nyatakan/opininya ditandai. Apakah pendapat positif atau negatif juga diidentifikasi. Jika pengguna tidak memberikan pendapat dalam sebuah kalimat, kalimat tidak ditandai seperti kita hanya tertarik pada kalimat dengan pendapat dalam pekerjaan ini. Semua hasil yang dihasilkan oleh sistem dibandingkan dengan tag secara manual. *Tagging* cukup mudah untuk kedua fitur produk dan opini. Sebuah komplikasi minor mengenai fitur tagging adalah bahwa fitur dapat eksplisit atau implisit dalam sebuah kalimat. Kebanyakan fitur muncul secara eksplisit dalam kalimat pendapat. Beberapa fitur mungkin tidak muncul dalam kalimat yang disebut fitur implisit. Kedua fitur eksplisit dan implisit mudah untuk mengidentifikasi dengan *tagger* manusia. Masalah lain adalah bahwa menilai pendapat dalam ulasan

dapat agak subjektif. Hal ini biasanya mudah untuk menilai apakah pendapat positif atau negatif jika kalimat jelas mengungkapkan pendapat. Pada penelitian ini tidak perlu dilakukan analisis secara manual, karena peneliti akan menggunakan data set yang sama persis dengan yang digunakan oleh Marrese-Taylor (2014).

Setelah itu akan dilakukan perhitungan *accuracy* (mengukur tingkat kedekatan antara hasil system dengan hasil manual), *precision* (menghitung tingkat ketepatan hasil algoritma), *recall* (mengukur tingkat keberhasilan sistem) dan *f-measure* (menampilkan timbal balik antara recall dan precision).

Untuk proses klasifikasi orientasi kalimat *review* menggunakan WEKA analisis hasil dapat dilakukan dengan lebih sederhana karena output hasil klasifikasi sudah terdapat hasil perhitungan *accuracy*, *precision*, *recall*, dan *f-measure*. Jadi peneliti tinggal membandingkan hasilnya untuk menyimpulkannya data set yang mana yang menghasilkan hasil prediksi paling bagus.

### **3.6. Penyusunan Kesimpulan dan Saran Pengembangan Penelitian**

Setelah seluruh rangkaian proses penelitian telah selesai dilaksanakan, maka kesimpulan terhadap hasil analisis penelitian disusun. Kesimpulan pada penelitian ini akan merangkum jawaban dari permasalahan dan menjawab tujuan yang telah ditentukan pada awal penelitian mulai dirancang. Selain itu, saran pengembangan terhadap penelitian ini juga disusun pada tahap ini.

## **BAB 4**

### **UJI COBA DAN ANALISIS HASIL**

Bab ini menjelaskan mengenai proses uji coba terdapat kinerja aplikasi yang telah dibangun dan analisis dari hasil yang didapatkan. Proses uji coba meliputi lingkungan uji coba, skenario uji coba, pelaksanaan dan hasil uji coba, serta analisis hasil uji coba.

#### **4.1. Data Uji Coba**

Data set yang digunakan dalam penelitian ini adalah data corpus yang dihasilkan oleh Marrese-Taylor (2014) dalam penelitiannya. Mereka mengunduh semua data *review* dari hotel dan restoran yang tertulis dalam bahasa Inggris tentang *Lake District* di TripAdvisor. Mereka memilih secara acak 100 *review* tentang hotel dan restoran dan memecahnya kedalam kalimat-kalimat dan setiap kalimat dilakukan ekstraksi objek secara manual lengkap dengan orientasi masing-masing objek. Setelah dipecah, data *review* hotel terdiri dari 789 kalimat dan data *review* restoran menjadi 470 kalimat. Hasil ekstraksi manual inilah yang nantinya digunakan untuk menguji performa dari *object-based opinion mining*. Jadi dalam penelitian ini peneliti tidak perlu lagi melakukan ekstraksi objek secara manual.

Lain halnya dengan uji coba pada proses klasifikasi untuk menentukan orientasi kalimat berdasarkan objek, untuk proses ini peneliti diharuskan melakukan klasifikasi secara manual dengan menandai setiap kalimat *review* apakah memiliki orientasi positif, netral, atau negatif. Dalam tahapan penelitian ini, perhitungan perfoma pengklasifikasi digunakan metode *cross validation* dengan jumlah *fold* adalah 10 *folds*. Pembagian data latih dan data uji dilakukan acak dan otomatis oleh aplikasi perangkat lunak WEKA 3.6.

#### **4.2. Lingkungan Uji Coba**

Lingkungan uji coba merupakan kriteria perangkat pengujian yang digunakan dalam menguji sistem yang dibangun pada penelitian ini. Lingkungan

uji coba terdiri dari perangkat keras dan perangkat lunak. Adapun perangkat keras yang digunakan ditunjukkan pada tabel 4.1.

**Tabel 4.1 Spesifikasi Lingkungan Uji Coba – Perangkat Keras**

Perangkat Keras	Spesifikasi
Jenis	PC
Processor	AMD A10
RAM	16 GB

Selain perangkat keras juga digunakan beberapa aplikasi perangkat lunak untuk uji coba dalam penelitian ini yang ditunjukkan pada tabel 4.2

**Tabel 4.2 Spesifikasi Lingkungan Uji Coba - Perangkat Lunak**

Perangkat Lunak	Spesifikasi
Sistem Operasi	Windows 7
Bahasa Pemrograman	Python dan PHP
Database	MySQL
Tools	<ul style="list-style-type: none"><li>• XAMPP Windows Ver 1.8.3</li><li>• Python 3.4.2</li><li>• NLTK 3.0</li><li>• Weka 3.6</li><li>• Protégé</li></ul>

### **4.3. Skenario Uji Coba**

Skenario uji coba dibuat agar dapat mencapai tujuan penelitian. Uji coba dalam penelitian ini secara umum yakni uji coba proses *object-based opinion mining*. Selain itu juga uji coba proses klasifikasi untuk memprediksi orientasi kalimat *review* berdasarkan objek hasil proses *object-based opinion mining*. Penjelasan lebih detil dijelaskan pada masing-masing bagian.

#### **4.3.1. Uji Coba Ekstraksi Objek**

Uji coba tahap ini dilakukan untuk mengukur performa dari *object-based opinion mining* dalam hal ekstraksi objek. Hasil objek yang berhasil diekstrak akan

dibandingkan dengan data objek yang diperoleh dari proses ekstraksi objek manual yang sudah tersedia pada data corpus yang digunakan dalam penelitian ini. Uji coba akan dilakukan berdasarkan lima pendekatan ekstraksi objek yang berbeda yang sudah dijelaskan pada bab 3.

#### **4.3.2. Uji Coba Klasifikasi Sentimen Objek**

Uji coba ini dilakukan untuk mengukur performa dari *object-based opinion mining* dalam hal penentuan orientasi sentimen untuk masing-masing objek dalam tiap kalimat *review*. Uji coba ini akan membandingkan antara hasil analisis *object sentiment classification* dengan data korpus. Pengujian akan dibagi menjadi dua tahap. Pertama, pengujian hanya akan menggunakan objek yang sesuai dengan objek yang ada pada data corpus saja. Kedua pengujian akan menggunakan seluruh objek yang berhasil terekstrak. Pembagian ini dilakukan dengan tujuan untuk membuktikan perbedaan hasil secara lebih detail dari penggunaan filter dalam *object-based opinion mining*. Uji coba ini akan menghitung kebenaran orientasi objek yang terdapat pada data korpus, apakah orientasi objek tersebut benar atau tidak pengklasifikasianya.

#### **4.3.3. Uji Coba Deteksi Kalimat Opini**

Uji coba ini dilakukan untuk mengukur performa dari *object-based opinion mining* dalam melakukan penentuan apakah suatu kalimat merupakan kalimat opini atau tidak. Uji coba ini akan membandingkan antara hasil analisis klasifikasi sentimen objek dengan data korpus. Kalimat yang berhasil mendeteksi object beserta orientasinya akan dianggap sebagai kalimat yang mengandung opini, begitu juga sebaliknya kalimat yang tidak terdapat objek dengan orientasi sentimennya akan dianggap sebagai kalimat yang tidak mengandung opini.

#### **4.3.4. Uji Coba Klasifikasi Sentimen Kalimat Review**

Pada tahap uji coba ini semua hasil pendekatan dalam *object-based opinion mining* akan dilakukan proses klasifikasi dengan menggunakan WEKA. Dalam uji coba ini peneliti akan membandingkan performa dari tiga teknik klasifikasi yang berbeda untuk memprediksi orientasi sentimen kalimat *review*. Metode yang

digunakan adalah Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Uji coba ini bertujuan untuk mengetahui teknik klasifikasi mana yang mendapatkan hasil lebih baik pada kasus ini.

#### **4.4. Pelaksanaan Uji Coba**

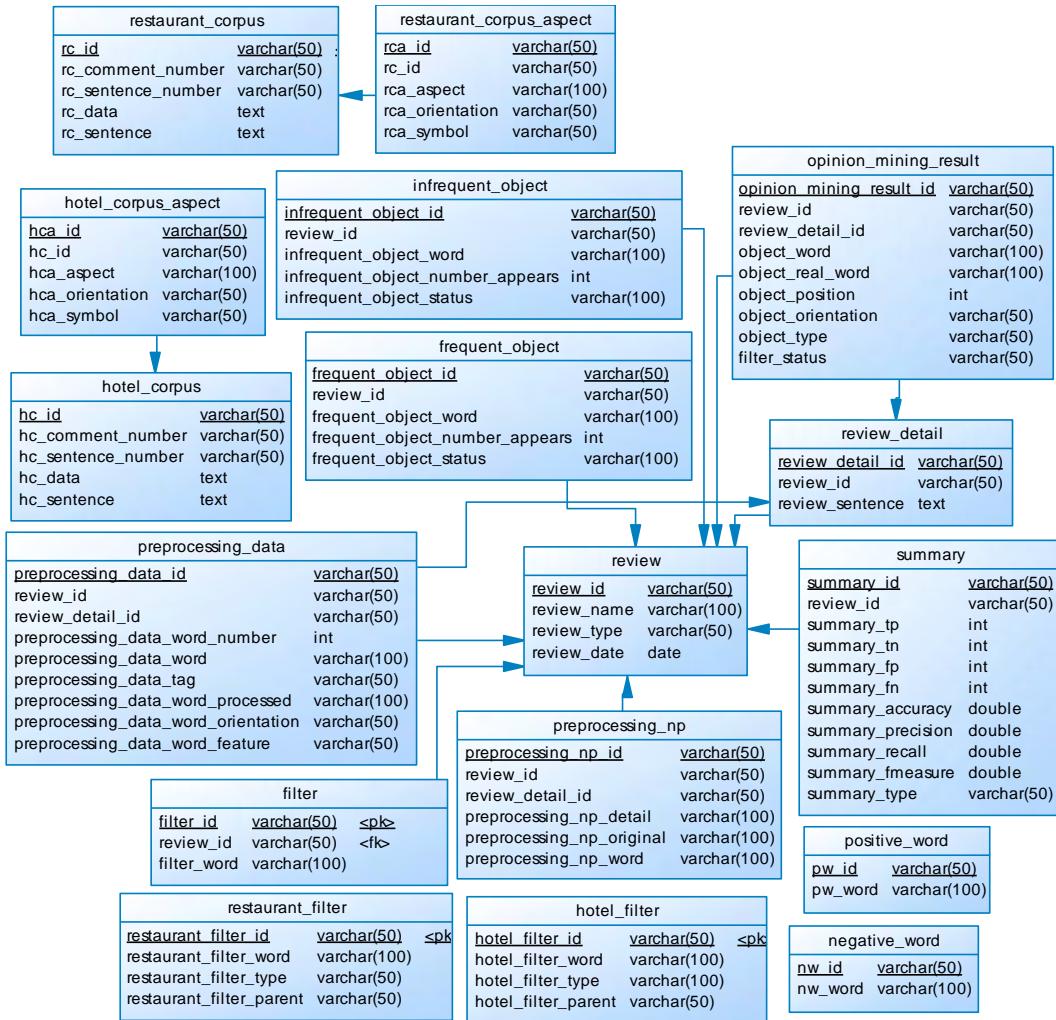
Pada bagian ini akan dijelaskan mengenai detail proses dari uji coba yang dilakukan sesuai dengan masing-masing skenario uji coba yang telah dijelaskan sebelumnya.

##### **4.4.1. Pembuatan Database dan Program Prototype**

Seperti yang telah dijelaskan pada bab sebelumnya, dalam penelitian ini peneliti akan membuat program prototype untuk melakukan *analisis object-based opinion mining*. Disini peneliti menggunakan bahasa pemrograman Python dan PHP untuk membangun program prototype ini, sedangkan untuk database peneliti menggunakan MySQL. Untuk mengeksekusi *script* Python peneliti menggunakan Python versi 3.4.2. Untuk dapat mengeksekusi *script* PHP maka kita akan membutuhkan apache server, disinilah peneliti menggunakan XAMPP versi 1.8.3, sedangkan untuk database server peneliti akan menggunakan MySQL yang terdapat pada XAMPP tersebut.

Untuk dapat melakukan rangkaian proses uji coba dan memudahkan dalam proses manajemen data, maka pertama-tama peneliti akan membuat database sebagai media penyimpanan data. Rancangan database dapat dilihat pada gambar 4.1, dimana rancangan disajikan dalam bentuk Physical Data Model (PDM).

Pembuatan program prototype dibagi menjadi dua bagian, yang pertama menggunakan bahasa pemrograman Python, dimana bagian ini bertugas untuk melakukan melakukna tahap praproses data. Program bagian kedua akan dibuat menggunakan bahasa pemrograman PHP, dimana bagian ini berfungsi untuk melakukan proses analisis *object-based opinion mining* sampai dengan perhitungan performanya.



Gambar 4.1 Rancangan Database

#### 4.4.2. Persiapan Data Uji Coba

Pada awal-awal sebelum uji coba dilakukan, peneliti perlu mempersiapkan data terlebih dahulu agar siap diolah. Data-data yang dibutuhkan adalah data review, data *opinion lexicon*, dan data filter. Data-data tersebut akan disimpan dalam database agar dapat siap dipakai untuk proses selanjutnya.

##### a. Data Review

Seperti yang telah dijelaskan pada bab sebelumnya bahwa peneliti akan menggunakan data corpus yang dihasilkan Marrese-Taylor (2014). Dalam data ini terdapat data review yang telah dibagi ke dalam bentuk kalimat beserta dengan objek yang telah diekstrak secara manual lengkap dengan orientasinya. Disini data yang diperoleh berupa file txt, dimana contoh

isinya dapat dilihat pada bab sebelumnya bagian 3.1. Untuk dapat diolah dengan mudah, maka peneliti memasukkannya ke dalam database dengan memisahkannya antara kalimat review dan objek nya ke dalam tabel terpisah pada database.

b. Opinion Lexicon

Data ini berisi daftar kata-kata positif dan negatif, dimana data ini diperoleh dari [www.cs.uic.edu](http://www.cs.uic.edu) dalam format txt. Untuk mempermudah penggunaannya dalam analisis maka peneliti memasukkannya ke dalam database. Dalam database data ini disimpan ke dalam tabel positive\_word dan negative\_word.

c. Data Filter

Data ini diperoleh dari beberapa sumber di internet. Yang menjadi tantangan adalah data filter tidak terdapat patokan mana yang lebih bagus dibanding dengan yang lain karena data ini akan disesuaikan dengan kebutuhan masing-masing sistem. Dalam penelitian ini peneliti melakukan penyusunan data filter dengan acuan data ontologi yang diperoleh dari internet dan disempurnakan dengan daftar objek yang ada dalam data corpus agar hasil analisis lebih maksimal. Data ontologi yang diperoleh dari internet berupa file owl dimana untuk dapat membacanya digunakanlah tool yang bernama Protégé. Data ontologi yang dianggap sesuai dengan kebutuhan sistem akan disimpan dalam tabel hotel\_filter dan restaurant\_filter pada database. Belum adanya konsep dan landasan ilmiah yang pasti dalam menyusun dan menentukan data filter yang sesuai, sehingga proses penyusunan data filter yang digunakan dalam penelitian ini dilakukan secara manual oleh peneliti, yaitu berdasarkan data ontologi yang didapat dari internet dan juga berdasarkan data objek yang ada pada data corpus. Oleh karena itu apabila ada peneliti lain mencoba melakukan penelitian serupa maka kemungkinan besar data filter yang dihasilkan akan berbeda. Hal ini termasuk merupakan kelemahan yang ada dalam penelitian ini dan diharapkan pada penelitian selanjutnya dapat diperbaiki dengan adanya landasan ilmiah yang pasti. Data filter yang digunakan pada penelitian ini secara rinci ditampilkan pada lampiran D.

#### **4.4.3. Praproses Data Uji Coba**

Setelah semua data yang dibutuhkan sudah tersimpan dengan baik dalam database, maka proses selanjutnya yang dilakukan adalah praproses data. Seperti dijelaskan sebelumnya, proses ini akan dilakukan dengan menggunakan bahasa pemrograman Python. Alasan kenapa tahap ini dipisahkan dari proses *object-based opinion mining* dengan bahasa pemrograman yang berbeda pula, dikarenakan dalam Python terdapat tool yang bernama NLTK. NLTK sendiri merupakan tool yang powerfull untuk mengolah data bahasa manusia (*natural language*), selain itu dikarenakan penelitian ini dilakukan untuk melakukan peningkatan hasil dari penelitian sebelumnya (Marrese-Taylor, 2014) dan pada penelitian tersebut juga menggunakan NLTK. Agar hasil perbandingan lebih sesuai maka digunakanlah NLTK dalam tahap praproses. Seperti dijelaskan pada bab sebelumnya bahwa pada tahapan ini dibagi menjadi lima tahapan yaitu Tokenizing, POS Tagging, Chunking, Stopping, dan Stemming. Script yang digunakan dalam proses ini dapat dilihat secara lengkap pada lampiran E.

Proses pertama adalah *tokenizing* dimana disini kalimat review akan dipecah ke dalam token-token. Contoh dari hasil proses *tokenizing* dapat dilihat pada tabel 4.3. Data review diambil dari tabel review dan review\_detail di database.

**Tabel 4.3 Contoh Hasil Tokenizing**

Data Review	Hasil Tokenizing
there was no hot water.	['there', 'was', 'no', 'hot', 'water', '.']
cannot fault the place although the tennis court is a little tired as is the equipment.	['can', 'not', 'fault', 'the', 'place', 'although', 'the', 'tennis', 'court', 'is', 'a', 'little', 'tired', 'as', 'is', 'the', 'equipment', '.']

Proses Selanjutnya adalah proses POS Tagging. Setelah data review berbentuk token kata, maka sekarang saatnya pemberian label kelas kata pada suatu kata dalam kalimat. Contoh dari hasil POS Tagging dapat dilihat pada tabel 4.4.

**Tabel 4.4 Contoh Hasil POS Tagging**

<b>Hasil Tokenizing</b>	<b>Hasil POS Tagging</b>
['there', 'was', 'no', 'hot', 'water', '.']	[('there', 'EX'), ('was', 'VBD'), ('no', 'DT'), ('hot', 'JJ'), ('water', 'NN'), ('.', '.')]
['can', 'not', 'fault', 'the', 'place', 'although', 'the', 'tennis', 'court', 'is', 'a', 'little', 'tired', 'as', 'is', 'the', 'equipment', '.']	[('can', 'MD'), ('not', 'RB'), ('fault', 'VB'), ('the', 'DT'), ('place', 'NN'), ('although', 'IN'), ('the', 'DT'), ('tennis', 'NN'), ('court', 'NN'), ('is', 'VBZ'), ('a', 'DT'), ('little', 'RB'), ('tired', 'JJ'), ('as', 'IN'), ('is', 'VBZ'), ('the', 'DT'), ('equipment', 'NN'), ('.', '.')]

Proses selanjutnya adalah chunking, dimana digunakan untuk menemukan kata benda dan frase kata benda atau *noun phrase* (NP). Contoh dari hasil POS Tagging dapat dilihat pada tabel 4.5. Proses ini mencari kata sifat yang diikuti dengan kata benda dan kata benda yang diikuti dengan kata benda untuk dijadikan sebagai kandidat objek dan disimpan pada tabel preprocessing\_np dalam database.

**Tabel 4.5 Contoh Hasil Chunking**

<b>Hasil POS Tagging</b>	<b>Hasil Chunking</b>
[('there', 'EX'), ('was', 'VBD'), ('no', 'DT'), ('hot', 'JJ'), ('water', 'NN'), ('.', '.')]	[('hot', 'JJ'), ('water', 'NN')]
[('can', 'MD'), ('not', 'RB'), ('fault', 'VB'), ('the', 'DT'), ('place', 'NN'), ('although', 'IN'), ('the', 'DT'), ('tennis', 'NN'), ('court', 'NN'), ('is', 'VBZ'), ('a', 'DT'), ('little', 'RB'), ('tired', 'JJ'), ('as', 'IN'), ('is', 'VBZ'), ('the', 'DT'), ('equipment', 'NN'), ('.', '.')]	[('tennis', 'NN'), ('court', 'NN')]

Tahap selanjutnya dalam praproses data adalah tahap *stopping*, dimana dari hasil proses POS Tagging akan dilakukan seleksi dan kata yang merupakan *stopword* akan dieliminasi. Contoh dari hasil *stopping* dapat dilihat pada tabel 4.6.

**Tabel 4.6 Contoh Hasil Stopping**

<b>Hasil POS Tagging</b>	<b>Hasil Stopping</b>
[('there', 'EX'), ('was', 'VBD'), ('no', 'DT'), ('hot', 'JJ'), ('water', 'NN'), ('.', '.')]	[('hot', 'JJ'), ('water', 'NN')]
[('can', 'MD'), ('not', 'RB'), ('fault', 'VB'), ('the', 'DT'), ('place', 'NN'), ('although', 'IN'), ('the', 'DT'), ('tennis', 'NN'), ('court', 'NN'), ('is', 'VBZ'), ('a', 'DT'), ('little', 'RB'), ('tired', 'JJ'), ('as', 'IN'), ('is', 'VBZ'), ('the', 'DT'), ('equipment', 'NN'), ('.', '.')]	[('fault', 'VB'), ('place', 'NN'), ('although', 'IN'), ('tennis', 'NN'), ('court', 'NN'), ('little', 'RB'), ('tired', 'JJ'), ('equipment', 'NN'), ('.', '.')]

Proses terakhir dalam rangkaian praproses data adalah *Stemming*, dimana pada kata-kata yang tersisa pada dokumen teks dilakukan penguraian untuk mendapatkan kata dasar. Metode *stemming* yang digunakan adalah Porter Stemmer yang sudah terdapat pada NLTK. Contoh dari hasil *stemming* dapat dilihat pada tabel 4.7. Setelah semua tahapan praproses selesai maka hasil tersebut akan disimpan pada tabel preprocessing\_data dalam database.

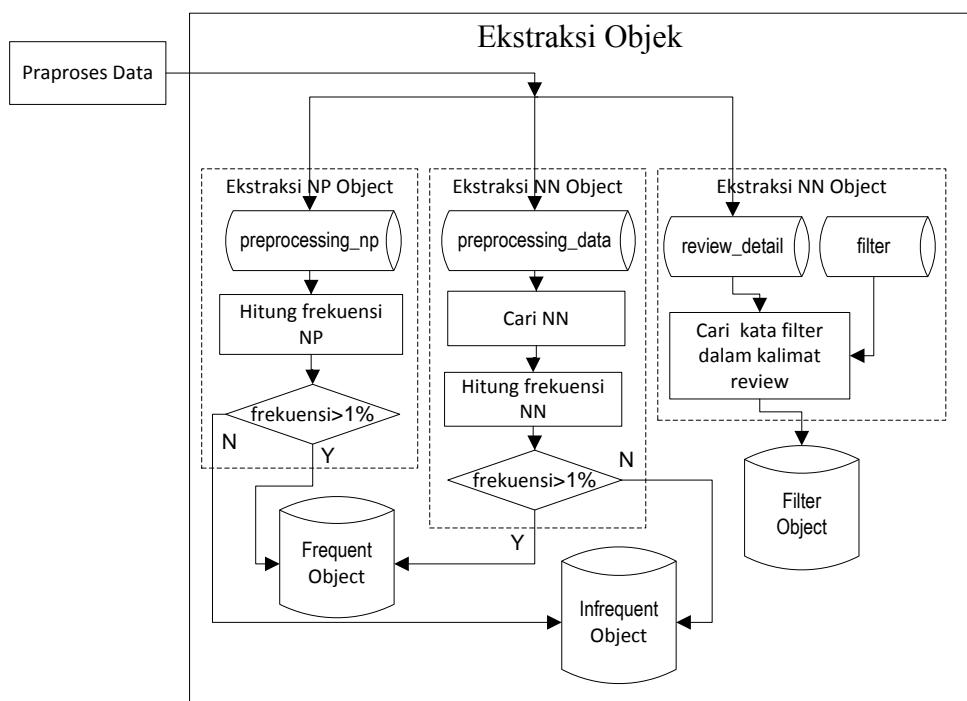
**Tabel 4.7 Contoh Hasil Stemming**

<b>Hasil Stopping</b>	<b>Hasil Stemming</b>
[('hot', 'JJ'), ('water', 'NN')]	[('hot', 'JJ'), ('water', 'NN')]
[('fault', 'VB'), ('place', 'NN'), ('although', 'IN'), ('tennis', 'NN'), ('court', 'NN'), ('little', 'RB'), ('tired', 'JJ'), ('equipment', 'NN'), ('.', '.')]	[('fault', 'VB'), ('place', 'NN'), ('although', 'IN'), ('tenni', 'NN'), ('court', 'NN'), ('littl', 'RB'), ('tire', 'JJ'), ('equip', 'NN'), ('.', '.')]

#### **4.4.4. Proses Uji Coba Ekstraksi Objek**

Mulai dari proses ini dan proses selanjutnya akan dilakukan menggunakan prototype program yang ditulis dalam bahasa pemrograman PHP. Pada proses uji coba tahapan ekstraksi objek ini, pertama-tama yang dilakukan adalah sistem akan melakukan scanning pada tabel preprocessing\_np untuk mencari frase kata benda

mana yang memenuhi batas minimal jumlah kemunculan untuk dianggap sebagai *frequent object*. Frase kata benda yang tidak memenuhi syarat akan dianggap sebagai *infrequent object*. Ketentuan minimal frequent object adalah jika kemunculannya lebih dari atau sama dengan 1% dari jumlah kalimat review yang dianalisis. Proses tersebut juga dilakukan pada data tabel preprocessing\_data karena selain frase kata benda, kandidat objek juga bisa berupa kata benda yang ada pada tabel preprocessing. Setelah proses tersebut selesai, maka hasilnya akan disimpan dalam tabel frequent\_object dan infrequent\_object. Alur proses ekstraksi objek dapat dilihat pada gambar 4.2.



Gambar 4.2 Proses Ekstraksi Objek

Sesuai dengan skenario yang ada dan merupakan salah satu inti dari penelitian ini, untuk dapat meningkatkan performa dari *object-based opinion mining* maka kandidat objek yang telah tersimpan dalam tabel frequent\_object dan infrequent\_object akan dilakukan filterisasi dengan cara mencocokkan dengan data filter yang telah disimpan dalam tabel hotel\_filter dan restaurant\_filter. Kandidat objek akan ditandai dalam database objek mana saja yang termasuk dalam data filter dan mana yang tidak. Objek yang termasuk dalam data filter akan dijadikan objek dalam skenario analisis *object-based opinion mining*. Proses filterisasi objek

menggunakan data filter akan dilakukan pada saat setelah orientasi masing-masing objek diketahui dan hasilnya akan disimpan dalam tabel `opinion_mining_result`. Hasil filterisasi hanya akan disimpan berupa keterangan saja dalam database apakah suatu objek tersebut termasuk dalam data filter atau tidak dan akan disimpan ke dalam hasil *object-based opinion mining* yang biasa.

Sesuai dengan skenario yang telah dijelaskan sebelumnya, yaitu terdapat skenario ekstraksi objek yang menggunakan data filter secara langsung untuk mengekstrak objek. Oleh sebab itu sistem akan melakukan *scanning* secara langsung terhadap setiap kalimat review, apakah dalam kalimat review terdapat kata yang termasuk dalam data filter. Proses ini memakan waktu yang cukup lama dikarenakan tiap kata pada kalimat review dilakukan pencocokan terhadap semua data filter. Semua kata yang sesuai akan dianggap sebagai kandidat objek dan akan disimpan pada tabel `filter_object` dalam database. Script yang digunakan dalam proses ini dapat dilihat secara lengkap pada lampiran E.

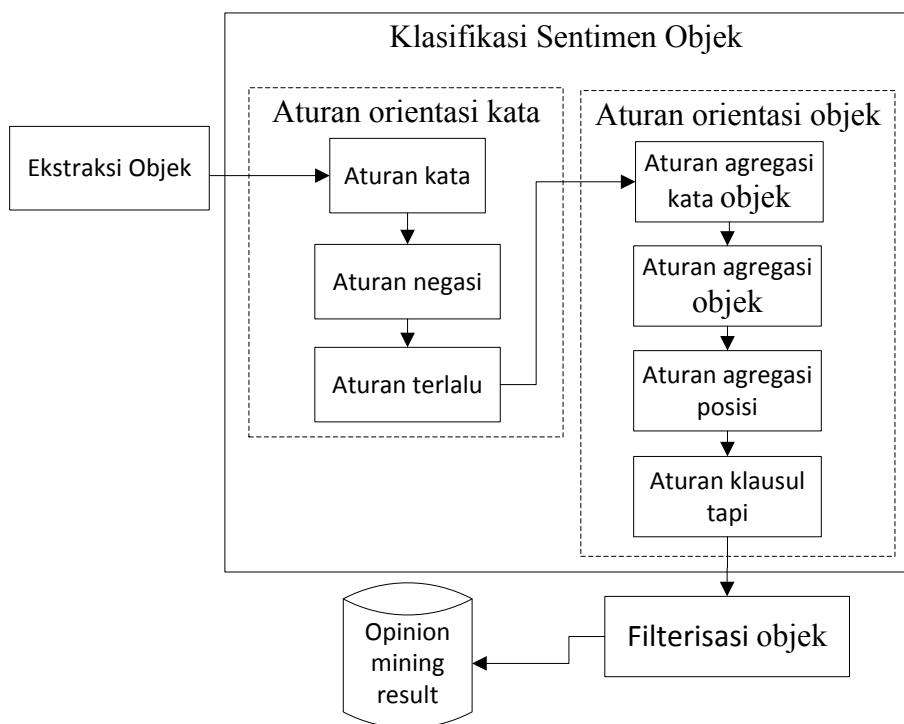
Setelah proses ekstraksi objek selesai maka untuk melakukan perhitungan performa dari ekstraksi objek akan dilakukan setelah proses klasifikasi sentimen objek selesai. Hal ini dilakukan karena kandidat objek yang telah didapatkan harus mengandung orientasi sentimen untuk dianggap sebagai objek dari suatu review dan kandidat objek yang tidak terdapat orientasi akan dieliminasi nantinya. Hasil dari perhitungan tersebut akan disimpan dalam tabel `summary` yang ada pada database. Perhitungan performa meliputi perhitungan nilai *accuracy*, *precision*, *recall*, dan *f-measure*.

#### 4.4.5. Proses Uji Coba Klasifikasi Sentimen Objek

Setelah proses ekstraksi objek selesai dilakukan, maka proses selanjutnya adalah melakukan proses klasifikasi sentimen objek. Pada proses ini objek-objek yang berhasil diekstrak nantinya akan diklasifikasikan dan dianalisis orientasinya, apakah positif atau negatif. Objek yang dianggap netral akan dieliminasi. Metode dan algoritma yang digunakan pada proses ini sama dengan apa yang digunakan oleh Marrese-Taylor pada tahun 2014.

Proses ini dibagi menjadi dua tahapan, dimana yang pertama adalah menerapkan aturan orientasi kata dan yang kedua baru akan menerapkan arutan

orientasi objek. Dimana secara garis besar aturan orientasi kata berfungsi untuk memberikan label orientasi sentimen ke semua kata yang lulus tahap praproses data. Kata-kata tersebut akan dicocokkan dengan orientasi yang ada pada daftar kata opini dalam tabel `positive_word` dan `negative_word`. Setelah itu hasilnya akan diproses dengan aturan negasi dan aturan terlalu untuk mendapatkan hasil yang lebih akurat. Setelah proses tersebut selesai, baru kita akan menganalisa orientasi dari semua objek yang berhasil diekstrak dengan acuan hasil dari aturan orientasi kata. Setelah orientasi sentimen masing-masing objek diketahui maka akan dilakukan proses filterisasi objek menggunakan data filter, baru setelah itu hasilnya akan disimpan dalam tabel `opinion_mining_result`. Proses filterisasi dilakukan saat terakhir karena peneliti menganggap saat inilah yang paling efisien dalam melakukan filterisasi dimana hasil *object-based opinion mining* sudah terlihat dan hasil filterisasi hanya berupa keterangan saja apakah suatu objek termasuk dalam filter atau tidak. Untuk penjelasan secara detail dari masing-masing aturan dapat dilihat pada laporan bagian 2.7, sedangkan secara garis besar tahapan proses klasifikasi sentimen objek dapat dilihat pada gambar 4.3. Script yang digunakan dalam proses ini dapat dilihat pada lampiran E.



**Gambar 4.3 Proses Klasifikasi Sentimen Objek**

Hasil dari proses klasifikasi sentimen objek dapat dilihat pada lampiran C. Dimana setelah didapatkan hasil dari klasifikasi sentimen objek, maka akan dilakukan perhitungan performa dari *object-based opinion mining* dibanding dengan data corpus yang merupakan hasil dari ekstraksi secara manual. Proses perhitungan performa dibagi menjadi dua tahap, dimana tahap pertama hanya objek yang benar dan sesuai dengan objek yang ada pada data corpus saja yang akan dianggap dan masuk perhitungan. Hal ini dimaksudkan untuk lebih mengukur tingkat keakuratan dari rule yang digunakan dalam hal menentukan orientasi suatu objek. Sedangkan perhitungan performa tahap kedua, semua objek yang dihasilkan oleh *object-based opinion mining* akan hitung. Hal ini dilakukan untuk menganalisa performa secara keseluruhan dari masing-masing pendekatan *object-based opinion mining* dalam hal menentukan orientasi sentimen objek. Hasil dari perhitungan tersebut akan disimpan dalam tabel summary yang ada pada database. Perhitungan performa meliputi perhitungan nilai *accuracy*, *precision*, *recall*, dan *f-measure*.

#### **4.4.6. Proses Uji Coba Deteksi Kalimat Opini**

Proses uji coba ini dilakukan setelah proses klasifikasi sentimen objek selesai dilakukan, dimana uji coba ini dilakukan untuk mengetahui performa dari hasil *object-based opinion mining* dalam hal pendekripsi kalimat opini dibandingkan dengan data corpus. Kalimat akan dianggap sebagai kalimat opini jika terdapat objek lengkap dengan orientasinya yang berhasil diekstrak dalam *object-based opinion mining*. Berdasarkan hal itu maka akan dilakukan perhitungan dengan acuan objek yang ada pada data corpus. Hasil dari perhitungan tersebut akan disimpan dalam tabel summary yang ada pada database. Perhitungan performa meliputi perhitungan nilai *accuracy*, *precision*, *recall*, dan *f-measure*. Script yang digunakan dalam proses ini dapat dilihat pada lampiran E.

#### **4.4.7. Proses Uji Coba Klasifikasi Sentimen Kalimat Review**

Proses ini dilakukan dengan menggunakan tool WEKA, dimana data yang digunakan adalah data hasil dari *object-based opinion mining*. Terdapat tujuh set data untuk masing-masing jenis produk yang akan dianalisis. Data tersebut adalah

kalimat review sebagai atribut klasifikasi, *frequent object* sebagai atribut klasifikasi, *frequent object* dan *infrequent object* sebagai atribut klasifikasi, *frequent object* yang telah terfilter data filter sebagai atribut klasifikasi, *frequent object* dan *infrequent object* yang telah terfilter data filter sebagai atribut klasifikasi, dan yang terakhir *filter object* sebagai atribut klasifikasi. Tujuan dari proses klasifikasi ini sendiri adalah untuk membuktikan apakah hasil output *object-based opinion mining* dapat digunakan untuk memprediksi orientasi kalimat review secara global. Kalimat akan diklasifikasikan ke dalam tiga kelas, yaitu positif, netral, dan negatif.

Untuk dapat melakukan proses klasifikasi menggunakan *machine learning* yang ada di WEKA maka harus dilakukan ekspor data dari database terlebih dahulu, mengingat format data yang bisa dieksekusi oleh WEKA adalah file berformat csv. Dalam proses ekspor data objek dan orientasi yang diekstrak akan ditransformasi ke dalam bentuk string tanpa spasi agar dapat diolah oleh WEKA dengan lebih mudah. Data review yang tidak terdapat objek akan ditandai dengan memberikan nilai statement sebagai atribut klasifikasi. Contoh perubahan format dapat dilihat pada tabel 4.8. Setelah data diekspor maka dilakukan identifikasi orientasi sentimen dari masing-masing kalimat review secara manual untuk dapat diolah di WEKA. Setelah itu data dibagi sesuai dengan tiap-tiap atribut klasifikasi yang akan dilakukan. Script eksport data dapat dilihat pada lampiran E.

**Tabel 4.8 Contoh perubahan format objek**

Objek Hasil Ekspor Data	Objek Hasil Transformasi
hot water[-]	waternegative
driveway[+], entrance[+], staff[+]	drivewaypositive, entrancepositive, staffpositive

Metode yang digunakan untuk proses klasifikasi sentimen kalimat review pada penelitian ini adalah Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Pada saat data pertama kali diimport dalam WEKA, sebelum dilakukan proses klasifikasi tiap atribut yang akan diolah terlebih dahulu difilter menggunakan StringToWordVector untuk dilakukan *preprocessing*. Untuk pengukuran performa klasifikasi pada WEKA dilakukan menggunakan *k-fold*

*cross-validation* yakni menggunakan 10 *folds*. Hasil output dari proses klasifikasi ini dapat dilihat secara lengkap pada lampiran B.

## 4.5. Hasil Uji Coba

Pada bagian ini akan dijelaskan mengenai detail hasil uji coba yang dilakukan sesuai dengan masing-masing skenario uji coba yang telah dijelaskan sebelumnya.

### 4.5.1. Uji Coba Ekstraksi Objek

Pada uji coba ini dilakukan berbagai jenis pengujian dengan berbagai pendekatan yang berbeda dalam proses ekstraksi objek. Hasil uji ekstraksi objek data hotel dapat dilihat pada tabel 4.9, sedangkan hasil uji ekstraksi objek data restoran dapat dilihat pada tabel 4.10.

**Tabel 4.9 Hasil Uji Ekstraksi Objek Hotel**

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	36.80%	45.70%	44.35%	45.01%
Frequent and Infrequent Object	31.82%	22.33%	59.97%	32.54%
Frequent Object with filter	38.74%	64.49%	43.52%	51.97%
Frequent and Infrequent Object with Filter	46.15%	63.02%	55.95%	59.27%
Filter Object	46.76%	62.36%	57.39%	59.77%

**Tabel 4.10 Hasil Uji Ekstraksi Objek Restoran**

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	35.86%	44.82%	45.40%	45.11%
Frequent and Infrequent Object	31.37%	21.60%	61.06%	31.91%
Frequent Object with filter	39.91%	64.61%	45.11%	53.13%
Frequent and Infrequent Object with Filter	48.93%	65.40%	59.20%	62.15%
Filter Object	49.16%	63.76%	60.92%	62.31%

Dan berikut ini merupakan waktu yang dihabiskan pada masing-masing pengujian untuk melakukan proses ekstraksi objek. Untuk waktu pengujian dapat dilihat pada tabel 4.11.

**Tabel 4.11 Waktu Pengujian Ekstraksi Objek**

Jenis Ekstraksi Objek	Waktu Ekstraksi Objek Hotel	Waktu Ekstraksi Objek Restoran
Frequent Object	0 detik	1 detik
Frequent and Infrequent Object	4 detik	3 detik
Frequent Object with filter	0 detik	1 detik
Frequent and Infrequent Object with Filter	4 detik	3 detik
Filter Object	3051 detik	1559 detik

#### 4.5.2. Uji Coba Klasifikasi Sentimen Objek

Pada uji coba ini dilakukan berbagai jenis pengujian dengan berbagai pendekatan yang berbeda dalam proses klasifikasi sentimen objek. Hasil uji klasifikasi sentimen objek hotel dengan hanya menggunakan objek yang ada pada data corpus dapat dilihat pada tabel 4.12, sedangkan hasil uji klasifikasi sentimen objek restoran dengan hanya menggunakan objek yang ada pada data corpus dapat dilihat pada tabel 4.13.

**Tabel 4.12 Hasil Uji Klasifikasi Sentimen Objek Hotel (Objek yang Benar)**

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	88.03%	90.81%	94.36%	92.55%
Frequent and Infrequent Object	87.61%	90.88%	93.00%	91.93%
Frequent Object with filter	88.15%	90.91%	94.51%	92.68%
Frequent and Infrequent Object with Filter	87.67%	91.33%	92.78%	92.05%
Filter Object	85.47%	90.02%	91.25%	90.63%

**Tabel 4.13 Hasil Uji Klasifikasi Sentimen Objek Restoran (Objek yang Benar)**

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	91.14%	95.27%	94.58%	94.92%
Frequent and Infrequent Object	88.71%	93.61%	93.09%	93.35%
Frequent Object with filter	91.08%	95.24%	94.55%	94.89%
Frequent and Infrequent Object with Filter	88.59%	93.41%	93.14%	93.27%
Filter Object	87.03%	91.94%	92.72%	92.33%

Hasil uji klasifikasi sentimen objek hotel dengan menggunakan semua objek yang berhasil terekstrak dapat dilihat pada tabel 4.14, sedangkan hasil uji klasifikasi sentimen objek restoran dengan menggunakan semua objek yang berhasil terekstrak dapat dilihat pada tabel 4.15.

**Tabel 4.14 Hasil Uji Klasifikasi Sentimen Objek Hotel (Semua Objek)**

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	40.23%	44.94%	65.22%	53.21%
Frequent and Infrequent Object	19.57%	21.59%	40.43%	28.15%
Frequent Object with filter	56.85%	60.91%	79.93%	69.14%
Frequent and Infrequent Object with Filter	55.25%	59.41%	76.27%	66.79%
Filter Object	53.29%	58.46%	73.79%	65.24%

**Tabel 4.15 Hasil Uji Klasifikasi Sentimen Objek Restoran (Semua Objek)**

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	40.85%	44.71%	73.80%	55.68%
Frequent and Infrequent Object	19.16%	21.06%	50.68%	29.76%
Frequent Object with filter	58.85%	63.26%	84.14%	72.22%
Frequent and Infrequent Object with Filter	57.94%	61.51%	84.24%	71.10%
Filter Object	55.49%	59.32%	82.75%	69.10%

Dan berikut ini merupakan waktu yang dihabiskan pada masing-masing pengujian untuk melakukan proses klasifikasi sentimen objek. Untuk waktu pengujian dapat dilihat pada tabel 4.16.

**Tabel 4.16 Waktu Pengujian Klasifikasi Sentimen Objek**

<b>Jenis Ekstraksi Objek</b>	<b>Waktu Klasifikasi Sentimen Objek Hotel</b>	<b>Waktu Klasifikasi Sentimen Objek Restoran</b>
Frequent Object	114 detik	75 detik
Frequent and Infrequent Object	202 detik	141 detik
Frequent Object with filter	114 detik	75 detik
Frequent and Infrequent Object with Filter	202 detik	141 detik
Filter Object	2700 detik	1291 detik

#### 4.5.3. Uji Coba Deteksi Kalimat Opini

Pada uji coba ini dilakukan berbagai jenis pengujian dengan berbagai pendekatan yang berbeda dalam proses deteksi kalimat opini. Hasil uji deteksi kalimat opini hotel dapat dilihat pada tabel 4.17, sedangkan hasil uji deteksi kalimat opini restoran dapat dilihat pada tabel 4.18.

**Tabel 4.17 Hasil Uji Deteksi Kalimat Opini Hotel**

<b>Jenis Ekstraksi Objek</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F-measure</b>
Frequent Object	73.38%	88.15%	75.70%	81.45%
Frequent and Infrequent Object	76.30%	85.76%	83.09%	84.40%
Frequent Object with filter	72.50%	91.35%	71.10%	79.96%
Frequent and Infrequent Object with Filter	74.90%	89.90%	76.03%	82.39%
Filter Object	75.16%	89.94%	76.35%	82.59%

**Tabel 4.18 Hasil Uji Deteksi Kalimat Opini Restoran**

<b>Jenis Ekstraksi Objek</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F-measure</b>
Frequent Object	71.70%	90.41%	71.54%	79.88%
Frequent and Infrequent Object	78.94%	89.71%	82.66%	86.04%
Frequent Object with Filter	69.57%	93.13%	66.12%	77.33%
Frequent and Infrequent Object with Filter	75.53%	93.20%	74.25%	82.65%
Filter Object	76.60%	93.31%	75.61%	83.53%

Dan berikut ini merupakan waktu yang dihabiskan pada masing-masing pengujian untuk melakukan proses ekstraksi kalimat opini. Untuk waktu pengujian dapat dilihat pada tabel 4.19.

**Tabel 4.19 Waktu Pengujian Deteksi Kalimat Opini**

Jenis Ekstraksi Objek	Waktu Ekstraksi Objek Hotel	Waktu Ekstraksi Objek Restoran
Frequent Object	22 detik	14 detik
Frequent and Infrequent Object	21 detik	14 detik
Frequent Object with Filter	21 detik	14 detik
Frequent and Infrequent Object with Filter	21 detik	14 detik
Filter Object	35 detik	21 detik

#### **4.5.4. Uji Coba Klasifikasi Sentimen Kalimat Review**

Pada uji coba ini dilakukan perbandingan terhadap hasil klasifikasi dari tiga teknik berbeda yang digunakan untuk memprediksi sentimen kalimat *review* berdasarkan data objek yang telah berhasil diekstrak dalam *object-based opinion mining*. Teknik klasifikasi yang digunakan adalah Naïve Bayes (NB), Support Vector Machine (SVM), dan Random Forest dengan menggunakan data atribut sesuai dengan pendekatan yang telah dilakukan pada proses *object-based opinion mining*. Seluruh proses klasifikasi dilakukan menggunakan tool WEKA.

Pada uji coba ini peneliti menggunakan tujuh set data yang berbeda yang akan digunakan sebagai atribut klasifikasi untuk masing-masing jenis produk review. Data tersebut adalah kalimat review asli, aspek data corpus, *frequent object*, gabungan *frequent* dan *infrequent object*, *frequent object* yang telah terfilter oleh data filter, gabungan *frequent* dan *infrequent object* yang telah terfilter oleh data filter, dan objek filter.

**Tabel 4.20 Hasil Uji Klasifikasi dengan Atribut Data Kalimat Review**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	56.40%	31.80%	56.40%	40.70%	64.26%	41.30%	64.3%	50.30%
SVM	56.40%	31.80%	56.40%	40.70%	64.26%	41.30%	64.3%	50.30%
RF	56.40%	31.80%	56.40%	40.70%	64.26%	41.30%	64.30%	50.30%

**Tabel 4.21 Hasil Uji Klasifikasi dengan Atribut Data Object Data Corpus**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	86.44%	89.10%	86.40%	83.90%	88.51%	90.30%	88.50%	85.20%
SVM	86.44%	89.10%	86.40%	83.90%	89.15%	90.10%	89.10%	86.50%
RF	86.44%	89.10%	86.40%	83.90%	89.15%	90.10%	89.10%	86.50%

**Tabel 4.22 Hasil Uji Klasifikasi dengan Atribut Data Frequent Object**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	56.91%	50.00%	56.90%	43.70%	63.83%	49.90%	63.80%	50.80%
SVM	60.20%	61.80%	60.20%	55.80%	62.77%	49.80%	62.80%	50.40%
RF	60.08%	61.70%	60.10%	55.70%	61.06%	48.90%	61.10%	51.70%

**Tabel 4.23 Hasil Uji Klasifikasi dengan Atribut Data Frequent object dan Infrequent Object**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	56.27%	31.80%	56.30%	40.60%	64.26%	41.30%	64.30%	50.30%
SVM	60.33%	47.40%	60.30%	53.00%	67.45%	58.70%	67.40%	62.70%
RF	60.33%	47.40%	60.30%	53.00%	67.45%	58.70%	67.40%	62.70%

**Tabel 4.24 Hasil Uji Klasifikasi dengan Atribut Data Frequent Object yang Terfilter**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	58.05%	46.80%	58.00%	45.70%	63.61%	47.40%	63.60%	50.70%
SVM	62.61%	67.00%	62.60%	60.60%	62.98%	51.00%	63.00%	51.90%
RF	62.36%	67.10%	62.40%	60.10%	63.19%	53.40%	63.20%	52.60%

**Tabel 4.25 Hasil Uji Klasifikasi dengan Atribut Data Frequent Object dan Infrequent Object yang Terfilter**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	55.89%	41.00%	55.90%	42.00%	62.98%	41.00%	63.00%	49.70%
SVM	61.09%	60.50%	61.10%	56.70%	63.40%	63.20%	63.40%	60.30%
RF	61.34%	62.20%	61.30%	57.10%	64.68%	65.50%	64.70%	61.70%

**Tabel 4.26 Hasil Uji Klasifikasi dengan Atribut Data Filter Object**

Classifier	Hotel				Restaurant			
	Acc	Prec	Rec	F-meas	Acc	Prec	Rec	F-meas
Naive Bayes	56.27%	42.40%	56.30%	41.90%	62.98%	41.00%	63.00%	49.70%
SVM	61.97%	63.10%	62.00%	57.40%	64.04%	63.40%	64.00%	60.80%
RF	61.85%	63.50%	61.90%	57.10%	65.32%	65.60%	65.30%	62.10%

## 4.6. Analisis Hasil Uji Coba

Tahap uji coba telah dilaksanakan dan menghasilkan berbagai nilai. Analisis dari hasil uji coba dipaparkan pada masing-masing bagian berikut ini sesuai dengan jenis uji coba.

### 4.6.1. Analisis Eksatraksi Objek

Dari hasil uji coba ini dapat diketahui bahwa pendekatan ekstraksi objek yang telah dilakukan tidak ada yang mendapatkan hasil dominan secara keseluruhan, akan tetapi penerapan filter dalam proses ekstraksi objek jelas dapat

meningkatkan presisi dengan cukup meyakinkan yaitu pada pendekatan *frequent object* yang semula 45.7% menjadi 64.49% pada data *review* hotel dan dari 44.82% menjadi 64.61% pada data *review* restoran, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 44.35% menjadi 43.52% pada data *review* hotel dan dari 45.4% menjadi 45.11% pada data *review* restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan *frequent object* dan *infrequent object* dimana pada kedua data *review* terjadi peningkatan hasil presisi yang cukup meyakinkan yaitu yang semula 22.33% menjadi 63.02% pada data *review* hotel dan dari 21.6% menjadi 65.4% pada data *review* restoran, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 59.97% menjadi 55.95% pada data *review* hotel dan dari 61.06% menjadi 59.2% pada data *review* restoran. Waktu yang dibutuhkan untuk keempat pendekatan tersebut tidak lebih dari 4 detik.

Hasil yang cukup menarik didapatkan dari hasil ekstraksi objek dengan menggunakan pendekatan data filter yang secara langsung digunakan untuk mengekstrak objek yang ada pada data *review*, dimana pendekatan ini mendapatkan hasil akurasi dan f-measure tertinggi baik pada kedua data *review* yaitu mendapatkan 46.76% akurasi dan 59.77% f-measure pada data *review* hotel dan mendapatkan 49.16% akurasi dan 62.31% f-measure pada data *review* restoran. Akan tetapi hasil tersebut dapat diperoleh dengan menghabiskan waktu yang jauh lebih lama dibandingkan dengan pendekatan yang lain yaitu 3051 detik pada data hotel dan 1559 detik pada data restoran, padahal pendekatan lain hanya memakan waktu tidak lebih dari 4 detik untuk melakukan ekstraksi objek. Hal ini terjadi dikarenakan pada pendekatan data filter secara langsung setiap kalimat yang sudah dipecah ke dalam token-token akan dilakukan scanning dan dicocokkan pada semua data filter yang ada, dimana tentu saja hal itu akan membutuhkan waktu yang cukup lama.

#### **4.6.2. Analisis Klasifikasi Sentimen Objek**

Dari hasil uji coba ini dapat diketahui bahwa untuk tahap uji coba yang pertama yaitu pengujian dengan menggunakan objek yang sesuai dengan objek yang ada pada data corpus saja menghasilkan hasil yang agak berbeda pada data

*review* hotel dan *review* restoran dimana pada data *review* hotel hasil dari penerapan filter pada *frequent object* mendapatkan hasil yang dominan dibanding dengan pendekatan yang lain walaupun hasil yang didapatkan tidak signifikan yaitu 88.15% akurasi, 90.91% presisi, 94.51% recall, dan 92.68% f-measure. Sedangkan pada data *review* restoran hasil dari penerapan filter pada *frequent object* mendapatkan hasil berbeda, dimana penerapan filter mengakibatkan turunnya sedikit hasil analisis yaitu 91.14% akurasi, 95.27% presisi, 94.58% recall, dan 94.92% f-measure tanpa filter dan mendapatkan hasil 91.08% akurasi, 95.24% presisi, 94.55% recall, dan 94.89% f-measure dengan filter. Sedikit penurunan pada tahap ini sebenarnya wajar saja mengingat objek yang dihitung hanya objek yang benar dan sesuai data corpus saja, sehingga peran dari filter belum kelihatan secara penuh. Oleh sebab itu peneliti juga melakukan uji coba tahap kedua dimana semua objek yang berhasil diekstrak dalam *object-based opinion mining* akan diperhitungkan sehingga peran filter akan benar-benar kelihatan.

Pada tahap uji coba kedua, pengujian dilakukan dengan menggunakan seluruh objek yang berhasil diekstrak dalam *object-based opinion mining*. Berbeda dengan hasil pada tahap sebelumnya yang hanya menggunakan objek yang sesuai dengan corpus saja, disini penggunaan filter dalam *object-based opinion mining* benar-benar terlihat penannya dengan mendapatkan hasil yang dominan secara keseluruhan dibanding dengan tanpa filter. Hasil tertinggi didapatkan pada pendekatan *frequent object* dengan filter, yaitu 56.85% akurasi, 60.91% presisi, 79.93% recall, 69.14% f-measure pada data *review* hotel, serta mendapatkan hasil 58.85% akurasi, 63.26% presisi, 84.14% recall, 72.22% f-measure pada data *review* restoran. Sedangkan untuk pendekatan data filter secara langsung, hasil yang diperoleh tidak lebih baik dari pendekatan lain yang menggunakan filter dan ditambah lagi waktu yang dibutuhkan juga jauh lebih lama yaitu 2700 detik pada data *review* hotel dan 1291 detik pada data *review* restoran. Padahal pendekatan lain hanya memakan waktu tidak lebih dari 202 detik.

#### **4.6.3. Analisis Deteksi Kalimat Opini**

Dari hasil uji coba ini dapat diketahui bahwa penerapan filter dalam proses deteksi kalimat opini hanya dapat meningkatkan sedikit saja presisi pada

pendekatan *frequent object* yang semula 88.15% menjadi 91.35% pada data *review* hotel dan dari 90.41% menjadi 93.13% pada data *review* restoran, dengan mengorbankan sedikit penurunan juga pada hasil recall yaitu yang semula 75.7% menjadi 71.1% pada data *review* hotel dan dari 71.54% menjadi 66.12% pada data *review* restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan *frequent object* dan *infrequent object* dimana pada kedua data *review* terjadi peningkatan sedikit saja presisi yaitu yang semula 85.76% menjadi 89.9% pada data *review* hotel dan dari 89.71% menjadi 93.2% pada data *review* restoran, dengan mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 83.09% menjadi 76.03% pada data *review* hotel dan dari 82.66% menjadi 74.25% pada data *review* restoran. Waktu yang dibutuhkan untuk keempat pendekatan tersebut tidak lebih dari 22 detik.

Untuk hasil proses deteksi kalimat opini dengan menggunakan pendekatan data filter yang secara langsung mendapatkan hasil cukup bagus yaitu 75.38% akurasi, 89.94% presisi, 76.35% recall, dan 82.59% f-measure pada data *review* hotel. Sedangkan untuk data *review* restoran mendapatkan hasil 76.6% akurasi, 93.31% presisi, 75.61% recall, dan 83.53% f-measure. Hasil tersebut mendapatkan hasil f-measure lebih baik jika dibandingkan dengan pendekatan lain yang menggunakan data filter. Seperti tahap analisis sebelumnya pendekatan ini memakan waktu lebih lama dibandingkan pendekatan lain yaitu 35 detik pada data hotel dan 21 detik pada data restoran.

#### **4.6.4. Analisis Klasifikasi Sentimen Kalimat Review**

Dari hasil uji coba ini dapat diketahui bahwa penerapan kalimat *review* secara langsung ke dalam proses klasifikasi sentimen menghasilkan hasil yang rendah yaitu 56.40% akurasi, 31.8% presisi, 56.4% recall, dan 40.7% f-measure pada data *review* hotel dan 64.26% akurasi, 41.3% presisi, 64.3% recall, 50.3% f-measure pada data *review* restoran. Hasil yang didapatkan semua sama persis pada ketiga teknik klasifikasi yang digunakan, hal ini terjadi karena semua kalimat hanya diklasifikasikan ke dalam kelas positif saja.

Hasil yang cukup bagus didapatkan pada uji coba dengan melakukan klasifikasi menggunakan objek yang ada pada data corpus, dimana hasil yang

diperoleh cukup bagus pada semua klasifikasi. Pada data *review* hotel ketiga *classifier* mendapatkan hasil sama persis yaitu 86.44% akurasi, 89.1% presisi, 86.4% recall, 83.9% f-measure. Hasil berbeda ditunjukkan pada klasifikasi data *review* restoran dimana SVM dan Random Forest mendapatkan akurasi lebih baik dibanding dengan Naïve Bayes yaitu 89.15% dibandingkan dengan 88.51%. Kesimpulan yang dapat ditarik dari hasil ini adalah bahwa penggunaan objek yang terekstrak dari kalimat *review* dapat digunakan untuk proses klasifikasi dalam menentukan orientasi sentimen kalimat *review*. Berhubung data aspek pada corpus adalah data uji yang digunakan dalam menentukan performa *object-based opinion mining*, maka dalam proses klasifikasi ini kita juga perlu menguji hasil dari masing-masing pendekatan yang telah dilakukan pada *object-based opinion mining* untuk mengetahui apakah hasil tersebut dapat menghasilkan hasil yang bagus jika dijadikan atribut klasifikasi penentuan orientasi setimen kalimat.

Pada uji coba klasifikasi dengan menggunakan objek yang ada pada hasil *object-based opinion mining* sebagai atribut menunjukkan bahwa hasil yang diperoleh tidaklah lebih bagus dibanding dengan hasil klasifikasi yang menggunakan aspek corpus sebagai atribut. Akan tetapi bila dibandingkan dengan hasil klasifikasi yang hanya menggunakan kalimat *review* saja sebagai atribut, maka bisa dibilang hasil klasifikasi dari objek hasil *object-based opinion mining* mendapatkan hasil yang lebih bagus.

Secara keseluruhan tidak ada teknik klasifikasi yang mendapatkan hasil yang dominan lebih baik dibanding dengan yang lain pada semua hasil pendekatan *object-based opinion mining*. Akan tetapi dapat disimpulkan bahwa klasifikasi yang lebih banyak mendapatkan hasil yang lebih baik adalah teknik klasifikasi *Random Forest*, dimana pada hasil klasifikasi dengan atribut data *frequent object* dan *infrequent object* dengan filter, Random Forest mendominasi seluruh hasil dibanding dengan Naïve Bayes dan SVM. Pada data *review* hotel Random Forest mendapatkan hasil 61.34% akurasi, 62.2% presisi, 61.3% recall, 57.1% f-measure dan pada data *review* restoran mendapatkan hasil 64.68% akurasi, 65.5% presisi, 64.7% recall, 61.7% f-measure.

Dari uji coba secara keseluruhan dari proses klasifikasi sentimen kalimat dapat ditarik kesimpulan bahwa penggunaan objek terekstraksi sebagai atribut

klasifikasi untuk menentukan orientasi kalimat mendapatkan hasil yang lebih baik dibanding dengan hanya menggunakan kalimat *review* secara langsung sebagai atribut klasifikasi. Walaupun hasil objek dari *object-based opinion mining* masih kurang dibandingkan dengan penggunaan aspek corpus sebagai atribut, akan tetapi hasil pemanfaatan *object-based opinion mining* dalam klasifikasi sentimen kalimat menghasilkan lebih baik dibanding menggunakan kalimat *review*. Selain itu yang paling penting adalah hasil dari proses klasifikasi menggunakan objek dari *object-based opinion mining* dengan filter mendapatkan hasil yang lebih baik dibanding dari tanpa menggunakan data filter sama sekali.

## LAMPIRAN A

Lampiran ini berisi mengenai hasil pengukuran performa *object-based opinion mining* secara lengkap. Nilai hasil pengukuran yang tertera berikut ini antara lain adalah nilai *accuracy*, *precision*, *recall*, *f-measure*, dan *confusion matrix*.

### A.1. Hasil Ekstraksi Objek Review Hotel (*Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	36.80%
Precision	45.70%
Recall	44.35%
F-Measure	45.01%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	585	734
	Negative	695	247

### A.2. Hasil Ekstraksi Objek Review Hotel (*Frequent Object* dan *Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	31.82%
Precision	22.33%
Recall	59.97%
F-Measure	32.54%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	791	528
	Negative	2751	739

#### A.3. Hasil Ekstraksi Objek Review Hotel (*Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	38.74%
Precision	64.49%
Recall	43.52%
F-Measure	51.97%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	574	745
	Negative	316	97

#### A.4. Hasil Ekstraksi Objek Review Hotel (*Frequent Object* dan *Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* dengan data filter sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	46.15%
Precision	63.02%
Recall	55.95%
F-Measure	59.27%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	738	581
	Negative	433	131

#### A.5. Hasil Ekstraksi Objek Review Hotel (*Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *filter object* sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	46.76%
Precision	62.36%
Recall	57.39%
F-Measure	59.77%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	757	562
	Negative	457	138

#### A.6. Hasil Ekstraksi Objek Review Restaurant (*Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	35.86%
Precision	44.82%
Recall	45.40%
F-Measure	45.11%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	316	380
	Negative	389	114

#### A.7. Hasil Ekstraksi Objek Review Restaurant (*Frequent Object* dan *Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	31.37%
Precision	21.60%
Recall	61.06%
F-Measure	31.91%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	425	271
	Negative	1543	404

#### A.8. Hasil Ekstraksi Objek Review Restaurant (*Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	39.91%
Precision	64.61%
Recall	45.11%
F-Measure	53.13%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	314	382
	Negative	172	54

#### A.9. Hasil Ekstraksi Objek Review Restaurant (*Frequent Object* dan *Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	48.93%
Precision	65.40%
Recall	59.20%
F-Measure	62.15%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	412	284
	Negative	218	69

#### A.10. Hasil Ekstraksi Objek Review Restaurant (*Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma ekstraksi objek data review hotel dengan pendekatan *filter object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	49.16%
Precision	63.76%
Recall	60.92%
F-Measure	62.31%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	424	272
	Negative	241	72

#### A.11. Hasil Deteksi Kalimat Opini Review Hotel (*Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	73.38%
Precision	88.15%
Recall	75.70%
F-Measure	81.45%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	461	148
	Negative	62	118

A.12. Hasil Deteksi Kalimat Opini Review Hotel (*Frequent Object* dan *Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* dan *infrequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	76.30%
Precision	85.76%
Recall	83.09%
F-Measure	84.40%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	506	103
	Negative	84	96

A.13. Hasil Deteksi Kalimat Opini Review Hotel (*Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* dengan data filter sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	72.50%
Precision	91.35%
Recall	71.10%
F-Measure	79.96%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	433	176
	Negative	41	139

#### A.14. Hasil Deteksi Kalimat Opini Review Hotel (*Frequent Object* dan *Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan performa deteksi kalimat opini data review hotel dengan pendekatan *frequent object* dan *infrequent object* dengan data filter sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	74.90%
Precision	89.90%
Recall	76.03%
F-Measure	82.39%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	463	146
	Negative	52	128

#### A.15. Hasil Deteksi Kalimat Opini Review Hotel (*Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *filter object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	75.16%
Precision	89.94%
Recall	76.35%
F-Measure	82.59%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	465	144
	Negative	52	128

#### A.16. Hasil Deteksi Kalimat Opini Review Restaurant (*Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	71.70%
Precision	90.41%
Recall	71.54%
F-Measure	79.88%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	264	105
	Negative	28	73

A.17. Hasil Deteksi Kalimat Opini Review Restaurant (*Frequent Object* dan *Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* dan *infrequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	78.94%
Precision	89.71%
Recall	82.66%
F-Measure	86.04%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	305	64
	Negative	35	66

A.18. Hasil Deteksi Kalimat Opini Review Restaurant (*Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	69.57%
Precision	93.13%
Recall	66.12%
F-Measure	77.33%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	244	125
	Negative	18	83

A.19. Hasil Deteksi Kalimat Opini Review Restaurant (*Frequent Object* dan *Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *frequent object* dan *infrequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	75.53%
Precision	93.20%
Recall	74.25%
F-Measure	82.65%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	274	95
	Negative	20	81

A.20. Hasil Deteksi Kalimat Opini Review Restaurant (*Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma deteksi kalimat opini data review hotel dengan pendekatan *filter object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	76.60%
Precision	93.31%
Recall	75.61%
F-Measure	83.53%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	279	90
	Negative	20	81

#### A.21. Hasil Klasifikasi Sentimen Objek Review Hotel (*Correct Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	88.03%
Precision	90.81%
Recall	94.36%
F-Measure	92.55%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	435	26
	Negative	44	80

#### A.22. Hasil Klasifikasi Sentimen Objek Review Hotel (*Correct Frequent Object* dan *Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* yang benar sesuai data corpus sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	87.61%
Precision	90.88%
Recall	93.00%
F-Measure	91.93%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	558	42
	Negative	56	135

#### A.23. Hasil Klasifikasi Sentimen Objek Review Hotel (*Correct Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* dengan filter yang benar sesuai data corpus sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	88.15%
Precision	90.91%
Recall	94.51%
F-Measure	92.68%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	430	25
	Negative	43	76

A.24. Hasil Klasifikasi Sentimen Objek Review Hotel (*Correct Frequent Object and Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* dengan data filter yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	87.67%
Precision	91.33%
Recall	92.78%
F-Measure	92.05%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	527	41
	Negative	50	120

A.25. Hasil Klasifikasi Sentimen Objek Review Hotel (*Correct Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *filter object* yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	85.47%
Precision	90.02%
Recall	91.25%
F-Measure	90.63%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	532	115
	Negative	59	128

A.26. Hasil Klasifikasi Sentimen Objek Review Restaurant (*Correct Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	91.14%
Precision	95.27%
Recall	94.58%
F-Measure	94.92%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	262	15
	Negative	13	26

A.27. Hasil Klasifikasi Sentimen Objek Review Restaurant (*Correct Frequent Object and Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	88.71%
Precision	93.61%
Recall	93.09%
F-Measure	93.35%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	337	25
	Negative	23	40

#### A.28. Hasil Klasifikasi Sentimen Objek Review Restaurant (*Correct Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan performa klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* dengan filter yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	91.08%
Precision	95.24%
Recall	94.55%
F-Measure	94.89%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	260	15
	Negative	13	26

A.29. Hasil Klasifikasi Sentimen Objek Review Restaurant (*Correct Frequent Object and Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *frequent object* dan *infrequent object* dengan filter yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	88.59%
Precision	93.41%
Recall	93.14%
F-Measure	93.27%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	326	24
	Negative	23	39

A.30. Hasil Klasifikasi Sentimen Objek Review Restaurant (*Correct Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan *filter object* yang benar sesuai data corpus sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	87.03%
Precision	91.94%
Recall	92.72%
F-Measure	92.33%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	331	26
	Negative	29	38

### A.31. Hasil Klasifikasi Sentimen Objek Review Hotel (*All Frequent Object*)

Berikut ini merupakan hasil perhitungan performa klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	40.23%
Precision	44.94%
Recall	65.22%
F-Measure	53.21%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	435	232
	Negative	533	80

### A.32. Hasil Klasifikasi Sentimen Objek Review Hotel (*All Frequent Object and Infrequent Object*)

Berikut ini merupakan hasil perhitungan performa klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* dan *infrequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	19.57%
Precision	21.59%
Recall	40.43%
F-Measure	28.15%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	558	822
	Negative	2027	135

#### A.33. Hasil Klasifikasi Sentimen Objek Review Hotel (*All Frequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	56.85%
Precision	60.91%
Recall	79.93%
F-Measure	69.14%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	430	108
	Negative	276	76

#### A.34. Hasil Klasifikasi Sentimen Objek Review Hotel (*All Frequent Object and Infrequent Object* dengan Filter)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* dan *infrequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	55.25%
Precision	59.41%
Recall	76.27%
F-Measure	66.79%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	527	164
	Negative	360	120

#### A.35. Hasil Klasifikasi Sentimen Objek Review Hotel (*All Filter Object*)

Berikut ini merupakan hasil perhitungan performa klasifikasi sentimen objek data review hotel dengan pendekatan semua *filter object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	53.29%
Precision	58.46%
Recall	73.79%
F-Measure	65.24%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	532	189
	Negative	378	128

### A.36. Hasil Klasifikasi Sentimen Objek Review Restaurant (*All Frequent Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	40.85%
Precision	44.71%
Recall	73.80%
F-Measure	55.68%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	262	93
	Negative	324	26

### A.37. Hasil Klasifikasi Sentimen Objek Review Restaurant (*All Frequent Object and Infrequent Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* dan *infrequent object* sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	19.16%
Precision	21.06%
Recall	50.68%
F-Measure	29.76%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	337	328
	Negative	1263	40

A.38. Hasil Klasifikasi Sentimen Objek Review Restaurant (*All Frequent Object dengan Filter*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* dengan data filter sebagai acuan ekstraksi.

Pengukuran	Nilai
Accuracy	58.85%
Precision	63.26%
Recall	84.14%
F-Measure	72.22%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	260	49
	Negative	151	26

A.39. Hasil Klasifikasi Sentimen Objek Review Restaurant (*All Frequent Object and Infrequent Object dengan Filter*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *frequent object* dan *infrequent object* dengan data filter sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	57.94%
Precision	61.51%
Recall	84.24%
F-Measure	71.10%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	326	61
	Negative	204	39

#### A.40. Hasil Klasifikasi Sentimen Objek Review Restaurant (*All Filter Object*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen objek data review hotel dengan pendekatan semua *filter object* sebagai acuan ekstraksi.

<b>Pengukuran</b>	<b>Nilai</b>
Accuracy	55.49%
Precision	59.32%
Recall	82.75%
F-Measure	69.10%

Dan berikut ini merupakan *confusion matrix*-nya

		Predicted	
		Positive	Negative
Actual	Positive	331	69
	Negative	227	38

*Halaman ini sengaja dikosongkan.*

## LAMPIRAN B

Lampiran ini berisi mengenai hasil pengukuran performa klasifikasi sentimen kalimat review secara lengkap. Nilai hasil pengukuran yang tertera berikut ini antara lain adalah nilai *correctly instance*, *incorrect instance*, akurasi, *precision*, *recall*, *f-measure*, AUC, dan *confusion matrix*.

### B.1. Hasil Klasifikasi Sentimen Kalimat *Review Hotel* (*Sentence Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan kalimat *review* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	445	445	445
Incorrectly Classified Instances	344	344	344
TP Rate	0.56	0.56	0.56
FP Rate	0.56	0.56	0.56
Akurasi (%)	56.40	56.40	56.40
Precision (%)	31.80	31.80	31.80
Recall (%)	56.40	56.40	56.40
F-Measure (%)	40.70	40.70	40.70
ROC Area (AUC)	0.49	0.50	0.49

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	445	0	0
	Neutral	181	0	0
	Negative	163	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	445	0	0
	Neutral	181	0	0
	Negative	163	0	0

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	445	0	0
	Neutral	181	0	0
	Negative	163	0	0

## B.2. Hasil Klasifikasi Sentimen Kalimat *Review Hotel* (*Corpus Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan *corpus object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	682	682	682
Incorrectly Classified Instances	107	107	107
TP Rate	0.86	0.86	0.86
FP Rate	0.17	0.17	0.18
Akurasi (%)	86.43	86.43	86.44
Precision (%)	89.10	89.10	89.10
Recall (%)	86.40	86.40	86.40
F-Measure (%)	83.90	83.90	83.90
ROC Area (AUC)	0.91	0.86	0.91

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	445	0	0
	Neutral	106	180	0
	Negative	1	0	57

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	445	0	0
	Neutral	106	180	0
	Negative	1	0	57

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	445	0	0
	Neutral	106	180	0
	Negative	1	0	57

### B.3. Hasil Klasifikasi Sentimen Kalimat *Review Hotel (Frequent Object Attribute)*

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan *frequent object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	449	475	474
Incorrectly Classified Instances	340	314	315
TP Rate	0.57	0.60	0.60
FP Rate	0.54	0.31	0.31
Akurasi	56.91	60.20	60.08
Precision	50.00	61.80	61.70
Recall	56.90	60.20	60.10
F-Measure	43.70	55.80	55.70
ROC Area (AUC)	0.57	0.64	0.66

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	436	5	4
	Neutral	176	2	3
	Negative	150	2	11

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	344	97	4
	Neutral	58	120	3
	Negative	94	58	11

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	343	98	4
	Neutral	58	120	3
	Negative	94	58	11

#### B.4. Hasil Klasifikasi Sentimen Kalimat *Review Hotel (Frequent and Infrequent Object Attribute)*

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan *frequent object* dan *infrequent object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	444	476	476
Incorrectly Classified Instances	345	313	313
TP Rate	0.56	0.603	0.60
FP Rate	0.56	0.382	0.38
Akurasi (%)	56.27	60.33	60.33
Precision (%)	31.80	47.40	47.40
Recall (%)	56.30	60.30	60.30
F-Measure (%)	40.60	53.00	53.00
ROC Area (AUC)	0.59	0.61	0.60

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	444	1	0
	Neutral	181	0	0
	Negative	163	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	380	65	0
	Neutral	85	96	0
	Negative	124	39	0

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	380	65	0
	Neutral	85	96	0
	Negative	124	39	0

#### B.5. Hasil Klasifikasi Sentimen Kalimat *Review Hotel* (*Filtered Frequent Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan *frequent object* yang telah terfilter sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	458	494	492
Incorrectly Classified Instances	331	295	297
TP Rate	0.58	0.63	0.62
FP Rate	0.52	0.24	0.24
Akurasi (%)	58.05	62.61	62.36
Precision (%)	46.80	67.00	67.10
Recall (%)	58.00	62.60	62.40
F-Measure (%)	45.70	60.60	60.10
ROC Area (AUC)	0.63	0.69	0.71

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	437	2	6
	Neutral	177	0	4
	Negative	138	4	21

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	327	111	7
	Neutral	36	139	6
	Negative	68	67	28

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	327	112	6
	Neutral	36	140	5
	Negative	68	70	25

#### B.6. Hasil Klasifikasi Sentimen Kalimat *Review Hotel (Filtered Frequent and Infrequent Object Attribute)*

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan *frequent object* dan *infrequent object* yang telah terfilter sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	441	482	484
Incorrectly Classified Instances	348	307	305
TP Rate	0.56	0.61	0.61
FP Rate	0.55	0.30	0.29
Akurasi (%)	55.89	61.09	61.34
Precision (%)	41.00	60.50	62.20
Recall (%)	55.90	61.10	61.30
F-Measure (%)	42.00	56.70	57.10
ROC Area (AUC)	0.67	0.66	0.68

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	434	4	7
	Neutral	179	0	2
	Negative	154	2	7

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	343	94	8
	Neutral	50	128	3
	Negative	94	58	11

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	343	95	7
	Neutral	50	129	2
	Negative	91	60	12

#### B.7. Hasil Klasifikasi Sentimen Kalimat *Review Hotel* (*Filter Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review hotel* dengan *filter object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	444	489	488
Incorrectly Classified Instances	345	300	301
TP Rate	0.56	0.62	0.62
FP Rate	0.55	0.31	0.29
Akurasi (%)	56.27	61.98	61.85
Precision (%)	42.40	63.10	63.50
Recall (%)	56.30	62.00	61.90
F-Measure (%)	41.90	57.40	57.10
ROC Area (AUC)	0.67	0.67	0.69

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	439	4	2
	Neutral	178	0	3
	Negative	154	4	5

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	355	88	2
	Neutral	54	122	5
	Negative	100	51	12

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	349	94	2
	Neutral	48	129	4
	Negative	95	58	10

#### B.8. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Sentence Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan kalimat *review* sebagai atribut.

<b>Pengukuran</b>	<b>Nilai</b>		
	<b>Naïve Bayes</b>	<b>SVM</b>	<b>Random Forests</b>
Correctly Classified Instances	302	302	302
Incorrectly Classified Instances	168	168	168
TP Rate	0.64	0.64	0.64
FP Rate	0.64	0.64	0.64
Akurasi (%)	64.26	64.26	64.26
Precision (%)	41.30	41.30	41.30
Recall (%)	64.30	64.30	64.30
F-Measure (%)	50.30	50.30	50.30
ROC Area (AUC)	0.49	0.50	0.49

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	302	0	0
	Neutral	102	0	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	302	0	0
	Neutral	102	0	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	302	0	0
	Neutral	102	0	0
	Negative	66	0	0

#### B.9. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Corpus Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan *corpus object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	416	419	419
Incorrectly Classified Instances	54	51	51
TP Rate	0.89	0.89	0.89
FP Rate	0.21	0.19	0.19
Akurasi (%)	88.51	89.15	89.15
Precision (%)	90.30	90.10	90.10
Recall (%)	88.50	89.10	89.10
F-Measure (%)	85.20	86.50	86.50
ROC Area (AUC)	0.91	0.86	0.916

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	302	0	0
	Neutral	1	101	0
	Negative	53	0	13

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	301	0	1
	Neutral	1	101	0
	Negative	49	0	17

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	301	0	1
	Neutral	1	101	0
	Negative	49	0	17

#### B.10. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Frequent Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan *frequent object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	300	295	287
Incorrectly Classified Instances	170	175	183
TP Rate	0.64	0.63	0.61
FP Rate	0.64	0.64	0.60
Akurasi (%)	63.83	62.77	61.06
Precision (%)	49.90	49.80	48.90
Recall (%)	63.80	62.80	61.10
F-Measure (%)	50.80	50.40	51.70
ROC Area (AUC)	0.48	0.55	0.63

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	298	3	1
	Neutral	100	2	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	293	3	6
	Neutral	99	2	1
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	278	17	7
	Neutral	93	8	1
	Negative	61	4	1

#### B.11. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Frequent and Infrequent Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan *frequent object* dan *infrequent object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	302	317	317
Incorrectly Classified Instances	168	153	153
TP Rate	0.64	0.67	0.67
FP Rate	0.64	0.37	0.37
Akurasi (%)	64.26	67.45	67.45
Precision (%)	41.30	58.70	58.70
Recall (%)	64.30	67.40	67.40
F-Measure (%)	50.30	62.70	62.70
ROC Area (AUC)	0.45	0.66	0.63

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	302	0	0
	Neutral	102	0	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	251	49	2
	Neutral	36	66	0
	Negative	51	15	0

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	251	49	2
	Neutral	36	66	0
	Negative	51	15	0

#### B.12. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Filtered Frequent Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan *frequent object* terfilter sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	299	296	297
Incorrectly Classified Instances	171	174	173
TP Rate	0.64	0.63	0.63
FP Rate	0.64	0.61	0.60
Akurasi (%)	63.62	62.98	63.19
Precision (%)	47.40	51.00	53.40
Recall (%)	63.60	63.00	63.20
F-Measure (%)	50.70	51.90	52.60
ROC Area (AUC)	0.53	0.56	0.65

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	297	5	0
	Neutral	100	2	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	290	4	8
	Neutral	98	2	2
	Negative	60	2	4

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	289	4	9
	Neutral	97	3	2
	Negative	59	2	5

### B.13. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Filtered Frequent and Infrequent Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan *frequent object* dan *infrequent object* yang telah terfilter sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	296	298	304
Incorrectly Classified Instances	174	172	166
TP Rate	0.63	0.63	0.65
FP Rate	0.65	0.34	0.30
Akurasi (%)	62.98	63.40	64.68
Precision (%)	41.00	63.20	65.50
Recall (%)	63.00	63.40	64.70
F-Measure (%)	49.70	60.30	61.70
ROC Area (AUC)	0.51	0.66	0.68

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	296	5	1
	Neutral	102	0	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	224	75	3
	Neutral	30	72	0
	Negative	45	19	2

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	219	80	3
	Neutral	19	83	0
	Negative	43	21	2

B.14. Hasil Klasifikasi Sentimen Kalimat *Review* Restoran (*Filter Object Attribute*)

Berikut ini merupakan hasil perhitungan perfoma klasifikasi sentimen kalimat *review* restoran dengan *filter object* sebagai atribut.

Pengukuran	Nilai		
	Naïve Bayes	SVM	Random Forests
Correctly Classified Instances	296	301	307
Incorrectly Classified Instances	174	169	163
TP Rate	0.63	0.64	0.65
FP Rate	0.65	0.35	0.30
Akurasi (%)	62.98	64.04	65.32
Precision (%)	41.00	63.40	65.60
Recall (%)	63.00	64.00	65.30
F-Measure (%)	49.70	60.80	62.10
ROC Area (AUC)	0.50	0.66	0.67

Berikut ini merupakan *confusion matrix* Naïve Bayes.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	296	5	1
	Neutral	102	0	0
	Negative	66	0	0

Berikut ini merupakan *confusion matrix* SVM.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	227	72	3
	Neutral	30	72	0
	Negative	46	18	2

Berikut ini merupakan *confusion matrix* Random Forests.

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	222	77	3
	Neutral	19	83	0
	Negative	45	19	2

*Halaman ini sengaja dikosongkan.*

## LAMPIRAN C

Lampiran ini berisi mengenai hasil output dari *object-based opinion mining* secara lengkap. Nilai hasil pengukuran yang tertera berikut ini antara lain adalah objek beserta jumlah kemunculan dalam review untuk masing-masing pendekatan yang ada.

### C.1. Hasil *Object-based Opinion Mining* Hotel (*Frequent Object*)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review hotel dengan pendekatan *frequent object* sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	hotel	109	35	38	everyth	9	3
2	room	103	37	39	rate	6	1
3	stay	39	12	40	shop	5	2
4	breakfast	45	11	41	minut	4	2
5	lake	40	5	42	montt	7	2
6	staff	41	7	43	coffe	7	2
7	view	42	3	44	lobbi	6	3
8	place	34	9	45	window	3	4
9	locat	30	1	46	bedroom	9	1
10	town	18	6	47	review	3	5
11	night	11	6	48	dinner	3	2
12	restaur	22	4	49	experi	4	4
13	servic	20	6	50	lot	5	1
14	area	18	4	51	peopl	8	1
15	walk	14	4	52	park	5	2
16	puerto	15	5	53	desk	1	4
17	water	12	7	54	trip	6	1
18	bed	18	0	55	recept	5	2
19	bit	7	12	56	side	3	3
20	vara	11	4	57	bread	5	2
21	time	8	4	58	distantc	4	1
22	pool	13	4	59	space	5	1
23	food	17	2	60	peulla	5	2
24	pucon	9	3	61	kitchen	8	0
25	problem	4	12	62	casino	7	1
26	bathroom	15	3	63	year	3	4
27	day	6	5	64	floor	5	2
28	center	8	1	65	shower	4	3
29	wifi	11	3	66	owner	5	2
30	citi	7	5	67	balconi	6	1
31	front	6	5	68	fireplac	5	0
32	thing	5	6	69	fruit	8	0
33	way	5	4	70	tv	3	2
34	street	6	3	71	garden	6	1
35	price	4	4	38	everyth	9	3
36	volcano	11	0	39	rate	6	1
37	car	6	3	40	shop	5	2

## C.2. Hasil *Object-based Opinion Mining* Hotel (*frequent and Infrequent Object*)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review hotel dengan pendekatan *frequent dan infrequent object* sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	hotel	107	37	52	price	5	3
2	room	104	36	53	everyth	10	2
3	stay	38	13	54	look	5	4
4	breakfast	43	13	55	rate	6	1
5	lake	41	4	56	simpl	9	1
6	staff	41	7	57	offer	7	2
7	view	43	2	58	use	4	4
8	place	34	9	59	coffe	7	2
9	clean	41	0	60	minut	4	2
10	locat	29	2	61	lobbi	6	3
11	town	18	6	62	montt	7	2
12	night	10	7	63	shop	4	3
13	beauti	28	1	64	feel	8	0
14	restaur	21	5	65	park	5	2
15	one	13	10	66	dinner	2	3
16	servic	19	7	67	peopl	8	1
17	help	23	2	68	bedroom	8	2
18	area	18	4	69	lot	4	2
19	puerto	14	6	70	review	3	5
20	walk	13	5	71	window	4	3
21	water	11	8	72	experi	4	4
22	excel	22	0	73	distanc	4	1
23	recommend	17	2	74	amaz	9	0
24	bed	18	0	75	trip	6	1
25	bit	7	12	76	serv	5	3
26	vara	10	5	77	tour	4	3
27	time	8	4	78	side	3	3
28	food	17	2	79	bread	5	2
29	pool	10	7	80	outsid	5	4
30	bathroom	14	4	81	recept	5	2
31	pucon	9	3	82	desk	1	4
32	problem	5	13	83	kitchen	8	0
33	day	6	5	84	tv	3	2
34	right	15	3	85	peulla	4	3
35	even	5	5	86	kind	7	0
36	wonder	15	1	87	year	3	4
37	arriv	7	2	88	balconi	6	1
38	warm	14	2	89	owner	5	2
39	center	8	1	90	fruit	8	0
40	need	3	9	91	take	4	3
41	citi	7	5	92	del	2	3
42	wifi	11	3	93	casino	6	2
43	front	6	5	94	shower	3	4
44	back	2	4	95	local	6	0
45	thing	5	6	96	internet	3	4
46	work	9	5	97	fireplac	5	0
47	street	6	3	98	space	5	1
48	way	5	4	99	floor	5	2
49	volcano	11	0	100	cross	1	0
50	perfect	9	3	101	averag	3	1
51	car	6	3	102	home	3	2

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
103	hot water	3	4	162	decor	4	0
104	chile	5	1	163	buffet	3	2
105	cold	1	6	164	run	2	2
106	book	2	2	165	noisi	2	3
107	facil	5	1	166	mirador	3	2
108	dog	1	2	167	point	3	2
109	access	4	3	168	disappoint	1	4
110	travel	4	2	169	mountain	4	1
111	option	5	2	170	towel	2	2
112	call	2	2	171	fact	3	1
113	dine	6	0	172	loung	4	1
114	ground	5	1	173	hill	3	0
115	heat	4	1	174	spa	5	0
116	garden	6	1	175	drink	2	1
117	visit	3	1	176	excurs	2	0
118	noth	1	1	177	usd	3	1
119	surpris	6	0	178	differ	3	0
120	surround	4	0	179	english	0	3
121	sur	2	3	180	accommod	4	0
122	read	3	2	181	lago	3	0
123	set	4	1	182	great place	5	0
124	build	3	1	183	invit	5	0
125	chair	4	1	184	activ	2	2
126	hous	6	0	185	qualiti	2	2
127	chees	5	0	186	light	3	2
128	attract	5	0	187	free wifi	4	1
129	road	2	2	188	nois	3	2
130	left	2	2	189	daughter	3	0
131	lunch	1	0	190	peac	4	0
132	live	6	0	191	detail	1	2
133	bu	3	0	192	downtown	1	1
134	min	2	1	193	villarica	3	0
135	tabl	4	0	194	import	2	1
136	morn	3	0	195	face	1	1
137	egg	5	1	196	puelto montt	3	1
138	etc	4	1	197	properti	3	0
139	start	1	3	198	rest	3	1
140	iva	2	0	199	jam	1	2
141	chang	3	2	200	group	1	2
142	la	4	1	201	famili	4	0
143	ok	2	2	202	cours	3	0
144	share	3	1	203	lack	0	4
145	suit	3	0	204	signal	2	1
146	door	3	1	205	swim	2	1
147	charm	3	1	206	osorno	4	0
148	natura	3	0	207	vacat	2	1
149	colono	2	2	208	posit	3	1
150	wood	5	0	209	channel	1	2
151	centr	3	2	210	smoke	0	4
152	hotel natura	3	0	211	cake	3	0
153	luxuri	3	2	212	claim	2	1
154	hour	2	1	213	citi center	3	1
155	choic	5	0	214	instant	2	1
156	guest	3	2	215	bar	2	1
157	valu	5	0	216	speak	0	2
158	coupl	3	1	217	smell	1	3
159	region	3	0	218	meal	3	0
160	good valu	5	0	219	cabin	3	0
161	castro	3	0	220	extra	3	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
221	weather	3	1	279	comput	2	1
222	level	1	3	280	site	1	0
223	lake cross	1	0	281	dirti	0	3
224	juic	3	1	282	touch	3	0
225	size	3	0	283	world	1	2
226	apart	2	1	284	lake view	3	0
227	ladi	3	1	285	wing	2	1
228	friend	3	1	286	lodg	1	1
229	solac	4	0	287	afternoon	2	1
230	decemb	1	1	288	desper need	0	3
231	terrac	3	1	289	someth	2	1
232	rental	1	1	290	heart	1	0
233	public	2	1	291	terrif	2	0
234	date	0	3	292	step	3	0
235	reason	3	1	293	mall	2	0
236	expect	1	3	294	ham	3	0
237	driver	3	0	295	p montt	1	1
238	accomod	3	0	296	gift	2	0
239	hostel	3	0	297	safe	3	0
240	bath	2	2	298	airi	2	1
241	basic	3	1	299	furnish	3	0
242	peso	1	1	300	ocean	1	0
243	palafito	4	0	301	nicer	3	0
244	rain	1	1	302	upgrad	2	1
245	check	1	1	303	court	0	2
246	tea	3	1	304	complain	0	2
247	comfi	3	0	305	tast	2	1
248	taxi	0	1	306	puelch	2	1
249	impress	1	0	307	base	2	0
250	villag	2	0	308	aspect	3	0
251	friendli staff	3	0	309	issu	1	2
252	ride	3	0	310	parti	1	1
253	end	2	1	311	cafe	1	0
254	manag	1	2	312	direct	1	0
255	dish	1	1	313	stop	1	0
256	filter	0	2	314	fire	3	0
257	dine room	2	0	315	antumal	1	1
258	superb	3	0	316	wine	3	0
259	block	2	0	317	money	2	0
260	hotel puelch	2	1	318	request	2	1
261	verd	3	0	319	sofa	1	0
262	cab	2	0	320	lo	3	0
263	cabana	1	0	321	mold	2	1
264	key	2	1	322	trail	2	0
265	corner	2	0	323	breakfast buffet	2	1
266	short walk	1	1	324	natur	2	0
267	spot	2	0	325	miss	1	2
268	hotel peulla	2	1	326	jorg	3	0
269	llanquihu	3	0	327	volcan	3	0
270	good hotel	3	0	328	bottl	1	1
271	cereal	3	0	329	petrohu	3	0
272	indoor	2	0	330	hand	1	1
273	anim	2	1	331	front desk	0	2
274	transport	2	1	332	star	0	2
275	exchang	1	1	333	drive	1	0
276	frutillar	3	0	334	room servic	3	0
277	linen	3	0	335	standard	1	1
278	tourist	1	2	336	driveway	2	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
337	equip	1	1	394	onli drawback	1	1
338	agenc	2	0	395	commun	1	1
339	littl cold	0	2	396	chilo	1	0
340	free park	2	0	397	corner room	2	0
341	atmospher	2	0	398	renov	1	1
342	cleanli	2	0	399	express	1	0
343	temuco	0	2	400	postcard	1	1
344	boat	1	0	401	excel hotel	2	0
345	downsid	0	2	402	lakesid	1	1
346	amaz view	2	0	403	question	1	1
347	rental car	1	0	404	book exchang	1	0
348	stone	0	1	405	luggag	0	1
349	person	0	1	406	ensenada	2	0
350	tenni	0	1	407	ton	0	1
351	elev	1	1	408	dark	0	2
352	mention	0	1	409	district	1	1
353	deck	1	0	410	background	2	0
354	live room	2	0	411	menu	1	0
355	heater	0	2	412	great staff	2	0
356	thin	1	0	413	mont verd	2	0
357	relax	2	0	414	fresh fruit	2	0
358	buse	0	1	415	dump	1	1
359	host	2	0	416	variети	1	1
360	santiago	1	1	417	entranc	1	0
361	tenni court	0	1	418	star hotel	0	1
362	first night	2	0	419	edg	0	1
363	gym	1	1	420	spice	2	0
364	inn	2	0	421	plenti	2	0
365	good breakfast	1	1	422	cup	1	0
366	anyth	0	1	423	hotel weisserhau	2	0
367	advisor	1	1	424	roll	1	1
368	match	0	2	425	buffet breakfast	1	1
369	brazilian	1	0	426	countri	1	1
370	tip	1	0	427	new year	0	1
371	holiday	2	0	428	bird	2	0
372	hotel frontera	1	0	429	silenc	1	1
373	write	1	0	430	nice view	2	0
374	busi	0	2	431	age	1	1
375	south	0	1	432	great locat	2	0
376	languag	0	1	433	event	2	0
377	spread	1	0	434	middl	0	1
378	santo	2	0	435	client	2	0
379	waterfal	1	0	436	meat	1	0
380	christma	2	0	437	bathhtub	1	1
381	refriger	1	1	438	attent	2	0
382	winter	1	0	439	flower	2	0
383	landscap	1	0	440	cost	0	2
384	long time	1	1	441	spring	2	0
385	phone	0	2	442	beach	0	2
386	drawback	1	1	443	great view	2	0
387	littl bit	2	0	444	boyfriend	2	0
388	continu	1	0	445	hit	0	1
389	custom	1	1	446	writer	1	0
390	susan	2	0	447	paper	1	0
391	chef	1	0	448	hotel staff	2	0
392	report	1	0	449	good locat	2	0
393	discount	1	0	450	hotel restaur	2	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
451	minut walk	0	2	507	glass	2	0
452	fixtur	2	0	508	stuff	1	0
453	frontera	1	0	509	small balconi	1	0
454	bike	0	1	510	personnel	2	0
455	beauti town	2	0	511	sceneri	1	1
456	cliff	1	1	512	yogurt	1	0
457	stroll	2	0	513	bag	1	1
458	part	1	1	514	receptionist	0	1
459	sheep	2	0	515	nice touch	2	0
460	style	2	0	516	cabl tv	1	0
461	barrier	0	1	517	lo santo	2	0
462	prior writer	1	0	518	piec	1	1
463	top	2	0	519	weekend	1	0
464	real coffe	2	0	520	lake district	1	1
465	maker	1	1	521	assist	2	0
466	someon	1	0	522	toaster	1	1
467	iron	1	1	523	weisserhau	2	0
468	attend	0	1	524	color	1	0
469	canopi	2	0	525	good select	2	0
470	languag barrier	0	1	526	note	1	0
471	wonder place	1	1	527	unfair bad rap	0	1
472	laundri	1	1	528	swimminpool	1	0
473	conveni	2	0	529	non smoke	0	1
474	ceil	0	1	530	bracelet	0	1
475	hike	1	0	531	shabbi	0	1
476	sheet	1	1	532	hint	1	0
477	outdoor pool	0	2	533	jump	0	1
478	suggest	1	0	534	brand	1	0
479	mattress	2	0	535	meringu pie	0	1
480	tower	1	0	536	superior doubl room	1	0
481	select	2	0	537	pastri	1	0
482	good choic	2	0	538	good spa	1	0
483	guid	0	1	539	fridg	1	0
484	bakeri	1	0	540	santa	1	0
485	rack	1	1	541	view room	1	0
486	cheaper	2	0	542	bay	1	0
487	music	1	1	543	greeneri	1	0
488	breakfast area	1	1	544	deterg	0	1
489	schedul	2	0	545	rec	1	0
490	river	2	0	546	anthoni	0	1
491	stilt	2	0	547	fool	0	1
492	mont	2	0	548	croissant	1	0
493	birthday	2	0	549	leftov	0	1
494	and	1	0	550	inner courtyard	1	0
495	larg room	1	0	551	nois problem	1	0
496	cabl	1	0	552	sir	1	0
497	nice hotel	2	0	553	reaction	1	0
498	anyon	1	1	554	exchang rate	0	1
499	addit	1	0	555	facilitit	0	1
500	internet access	1	1	556	sun	1	0
501	stove	2	0	557	comfort bed	1	0
502	dollar	1	0	558	health	0	1
503	patio	2	0	559	below averag	1	0
504	larg bathroom	2	0	560	pleasant experi	1	0
505	husband	1	0	561	beauti region	1	0
506	plu	2	0	562	averag food	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
563	furnitur	1	0	614	spite	0	1
564	doubl bedroom	1	0	615	loan	0	1
565	perfect spot	1	0	616	memori	1	0
566	sanctuari	0	1	617	strong german influenc	1	0
567	littl run	0	1	618	funni thing	0	1
568	board locat	1	0	619	hustl	1	0
569	great- hardwood	1	0	620	creativ write	1	0
570	mother	1	0	621	girl	1	0
571	ball	0	1	622	list	1	0
572	damag	1	0	623	ultra modern design	1	0
573	beauti garden	1	0	624	muzak	0	1
574	dingi	0	1	625	event driveway	1	0
575	eas	1	0	626	nice shop	1	0
576	strong coffe	0	1	627	moneyplac	0	1
577	getaway	1	0	628	tini room	0	1
578	oregan wood	1	0	629	condiment	1	0
579	stage	0	1	630	bookingscomw	0	1
580	marmelad	1	0	631	robe	1	0
581	place wonder	1	0	632	shower wand	0	1
582	bbc world	0	1	633	bookcas	1	0
583	rate staff	1	0	634	tv guid	0	1
584	exampl	1	0	635	vulcano	1	0
585	nice overview	1	0	636	nice bed	1	0
586	good heat	1	0	637	third day	0	1
587	itinerari	0	1	638	couch	1	0
588	mini-bar	1	0	639	rabbit	1	0
589	hand towel	0	1	640	fluenci	0	1
590	improv	1	0	641	fun	1	0
591	bet	1	0	642	item	0	1
592	old hous	1	0	643	dond	0	1
593	maze	1	0	644	cleaner	1	0
594	postcard volcan osorno	1	0	645	own pace	0	1
595	rec room	1	0	646	wifi signal	1	0
596	gordito	0	1	647	pool area	1	0
597	fraction	0	1	648	refresh time	1	0
598	bit noisi	0	1	649	fettish	1	0
599	manybett	1	0	650	jungl	0	1
600	action	0	1	651	degre	0	1
601	doubt	1	0	652	stun place	1	0
602	entir stay	1	0	653	game	1	0
603	stung	0	1	654	king	1	0
604	no-smok hotel	0	1	655	mix bag	1	0
605	stop point	1	0	656	contemporari hotel	1	0
606	strip	1	0	657	room price	1	0
607	wonder locat	1	0	658	tradit rustic furnitur	1	0
608	good food	1	0	659	great atmospher	1	0
609	atrium	1	0	660	instant espresso	1	0
610	uerto vara	1	0	661	rosario	1	0
611	puerto vara	1	0	662	amaz room	1	0
612	waiter	1	0	663	fond	0	1
613	surgeri	0	1	664	resort	0	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
665	volunt	1	0	719	sign	1	0
666	hotel principado	0	1	720	architectur masterpiec	1	0
667	short ride	1	0	721	extens mold damag	1	0
668	sunday	1	0	722	first room	1	0
669	masterpiec	1	0	723	busi travel	0	1
670	nicer room	1	0	724	tad	1	0
671	perfect day	1	0	725	excel coffe	1	0
672	hall	1	0	726	optionsridicul bracelet	0	1
673	al	0	1	727	simpl fact	1	0
674	grilla	1	0	728	coffe maker	1	0
675	brilliant option	1	0	729	brain	0	1
676	locationview	0	1	730	switch	0	1
677	littl shabbi	0	1	731	larg fireplac	1	0
678	pot	1	0	732	plug	0	1
679	grass	1	0	733	cargo	1	0
680	sur mirador hotel	0	1	734	larg deck	1	0
681	spa feel	1	0	735	liucuria river bridg	1	0
682	travel agent	1	0	736	vicent	1	0
683	decis maker	0	1	737	great breakfast	1	0
684	tour offic	1	0	738	deliveri	1	0
685	platform	1	0	739	wrap	1	0
686	regret	0	1	740	dond el gordito	0	1
687	disappoint qualiti	0	1	741	condition4-	1	0
688	simpl meal	1	0	742	park area	1	0
689	bridg	1	0	743	glori	1	0
690	invit space	1	0	744	themVeri	1	0
691	fantast experi	1	0	745	gener level	0	1
692	onli neg item	0	1	746	hole	0	1
693	good start	1	0	747	osorno volcano	1	0
694	anthoni bourdain	0	1	748	respect	1	0
695	littl cleaner	1	0	749	comfort way	1	0
696	bonu	1	0	750	rud	0	1
697	main area	0	1	751	terribl mold	0	1
698	station	1	0	752	wast water	1	0
699	pression	0	1	753	cab ride	1	0
700	curtain	0	1	754	ski	1	0
701	actual town	0	1	755	woman	1	0
702	loung area	1	0	756	king size bed	1	0
703	spa fettish	1	0	757	free broadband	1	0
704	small hotel	1	0	758	rang	1	0
705	dust	0	1	759	other hotel	0	1
706	despar	0	1	760	read nook	1	0
707	bee	0	1	761	quiet street	1	0
708	restaur la	1	0	762	littl balconi	1	0
709	sur mirador	0	1	763	seat area	1	0
710	everybodi	0	1	764	amaz place	1	0
711	cloud	1	0	765	amenit	1	0
712	floor atrium	1	0	766	larg hotel	0	1
713	botiqu	1	0	767	outstand food	1	0
714	fine pale wood	1	0	768	albergu	1	0
715	excel breakfast	1	0	769	luxuri hotel	1	0
716	welfar	0	1	770	excel servic	1	0
717	squash court	0	1	771	excel dine	1	0
718	next day	1	0	772	makeovertini elev restaur	0	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
773	time i	1	0	826	mould	0	1
774	pleasant town	1	0	827	easi chair	1	0
775	instant access	0	1	828	stretch	1	0
776	second tabl	1	0	829	them24	1	0
777	shelf	0	1	830	termin	1	0
778	bath room	1	0	831	close by	1	0
779	bar area comfi	1	0	832	room key	1	0
780	wonder breakfast	1	0	833	tourist accommod	1	0
781	good view	1	0	834	birthday weekend	1	0
782	onli problem	0	1	835	colleagu	0	1
783	machin	0	1	836	cookwar	1	0
784	right wing	1	0	837	opposit side	0	1
785	weak wifi	0	1	838	tre	1	0
786	box	1	0	839	great food	1	0
787	good local	1	0	840	disabl	1	0
788	great bathroom	1	0	841	themVeri conveni	1	0
789	nice kitchen	1	0	842	kind ladi	1	0
790	lake llanquihu	1	0	843	hat	1	0
791	old wing	0	1	844	system	0	1
792	free good wifi	1	0	845	sum	0	1
793	next morn	1	0	846	staff treat	0	1
794	footprint	1	0	847	check-in	1	0
795	good basic accomod	1	0	848	christin	1	0
796	curtain rail	0	1	849	whole properti	1	0
797	attendand	1	0	850	din	1	0
798	what	0	1	851	meringu	0	1
799	foreign	1	0	852	car park	0	1
800	high season	1	0	853	doorman	1	0
801	other inform	1	0	854	nice option	1	0
802	thing lack	0	1	855	small patio	1	0
803	mushroom	1	0	856	prix fix menu	1	0
804	mail	0	1	857	water incastro	1	0
805	recept procesa	0	1	858	none	1	0
806	lobbi area	0	1	859	hotel solac	1	0
807	sweep meadow	1	0	860	sky	1	0
808	fix	1	0	861	countri person	0	1
809	new build	0	1	862	breakfast servic	1	0
810	wonder view	1	0	863	turquois	1	0
811	good cereal	1	0	864	adventur	1	0
812	atm	1	0	865	huge problem	0	1
813	mirador lo	1	0	866	beautifull bathroom	1	0
814	attract room	1	0	867	half	0	1
815	total silenc	0	1	868	third floor	1	0
816	love view	1	0	869	refresh chang	1	0
817	great bedroom	1	0	870	balanc	0	1
818	daughter susan	1	0	871	like crew	1	0
819	champagn	1	0	872	crisp	1	0
820	albergu latino	1	0	873	paint	0	1
821	beautfiul	1	0	874	ie	0	1
822	courtyard	1	0	875	junior suit	1	0
823	transmitt	1	0	876	world class	0	1
824	affair	1	0	877	pass	1	0
825	strang thing	0	1	878	latter	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
879	tranquil spot	1	0	933	cart	0	1
880	dessert	1	0	934	argentinian agenc	1	0
881	environ	0	1	935	colono del sur mirador	1	0
882	garden suffer	0	1	936	batteri	0	1
883	littl cheaper	1	0	937	proprietor	1	0
884	lago llanquihu	1	0	938	plu side	1	0
885	spectacular place	1	0	939	overview	1	0
886	delici piec	1	0	940	lakesid road	0	1
887	bad:1-	1	0	941	bright	1	0
888	cold lobbi	0	1	942	poor qualiti	0	1
889	grill	1	0	943	horseback	1	0
890	larg place	1	0	944	beauti bay	1	0
891	smallish side	0	1	945	tourist citi	0	1
892	park system	0	1	946	sink	0	1
893	coffe smell	0	1	947	nice pool	1	0
894	clerk	0	1	948	man	0	1
895	din room	1	0	949	member	0	1
896	hot food	1	0	950	griddl	1	0
897	littl warm	1	0	951	tree	1	0
898	staurant	1	0	952	period	1	0
899	rustic decor	1	0	953	sea	1	0
900	venu	1	0	954	difficult balanc	0	1
901	great littl boutiqu hotel	1	0	955	problem sir	1	0
902	accompani	1	0	956	tingl	1	0
903	snowcap	1	0	957	tv remot	0	1
904	month	1	0	958	crazi low peso	0	1
905	usual basic	0	1	959	cafeteria	1	0
906	worth visit	1	0	960	habit	1	0
907	spong	0	1	961	convient	1	0
908	basketbal	1	0	962	main point	1	0
909	bourdain	0	1	963	gem	1	0
910	pool dirti	0	1	964	colour	1	0
911	nice place	1	0	965	transport i	1	0
912	tea bag	0	1	966	april	1	0
913	hotel shouldthen	0	1	967	summer	0	1
914	signal right	1	0	968	del sur i	0	1
915	free upgrad	1	0	969	ball point	0	1
916	remot	0	1	970	bustle2-	1	0
917	euro	1	0	971	curios	1	0
918	expens option	1	0	972	lift	1	0
919	good place	1	0	973	littl dark	0	1
920	excel restaur	1	0	974	design	1	0
921	avarag hotel	1	0	975	welcom staff	1	0
922	featur	1	0	976	makeovertini	0	1
923	loaf bread	1	0	977	slow toaster	0	1
924	decent hostal	1	0	978	romant getaway	1	0
925	much nicer	1	0	979	spa excurs	1	0
926	board	1	0	980	doubl bed	1	0
927	hardwood	1	0	981	treatment	0	1
928	clean hotel	1	0	982	front desk clerk	0	1
929	influenc	1	0	983	wand	0	1
930	pen	0	1	984	agent	1	0
931	good news	1	0	985	singl channel	0	1
932	nice grill	1	0	986	rate facil	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
987	tradit hotel	1	0	1040	mile	1	0
988	pub	1	0	1041	ozorno vulcano	1	0
989	tour itinerari	0	1	1042	offic	1	0
990	microwav	1	0	1043	terrif view	1	0
991	afternoon shower water	0	1	1044	much fun	1	0
992	front-desk servic	1	0	1045	entir apart	0	1
993	fire place	1	0	1046	salto	1	0
994	tripit	1	0	1047	folk	1	0
995	hostel style accommod	1	0	1048	attempt	1	0
996	great kitchen	1	0	1049	comfi feel	1	0
997	neverend maze	1	0	1050	reall hit	0	1
998	maid	0	1	1051	rail	0	1
999	boat trip	1	0	1052	good reason	1	0
1000	el	0	1	1053	quiet neighbourhood	1	0
1001	hotel staff member	0	1	1054	allergi	0	1
1002	pace	0	1	1055	break	0	1
1003	meter	1	0	1056	great common room	1	0
1004	sauna	1	0	1057	desk servic	1	0
1005	neighbourhood	1	0	1058	broadband	1	0
1006	hamburg	1	0	1059	gran hotel	0	1
1007	junk	0	1	1060	soap	0	1
1008	immun reaction	1	0	1061	shirt	0	1
1009	fear	0	1	1062	hidden gem	1	0
1010	superb set	1	0	1063	bustl	1	0
1011	top-qual hotel	1	0	1064	modern take	1	0
1012	sun set	1	0	1065	welcom warm	1	0
1013	main strip	1	0	1066	everyon	1	0
1014	team	0	1	1067	stare	0	1
1015	specif health	0	1	1068	beach servic	0	1
1016	anyon i	1	0	1069	corp	1	0
1017	nook	1	0	1070	bad experi	0	1
1018	cosi environ	0	1	1071	local bu	1	0
1019	conveni i	1	0	1072	construct work	1	0
1020	espresso	1	0	1073	lake villarica	1	0
1021	breakfast room	1	0	1074	origin room	1	0
1022	abov averag	1	0	1075	hotel i	1	0
1023	vet	0	1	1076	right door	1	0
1024	whole area	1	0	1077	u s	0	1
1025	principado	0	1	1078	al mar	0	1
1026	construct	1	0	1079	splurg	1	0
1027	standard room	1	0	1080	rariti	1	0
1028	curios i	1	0	1081	hotel guest	0	1
1029	privat car	1	0	1082	blade	1	0
1030	europ	0	1	1083	tasti food	1	0
1031	mar	0	1	1084	great surpris	1	0
1032	tierra	1	0	1085	number	1	0
1033	everywherecon	0	1	1086	except small hotel	1	0
1034	ridicul sum	0	1	1087	onli regret	0	1
1035	pan	1	0	1088	truck	1	0
1036	convers	1	0	1089	terribl servic	0	1
1037	same price	0	1	1090	full access	1	0
1038	cushi robe	1	0	1091	liucuria	1	0
1039	decis	0	1	1092	live style	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1093	huge spread	1	0	1146	live area	1	0
1094	nice accomod	1	0	1147	updat	0	1
1095	fat man	0	1	1148	big bonu	1	0
1096	afternoon swim	1	0	1149	love surpris	1	0
1097	line	1	0	1150	blooper	1	0
1098	short stroll	1	0	1151	carpet	0	1
1099	i pression	1	0	1152	fantast view	1	0
1100	decent view	1	0	1153	patagonia	1	0
1101	actual hotel	1	0	1154	treat	0	1
1102	cute kitchen	1	0	1155	season	1	0
1103	way i	1	0	1156	inform	1	0
1104	opinion	0	1	1157	top floor	1	0
1105	wast	1	0	1158	comput room	1	0
1106	littl like home	1	0	1159	simpl argentinean breakfast	1	0
1107	great shower	1	0	1160	mistak	0	1
1108	mimibar	0	1	1161	procesa	0	1
1109	town centr	1	0	1162	meadow	1	0
1110	trip advisor review team	0	1	1163	botiqu hotel	1	0
1111	mofo	0	1	1164	favorit space	1	0
1112	hotel right	1	0	1165	amaz hotel	1	0
1113	background music	1	0	1166	pleasant staff	1	0
1114	noisi centr	1	0	1167	bed linen	1	0
1115	incastro	1	0	1168	comment	0	1
1116	comfort place	1	0	1169	unforgett vacat	1	0
1117	porridg	1	0	1170	place i	1	0
1118	great space	1	0	1171	good experi	1	0
1119	cristian	0	1	1172	crockeri	0	1
1120	tour group	0	1	1173	own transport	1	0
1121	expens shop	0	1	1174	reclin	1	0
1122	peac corp volunt	1	0	1175	lakefre wifi	0	1
1123	good sign	1	0	1176	ingredi	1	0
1124	beauti hotel	1	0	1177	towel rack	1	0
1125	latino	1	0	1178	ground level	0	1
1126	main thing	1	0	1179	posit side	1	0
1127	staff i	1	0	1180	off-season discount	1	0
1128	hair	0	1	1181	coffe tabl	1	0
1129	delight hotel	1	0	1182	drawer	0	1
1130	expens side	1	0	1183	class	0	1
1131	la cart	0	1	1184	canopi platform	1	0
1132	airport	0	1	1185	rush	0	1
1133	bait	0	1	1186	small kitchen	1	0
1134	built-in bookcas	1	0	1187	by	1	0
1135	hanger	0	1	1188	complementari bottl	1	0
1136	fabul place	1	0	1189	interest hotel	1	0
1137	indoor pool	1	0	1190	cigarett	0	1
1138	cascada	1	0	1191	lake access	0	1
1139	fetal posit	0	1	1192	lech	1	0
1140	enorm fire	1	0	1193	wonder staff	1	0
1141	eye	0	1	1194	hostel2-	0	1
1142	chocol	1	0	1195	bottom line	1	0
1143	don vicent	1	0	1196	lake disappoint	1	0
1144	fair price	1	0	1197	steep hill	1	0
1145	next room	0	1	1198	big hole	0	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1199	wonder room	1	0	1250	warmth	0	1
1200	good dine	1	0	1251	posit part	1	0
1201	charm boutiqu hotel	1	0	1252	drop	1	0
1202	bad face	0	1	1253	practic i	0	1
1203	cinnamon	1	0	1254	kid	0	1
1204	veri end	1	0	1255	perfum	1	0
1205	pie	0	1	1256	thank	1	0
1206	bellhop	1	0	1257	snow	1	0
1207	de date	0	1	1258	quaint old hous	1	0
1208				1259	fantast landscap	1	0
1209	prix beauti place	0	1	1260	cheescak	0	1
1210	hotel room	1	0	1261	twin room	0	1
1211	delici home	1	0	1262	name	0	1
1212	day trip	1	0	1263	neighboor	1	0
1213	other	1	0	1264	ground floor	1	0
1214	first coupl	0	1	1265	peac chalet-styl hotel	1	0
1215	custom servic	1	0	1266	renov palafito	1	0
1216	owner jorg	1	0	1267	beauti build	1	0
1217	pictur	1	0	1268	heard	1	0
1218	spotlessli clean	1	0	1269	mind	1	0
1219	pasta	1	0	1270	swim pool	1	0
1220	huge disappoint	0	1	1271	muesli	1	0
1221	mani hotel	1	0	1272	hoop	1	0
1222	manag problem	0	1	1273	street nois	0	1
1223	inner	1	0	1274	wireless	0	1
1224	router	1	0	1275	crispi	1	0
1225	polici	1	0	1276	beauti view	1	0
1226	cours affair	1	0	1277	dirti crockeri	0	1
1227	crew	1	0	1278	shouldthen	0	1
1228				1279	router transmitt	1	0
1229	spartan	1	0	1280	wooden furnish	1	0
	hallway	0	1				
1230	shelf space	0	1	1281	extent	0	1
1231	welcom drink	1	0	1282	great servic	1	0
1232	pto	1	0	1283	materi	1	0
1233	locat right	1	0	1284	receipt	1	0
1234				1285	huge comfort bed	1	0
	crosser	1	0				
1235	brazilian breakfast	1	0	1286	trip advisor	1	0
1236				1287	snow-cov volcano	1	0
	crazi cab	1	0				
1237	internet termin	1	0	1288	restaur food	1	0
1238	slope	1	0	1289	pto montt	1	0
1239	qualiti linen	1	0	1290	colon	0	1
1240	great addit	1	0	1291	quiet littl villag	1	0
1241	usa	1	0	1292	good mall	1	0
1242	suffer	0	1	1293	present	1	0
1243	cigarett smoke	0	1	1294	left wing	1	0
1244	second floor	0	1	1295	hip	1	0
1245	applianc	1	0	1296	paper thin	1	0
1246	botan center	1	0	1297	hostal	1	0
1247	news	1	0	1298	colectivo	0	1
1248	cheap hostel2-	0	1	1299	warm weather	1	0
1249	rap	0	1	1300	long afternoon	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1301	immigr	1	0	1302	spectacular hotel	1	0

### C.3. Hasil *Object-based Opinion Mining Hotel (Frequent Object dengan Filter)*

Berikut ini merupakan hasil output dari *object-based opinion mining* data review hotel dengan pendekatan *frequent object* yang terfilter sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	hotel	1	hotel	20	rate	6	1
2	room	2	room	21	lobbi	6	3
3	stay	3	stay	22	coffe	7	2
4	breakfast	4	breakfast	23	window	3	4
5	lake	5	lake	24	bedroom	9	1
6	staff	6	staff	25	dinner	3	2
7	view	7	view	26	park	5	2
8	place	8	place	27	bread	5	2
9	servic	9	servic	28	recept	5	2
10	locat	10	locat	29	space	5	1
11	water	11	water	30	owner	5	2
12	bed	12	bed	31	tv	3	2
13	pool	13	pool	32	fruit	8	0
14	food	14	food	33	shower	4	3
15	bathroom	15	bathroom	34	kitchen	8	0
16	wifi	16	wifi	35	casino	7	1
17	restaur	17	restaur	36	garden	6	1
18	price	18	price	37	balconi	5	1
19	shop	19	shop	38	fireplac	5	0

### C.4. Hasil *Object-based Opinion Mining Hotel (Frequent Object and Infrequent Object dengan Filter)*

Berikut ini merupakan hasil output dari *object-based opinion mining* data review hotel dengan pendekatan *frequent object* dan *infrequent object* yang terfilter sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	hotel	107	37	13	pool	10	7
2	room	104	36	14	food	17	2
3	stay	38	13	15	bathroom	14	4
4	breakfast	43	13	16	wifi	11	3
5	lake	41	4	17	restaur	12	4
6	staff	41	7	18	price	5	3
7	view	43	2	19	lobbi	6	3
8	place	34	9	20	coffe	7	2
9	servic	18	7	21	rate	6	1
10	locat	21	1	22	shop	4	3
11	water	11	8	23	window	4	3
12	bed	18	0	24	bedroom	8	2

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
25	dinner	2	3	83	beach	0	2
26	park	5	2	84	yogurt	1	0
27	tour	4	3	85	refriger	1	1
28	recept	5	2	86	meat	1	0
29	bread	5	2	87	receptionist	0	1
30	owner	5	2	88	menu	1	0
31	tv	3	2	89	laundri	1	1
32	shower	3	4	90	entranc	1	0
33	internet	3	4	91	bathtub	1	1
34	kitchen	8	0	92	manag	0	2
35	fruit	8	0	93	ceil	0	1
36	space	5	1	94	bike	0	1
37	casino	6	2	95	phone	0	2
38	garden	6	1	96	parti	0	1
39	access	4	3	97	cost	0	2
40	hot water	3	4	98	tenni court	0	1
41	dine	6	0	99	anim	1	1
42	ground	5	1	100	flower	2	0
43	balconi	5	1	101	sheet	1	1
44	fireplac	5	0	102	comput	1	1
45	egg	5	1	103	breakfast area	1	1
46	lunch	1	0	104	match	0	2
47	chair	4	1	105	host	2	0
48	chees	4	0	106	toaster	1	1
49	loung	4	1	107	cleanli	2	0
50	decor	4	0	108	heater	0	2
51	tabl	3	0	109	elev	1	1
52	build	3	1	110	stove	2	0
53	light	3	2	111	driveway	2	0
54	suit	3	0	112	iron	1	1
55	qualiti	2	2	113	chef	1	0
56	spa	5	0	114	gym	1	1
57	towel	2	2	115	mattress	2	0
58	driver	3	0	116	internet access	1	1
59	downtown	1	1	117	lakesid	1	1
60	meal	3	0	118	pot	1	0
61	bar	2	1	119	din room	1	0
62	cabin	3	0	120	soap	0	1
63	cake	3	0	121	espresso	1	0
64	jam	1	2	122	balanc	0	1
65	bath	2	2	123	spong	0	1
66	tea	3	1	124	environ	0	1
67	ham	3	0	125	croissant	1	0
68	court	0	2	126	front-desk servic	1	0
69	site	1	0	127	park system	0	1
70	sofa	1	0	128	comput room	1	0
71	cereal	3	0	129	sauna	1	0
72	lake view	3	0	130	cookwar	1	0
73	spot	2	0	131	tv guid	0	1
74	dine room	2	0	132	pasta	1	0
75	cafe	1	0	133	microwav	1	0
76	room servic	3	0	134	wifi signal	1	0
77	heat	2	1	135	fridg	1	0
78	mall	2	0	136	squash court	0	1
79	wine	3	0	137	drawer	0	1
80	terrac	2	0	138	hamburg	1	0
81	live room	2	0	139	doorman	1	0
82	accomod	2	0	140	wireless	0	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
141	seat area	1	0	169	carpet	0	1
142	easi chair	1	0	170	beach servic	0	1
143	cafeteria	1	0	171	hanger	0	1
144	dessert	1	0	172	lift	1	0
145	hall	1	0	173	waiter	1	0
146	din	1	0	174	coffe tabl	1	0
147	rec room	1	0	175	crockeri	0	1
148	couch	1	0	176	shelf space	0	1
149	juic	1	0	177	facil	1	0
150	coffe maker	1	0	178	standard room	1	0
151	lake access	0	1	179	atmospher	1	0
152	paint	0	1	180	game	1	0
153	neighbourhood	1	0	181	landscap	1	0
154	airport	0	1	182	pen	0	1
155	local	1	0	170	beach servic	0	1
156	deterg	0	1	171	hanger	0	1
157	broadband	1	0	172	lift	1	0
158	tv remot	0	1	173	waiter	1	0
159	furnitur	1	0	174	coffe tabl	1	0
160	room price	1	0	175	crockeri	0	1
161	hot food	1	0	176	shelf space	0	1
162	park area	1	0	177	facil	1	0
163	mini-bar	1	0	178	standard room	1	0
164	equip	0	1	179	atmospher	1	0
165	bookcas	1	0	180	game	1	0
166	curtain	0	1	181	landscap	1	0
167	lobbi area	0	1	182	pen	0	1
168	design	1	0				

### C.5. Hasil *Object-based Opinion Mining* Hotel (*Filter Object*)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review hotel dengan pendekatan *filter object* sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	hotel	108	36	22	shop	4	3
2	room	102	38	23	window	3	4
3	stay	39	12	24	park	5	2
4	breakfast	43	13	25	bedroom	9	1
5	lake	40	5	26	dinner	3	2
6	staff	42	6	27	reception	5	2
7	view	42	3	28	tour	3	4
8	place	33	10	29	bread	4	3
9	service	17	8	30	shower	4	3
10	location	20	2	31	owner	5	2
11	water	12	7	32	casino	6	2
12	bed	16	2	33	space	5	1
13	food	17	2	34	tv	3	2
14	pool	11	6	35	internet	3	4
15	bathroom	15	3	36	fruit	6	2
16	restaurant	12	4	37	kitchen	8	0
17	wifi	12	2	38	hot water	3	4
18	price	4	4	39	ground	5	1
19	rate	6	1	40	balcony	4	2
20	lobby	7	2	41	fireplace	5	0
21	coffee	4	5	42	dining	5	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
43	access	4	3	101	receptionist	0	1
44	garden	6	1	102	standard room	1	1
45	chair	4	1	103	bike	0	1
46	egg	4	2	104	bathtub	1	1
47	lunch	1	0	105	computer	1	1
48	towel	2	2	106	mattress	2	0
49	cheese	3	1	107	tennis court	1	0
50	building	3	1	108	party	0	1
51	spa	5	0	109	driveway	2	0
52	lounge	3	2	110	chef	1	0
53	suite	3	0	111	living room	2	0
54	decor	4	0	112	menu	1	0
55	lake view	5	0	113	heater	0	2
56	table	3	0	114	gym	1	1
57	parking	3	1	115	sheets	1	1
58	suit	3	0	116	ceiling	0	1
59	quality	1	3	117	breakfast area	1	1
60	light	3	2	118	elevator	2	0
61	jam	0	3	119	refrigerator	1	1
62	tea	2	2	120	host	2	0
63	cabin	3	0	121	wifi signal	2	0
64	shopping	2	0	122	gardens	2	0
65	bath	2	2	123	beach	0	2
66	dining room	2	1	124	stove	2	0
67	driver	3	0	125	phone	1	1
68	cake	2	1	126	cost	0	2
69	meal	3	0	127	tv remote	0	1
70	bar	2	1	128	shelf space	0	1
71	downtown	1	1	129	room price	1	0
72	spot	2	0	130	wireless	0	1
73	site	1	0	131	public space	1	0
74	room service	3	0	132	lounge chairs	0	1
75	sofa	1	0	133	bike loan	0	1
76	cereal	2	1	134	cafeteria	1	0
77	mall	2	0	135	airport	0	1
78	court	0	2	136	microwave	1	0
79	ham	2	1	137	brunch	1	0
80	cafe	1	0	138	beach service	0	1
81	heating	2	1	139	easy chair	1	0
82	wine	3	0	140	pen	0	1
83	public room	1	1	141	pot	1	0
84	hot spring	2	0	142	croissant	1	0
85	lobby area	2	0	143	food options	0	1
86	internet access	1	1	144	ping pong table	1	0
87	yogurt	1	0	145	hamburger	1	0
88	matches	0	2	146	AC	0	1
89	entrance	1	0	147	dining area	1	0
90	flower	2	0	148	parking area	1	0
91	meat	1	0	149	twin	0	1
92	accomodation	2	0	150	computer room	1	0
93	terrace	2	0	151	common area	1	0
94	sheet	1	1	152	localization	1	0
95	lakeside	2	0	153	curtain	0	1
96	laundry	1	1	154	parking system	0	1
97	iron	2	0	155	sponge	0	1
98	toaster	1	1	156	hot food	1	0
99	animal	1	1	157	front desk service	1	0
100	cleanliness	2	0	158	carpet	0	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
159	management	0	1	185	balance	1	0
160	waiter	1	0	186	manager	0	1
161	furniture	1	0	187	coffee table	1	0
162	landscape	1	0	188	rec room	1	0
163	heated swimming pool	1	0	189	facility	1	0
164	espresso	1	0	190	coffee maker	1	0
165	design	1	0	191	equipment	1	0
166	paint	0	1	192	bookcase	1	0
167	dinning	1	0	193	basketball hoop	1	0
168	cookware	1	0	194	bar area	1	0
169	hanger	0	1	195	sauna	1	0
170	bus station	1	0	196	neighbourhood	1	0
171	lift	1	0	197	couch	1	0
172	dessert	1	0	198	double room	1	0
173	squash court	0	1	199	lake access	0	1
174	drawer	0	1	200	broadband	1	0
175	hall	1	0	201	decoration	1	0
176	mini-bar	1	0	202	detergent	0	1
177	mountain view	1	0	203	tv guide	0	1
178	fridge	1	0	204	crockery	0	1
179	lounge chair	0	1	205	seating area	1	0
180	game	1	0	206	environment	1	0
181	sitting room	1	0	207	pasta	1	0
182	juice	1	0	208	dinning room	1	0
183	soap	0	1	209	front-desk service	1	0
184	doorman	1	0	210	atmosphere	1	0

#### C.6. Hasil *Object-based Opinion Mining* Reatoran (*Frequent Object*)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review restoran dengan pendekatan *frequent object* sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	food	48	9	23	atmospher	12	0
2	place	37	8	24	fish	7	1
3	restaur	25	10	25	view	10	0
4	servic	27	4	26	peopl	5	3
5	menu	16	4	27	beer	8	0
6	wine	18	1	28	locat	5	1
7	chile	9	3	29	coffe	6	2
8	puerto	13	1	30	bit	3	5
9	recommend	16	0	31	hotel	6	2
10	night	11	2	32	steak	8	0
11	vara	12	2	33	tast	6	1
12	pizza	12	2	34	coupl	3	1
13	staff	16	0	35	dinner	3	1
14	seafood	14	1	36	room	4	3
15	meal	13	1	37	qualiti	6	2
16	day	8	3	38	owner	4	0
17	dish	12	0	39	area	7	1
18	salad	8	2	40	experi	5	2
19	tabl	7	2	41	choic	5	1
20	meat	9	2	42	town	4	0
21	price	10	3	43	chees	3	2
22	time	5	3	44	friend	4	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
45	hous	4	1	61	flavour	5	0
46	lot	5	1	62	pasta	5	0
47	someth	4	3	63	cours	2	2
48	reserv	1	2	64	pucon	4	0
49	week	4	0	65	lunch	4	0
50	sauc	4	1	66	great place	5	0
51	chocol	6	0	67	month	1	2
52	la	5	0	68	kind	5	0
53	bread	4	1	69	bland	0	5
54	side	4	1	70	chef	4	1
55	cream	6	0	71	lake	5	0
56	breakfast	4	0	72	fact	4	1
57	salmon	2	2	73	bar	3	2
58	sandwich	3	0	74	way	3	2
59	select	3	0	75	drink	2	1
60	decor	5	0				

#### C.7. Hasil *Object-based Opinion Mining* Restoran (*frequent and Infrequent Object*)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review restoran dengan pendekatan *frequent dan infrequent object* sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	food	48	9	33	differ	5	1
2	place	37	8	34	help	7	3
3	restaur	25	10	35	peopl	6	2
4	servic	26	5	36	local	6	1
5	menu	15	5	37	amaz	9	0
6	wine	19	1	38	coffe	6	2
7	one	12	8	39	love	9	0
8	chile	9	3	40	locat	4	2
9	puerto	13	1	41	stay	4	0
10	recommend	16	0	42	wonder	9	0
11	tri	11	2	43	coupl	3	1
12	eat	11	5	44	bit	3	5
13	pizza	11	3	45	hotel	6	2
14	night	11	2	46	tast	6	1
15	vara	12	2	47	beer	8	0
16	staff	16	0	48	steak	8	0
17	seafood	13	2	49	room	4	3
18	meal	12	2	50	owner	4	0
19	dish	12	0	51	qualiti	6	2
20	day	8	3	52	serv	2	4
21	salad	8	2	53	dinner	3	1
22	tabl	7	2	54	choic	5	1
23	price	10	3	55	english	4	1
24	meat	9	2	56	experi	5	2
25	time	5	3	57	area	7	1
26	atmospher	11	1	58	feel	6	1
27	reason	11	0	59	friend	4	0
28	order	5	3	60	someth	4	3
29	even	5	2	61	look	4	1
30	chilean	6	2	62	hous	4	1
31	fish	7	1	63	ate	5	0
32	view	10	0	64	chees	3	2

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
65	town	3	1	124	cake	3	0
66	lot	5	1	125	bottl	3	1
67	dine	6	0	126	great food	4	0
68	reserv	1	2	127	montt	2	0
69	breakfast	4	0	128	speak	1	1
70	la	5	0	129	sure	0	2
71	side	4	1	130	surpris	1	1
72	outsid	4	1	131	thing	2	2
73	beauti	6	0	132	bake	4	0
74	south	3	1	133	castro	2	0
75	ice	5	1	134	main	3	1
76	chocol	5	1	135	start	2	1
77	salmon	2	2	136	pisco	3	1
78	sauc	4	1	137	mix	2	0
79	top	4	1	138	cevich	4	0
80	bread	4	1	139	waiter	2	2
81	cream	6	0	140	dessert	2	0
82	week	4	0	141	corner	2	2
83	attent	6	0	142	style	4	0
84	shop	4	0	143	visit	3	0
85	flavour	5	0	144	sign	1	1
86	bland	0	5	145	argentina	2	0
87	person	4	1	146	portion	3	0
88	cours	2	2	147	crab	2	0
89	fact	3	2	148	xic dali	1	0
90	great place	5	0	149	life	2	1
91	sandwich	3	0	150	name	2	0
92	select	3	0	151	thank	2	0
93	lunch	4	0	152	sour	2	1
94	spend	4	0	153	play	2	0
95	pucon	4	0	154	cold	0	3
96	perfect	4	0	155	set	2	0
97	vegetarian	3	0	156	entre	3	0
98	plu	1	0	157	rush	1	1
99	month	1	2	158	walk	2	0
100	hope	2	1	159	cuisin	3	0
101	travel	3	0	160	crepe	2	0
102	lake	5	0	161	relax	2	0
103	chef	4	1	162	fruitillar	1	1
104	kind	5	0	163	everyth	3	0
105	bar	3	2	164	world	2	0
106	way	3	2	165	prepar	1	1
107	pasta	5	0	166	dali	1	0
108	decor	5	0	167	chilo	2	0
109	kitchen	2	2	168	right	3	0
110	pricey	0	4	169	tomato	1	0
111	chicken	1	3	170	homemad	3	0
112	special	2	2	171	gringa	2	0
113	husband	2	1	172	restor	3	0
114	plate	2	1	173	mean	1	0
115	beef	3	0	174	starter	1	0
116	farm	4	0	175	custom	1	2
117	juic	3	0	176	base	1	1
118	home	4	0	177	alessandro	3	0
119	insid	3	1	178	sort	3	0
120	drink	2	1	179	ambianc	3	0
121	ingredi	3	1	180	imagin	3	0
122	part	3	0	181	list	2	0
123	review	1	3	182	cassi	2	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
183	nice staff	3	0	242	season	1	1
184	cozi	3	0	243	risotto	2	0
185	del	1	0	244	book	1	0
186	pastri	2	1	245	outdoor	2	0
187	ale	3	0	246	addit	2	0
188	soup	2	0	247	idea	2	0
189	burger	1	2	248	none	1	1
190	tea	2	0	249	coff	0	1
191	empanada	3	0	250	pork	2	0
192	level	1	1	251	deer	2	0
193	waitress	2	1	252	rest	1	0
194	except	1	0	253	cut	2	0
195	potato	2	0	254	wifi	1	0
196	mushroom	2	0	255	street	1	0
197	flavor	1	0	256	same time	1	1
198	park	2	1	257	water	0	1
199	smoke	1	1	258	amount	1	1
200	rice	1	1	259	easi	2	0
201	block	2	0	260	good beer	2	0
202	manag	2	0	261	pita	1	0
203	wow	1	0	262	homey	2	0
204	excel food	2	0	263	door	0	1
205	abalon	2	0	264	variat	2	0
206	cooki	2	0	265	spectacular	1	0
207	hostel	2	0	266	platter	1	1
208	tv	1	1	267	wait	2	0
209	american	1	0	268	complaint	0	2
210	juici	0	1	269	citi	2	0
211	crisp	2	0	270	roast	2	0
212	goat	1	0	271	bill	2	0
213	second night	1	1	272	pleas	1	1
214	flat	0	1	273	congrio	2	0
215	advisor	0	1	274	ham	1	0
216	onli complaint	0	2	275	trip	0	1
217	frutillar	1	0	276	pio	2	0
218	la gringa	2	0	277	music	2	0
219	thai	0	2	278	good food	2	0
220	good english	2	0	279	hour	0	1
221	state	1	1	280	fresh fish	2	0
222	serrano	1	0	281	type	2	0
223	puerto vara	1	1	282	marion	1	0
224	trip advisor	0	1	283	chose	2	0
225	plenti	1	0	284	environ	2	0
226	good restaur	2	0	285	onion	1	1
227	standard	0	2	286	fri	1	1
228	peso	2	0	287	deal	2	0
229	boutiqu	1	0	288	da	2	0
230	brew	1	1	289	stuff	1	0
231	gnocchi	1	1	290	variети	2	0
232	touch	2	0	291	america	2	0
233	xime	2	0	292	chicken burger	0	2
234	fruit	1	0	293	cocktail	2	0
235	minut	0	2	294	fire	1	0
236	toast	2	0	295	catalan	1	0
237	arrebol	1	0	296	nice place	2	0
238	mountain	2	0	297	design	2	0
239	server	1	1	298	nice view	2	0
240	dine room	1	0	299	bite	2	0
241	Balai	1	0	300	front	1	1

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
301	video	2	0	353	nice bottl	1	0
302	great view	2	0	354	pan amasado	1	0
303	barbecu	1	0	355	melt-in-your-mouth deer	1	0
304	coffe hous	1	1	356	south america	1	0
305	eleg	1	1	357	delici food	1	0
306	pastagonia	2	0	358	row	0	1
307	cappuccino	0	1	359	pc	1	0
308	jewel	1	0	360	germani	1	0
309	marina	1	0	361	easi dish	1	0
310	vegetarian restaur	1	0	362	di mare	1	0
311	enough fish	0	1	363	seafood platter	1	0
312	wine room	1	0	364	scarf	1	0
313	organ green salad	1	0	365	staff somehow	1	0
314	place kind	1	0	366	someth i	1	0
315	entre easi	1	0	367	edg	1	0
316	seat	1	0	368	funni look	0	1
317	heart	1	0	369	omelet	1	0
318	expens side	0	1	370	homey decor	1	0
319	addict	0	1	371	friendli servic	1	0
320	simpl campfir	1	0	372	great atmospher	1	0
321	rain	1	0	373	la marzocco fb70	1	0
322	bad merlot	0	1	374	punta	1	0
323	llama	1	0	375	everyth right	1	0
324	hurri	0	1	376	outstand servic	1	0
325	marriott	1	0	377	tourist	1	0
326	cool outsid	0	1	378	beauti farm	1	0
327	wine bar area	1	0	379	tip	1	0
328	co-ordin	1	0	380	much tabl	1	0
329	marzocco	1	0	381	attend	1	0
330	brava	1	0	382	guest	1	0
331	main dish	1	0	383	cappuccino machin	0	1
332	beer list	1	0	384	wifi-	1	0
333	expens dish	1	0	385	vulcano	1	0
334	hole	1	0	386	valencian rice	1	0
335	last coupl	0	1	387	tour	1	0
336	snob	0	1	388	currenc	0	1
337	world clase	1	0	389	snow-cap volcano	1	0
338	market idea	1	0	390	plancha	1	0
339	vine select	1	0	391	time i	0	1
340	weisserhau hotel	1	0	392	botero anonimo	1	0
341	nectar	0	1	393	speak famili	1	0
342	lunch menu	1	0	394	anybodi	1	0
343	guess everyon	0	1	395	french style	1	0
344	foie	1	0	396	carrot soup	1	0
345	great pasta	1	0	397	detail	1	0
346	heat	0	1	398	tomato flavour	1	0
347	high side	1	0	399	outdoor seat	1	0
348	hotel marina del fuy	1	0	400	wednesday	1	0
349	papillot	1	0	401	offens thing	0	1
350	announc	1	0	402	case	1	0
351	botero	1	0	403	good steak	1	0
352	famili	1	0	404	internet	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
405	yogurt mint sauc	1	0	455	thai restaur	0	1
406	olla	1	0	456	pepper	1	0
407	spain	1	0	457	cotta	1	0
408	ice cream	1	0	458	sorpris	1	0
409	drag	1	0	459	properti	1	0
410	children	1	0	460	hell	1	0
411	big mall	1	0	461	good capuccino-free wifi-	1	0
412	comfort restaur	1	0	462	great cuisin	1	0
413	brit	1	0	463	cod cheek	1	0
414	success	1	0	464	rancho espatapajaro	1	0
415	earthi	1	0	465	life span	0	1
416	prefer	1	0	466	nice chilean wine	1	0
417	previou review	0	1	467	waitor	0	1
418	oil	1	0	468	limit	0	1
419	counter	1	0	469	window	1	0
420	authent cultur chilean	1	0	470	good qualiti	1	0
421	culinari experi	0	1	471	guid	1	0
422	rum	1	0	472	delici experi	1	0
423	tea hous	1	0	473	macha	1	0
424	cuppa	1	0	474	mushroom risotto	1	0
425	centr	1	0	475	smile	1	0
426	off day	1	0	476	wasabi mayo	1	0
427	wonder food	1	0	477	upstair	1	0
428	vegetable-light food	1	0	478	mech	1	0
429	cheek	1	0	479	champagn	1	0
430	rancho	1	0	480	money	1	0
431	local currenc	0	1	481	jazz	1	0
432	lake view	1	0	482	hotel eateri	0	1
433	congrio plancha	1	0	483	tonight	1	0
434	homemad pasta	1	0	484	splurg	1	0
435	friendli restaur	1	0	485	daughter	0	1
436	arturo prat	1	0	486	good meal	1	0
437	host	1	0	487	pto	1	0
438	cogrio	1	0	488	crust	1	0
439	blackboard	0	1	489	cherri	1	0
440	good artesen beer	1	0	490	map locat	0	1
441	excel servic	1	0	491	tasti side	1	0
442	person care	1	0	492	combin	1	0
443	artefact	1	0	493	local bake salmon dish	1	0
444	smoke chees	1	0	494	compet chef	1	0
445	catalunya	1	0	495	pretti decent final number	1	0
446	picki	0	1	496	new name	1	0
447	midnight	0	1	497	amaz fresh juic	1	0
448	wonder restaur	1	0	498	hearti breakfast	1	0
449	patti	0	1	499	hood	1	0
450	ride	1	0	500	raw	1	0
451	hotel staff	1	0	501	creami	1	0
452	amasado	1	0	502	big room	0	1
453	red wine	1	0	503	casual night	1	0
454	seafood establish	0	1	504	good select	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
505	unfilt ale	1	0	530	panna cotta	1	0
506	subject	1	0	531	littl pricey	0	1
507	tripadvisor	0	1	532	pleac	1	0
508	la marca	1	0	533	tuna steak i	1	0
509	appet	1	0	534	mint	1	0
510	sushi	0	1	535	putanesca	0	1
511	crab sauc	1	0	536	quaint	1	0
512	maitr	1	0	537	chorizo	1	0
513	mind	1	0	538	glass	0	1
514	goodby	1	0	539	confus start	0	1
515	feet	1	0	540	vanilla	1	0
516	creativ mix	1	0	541	to-die-for-muffin	1	0
517	fresh marin	1	0	542	wheat toast	1	0
518	cellar	1	0	543	wine area	1	0
519	main drag	1	0	544	bread bake	1	0
520	sauerkraut	0	1	545	abierta	1	0
521	bife	1	0	546	downsid	0	1
522	barman	1	0	547	wonder bargain	1	0
523	plaza	0	1	548	jasmin	1	0
524	sanguch	1	0	549	femal maitr d	1	0
525	baguett	1	0	550	love southern town	1	0
526	huerta	1	0	551	nice atmospher	1	0
527	great restaur	1	0	552	lomito	1	0
528	torremolino	1	0	553	slab	0	1
529	crepe seafood	1	0	554	delici salad	1	0
530	panna cotta	1	0	555	manag chose	1	0
531	littl pricey	0	1	556	pud	1	0
532	pleac	1	0	557	meati	1	0
533	tuna steak i	1	0	558	wine bar	0	1
534	mint	1	0	559	support	1	0
505	unfilt ale	1	0	560	privileg locat	1	0
506	subject	1	0	561	path	0	1
507	tripadvisor	0	1	562	top-draw	1	0
508	la marca	1	0	563	shrimp look	1	0
509	appet	1	0	564	mood	0	1
510	sushi	0	1	565	miserli	0	1
511	crab sauc	1	0	566	pleasant experi	1	0
512	maitr	1	0	567	treat	1	0
513	mind	1	0	568	chutney	1	0
514	goodby	1	0	569	hood restaur	1	0
515	feet	1	0	570	huge plate	1	0
516	creativ mix	1	0	571	sport	1	0
517	fresh marin	1	0	572	chocol powder	1	0
518	cellar	1	0	573	delici coffe	1	0
519	main drag	1	0	574	fresh fruit	1	0
520	sauerkraut	0	1	575	slick	1	0
521	bife	1	0	576	tast risotto	1	0
522	barman	1	0	577	loung	1	0
523	plaza	0	1	578	veggi	1	0
524	sanguch	1	0	579	nice sidewalk	1	0
525	baguett	1	0	580	gustacion	1	0
526	huerta	1	0	581	good idea	1	0
527	great restaur	1	0	582	cala	1	0
528	torremolino	1	0	583	wine recommend	1	0
529	crepe seafood	1	0	584	pork i	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
585	cappuccino	0	1	641	firm	1	0
586	arrebol hotel	1	0	642	nacho	0	1
587	guess	0	1	643	beautiful area	1	0
588	style pizza	1	0	644	jewellri	1	0
589	desert	1	0	645	mayo	1	0
590	dine experi	1	0	646	somehow	1	0
591	delici bottl	1	0	647	product	1	0
592	fuy	1	0	648	opinnion	1	0
593	custard pud	1	0	649	eleg meal	0	1
594	chilean food	1	0	650	donkey	1	0
595	content	0	1	651	mall	1	0
596	great coffe	1	0	652	ice chocol	0	1
597	neighbor	1	0	653	bland chees	0	1
598	prat	1	0	654	contemporari decor	1	0
599	basil ice cream	1	0	655	servic co-ordin	1	0
600	attent servic	1	0	656	bean	1	0
601	poor lodg	1	0	657	tacki giant tv	0	1
602	bad mood	0	1	658	papa	1	0
603	top qualiti	0	1	659	vegetarian entra	1	0
604	spaghetti	1	0	660	good coffe	1	0
605	great servic	1	0	661	food prepar	1	0
606	beautifi pucon	1	0	662	c ]	1	0
607	favor	1	0	663	horribl pizza	0	1
608	regular hous	1	0	664	km	1	0
609	amaz food	1	0	665	person touch	1	0
610	emu	1	0	666	nois	1	0
611	good sign	1	0	667	instant coff	0	1
612	servic amaz	1	0	668	cakesMarion	1	0
613	chocol i	1	0	669	background	1	0
614	good german food	1	0	670	true jewel	1	0
615	fb70	1	0	671	savori seafood	1	0
616	tabl abierta	1	0	672	italian restaur	1	0
617	chilenen	1	0	673	district	1	0
618	wall	1	0	674	span	0	1
619	nice choic	1	0	675	pushi	0	1
620	pizza cala	1	0	676	person servic	1	0
621	favourit	0	1	677	volcano	1	0
622	textur	1	0	678	merluza	1	0
623	market	1	0	679	dissapoint	0	1
624	popular place	1	0	680	amaz deal	1	0
625	same tast	1	0	681	waitstaff	1	0
626	same street	1	0	682	amaz steak	1	0
627	refresh	0	1	683	flouri crust	1	0
628	excel seafood	1	0	684	secret ingredi	1	0
629	lamp	0	1	685	solid pizza	1	0
630	tab	1	0	686	italian food	0	1
631	convers	1	0	687	poncho	1	0
632	magnific place	1	0	688	marin	1	0
633	much veget content	0	1	689	nice food	1	0
634	anonimo	1	0	690	wonder place	1	0
635	mouth	0	1	691	invit	1	0
636	carbener wine	1	0	692	adult	1	0
637	meat i	1	0	693	make sure	0	1
638	perfect servic	1	0	694	fuego	1	0
639	flight	0	1	695	breakfast place	1	0
640	valu	1	0	696	aussi	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
697	pm	1	0	751	delici homemad	1	0
698	nice ancient hostel	1	0	752	chilean twist	1	0
699	coffe bean	1	0	753	cod	1	0
700	papa brava	1	0	754	lake district	1	0
701	cocina	1	0	755	heat lamp	0	1
702	focaccia	1	0	756	nice environ	1	0
703	chilenean	1	0	757	so-so	0	1
704	tourist dine	1	0	758	good meat	1	0
705	food snob	0	1	759	help host	1	0
706	number	1	0	760	parmesan	1	0
				761	blackboard sign	0	1
707	florentini	1	0	762	last bite	1	0
708	park attend	1	0	763	wednesday night	1	0
709	good dinner	1	0	764	wife	1	0
710	spree	1	0	765	strong tast	1	0
711	hearti	1	0	766	chilean poncho	1	0
712	clear day	1	0	767	chicken pita	1	0
713	twist	1	0	768	la olla	1	0
714	non-seafood food	1	0	769	kika	1	0
715	organ beef	1	0	770	recommend spend	1	0
716	espatapajaro	1	0	771	wish	1	0
717	upper area	0	1	772	celebr	1	0
718	panoram view	1	0	773	cafeteria	0	1
719	pate	1	0	774	cab	1	0
720	fresh seafood	1	0	775	thousand	1	0
721	next day	1	0	776	week day	1	0
722	amaz cream	1	0	777	pan	1	0
723	great detail	1	0	778	farm-styl room	1	0
724	vegetarian pizza	1	0	779	note	0	1
725	good valu	1	0	780	superb chilenean	1	0
726	effici staff	1	0	781	stuf	1	0
727	care	1	0	782	grill meat	0	1
728	art	1	0	783	g ]	1	0
729	mango chutney	1	0	784	panna	1	0
730	other seafood	1	0	785	te same chef	1	0
731	recip	1	0	786	beer ice cream	1	0
732	year	0	1	787	excel caesar	1	0
733	nice casual restaur	1	0	788	german bread	1	0
734	eateri	0	1	789	suck	0	1
735	breakfast menu	1	0	790	lemon	1	0
736	hot food	1	0	791	stingi serv	0	1
737	beefburg	0	1	792	qinoua	1	0
738	vibe	1	0	793	santiago	1	0
739	greasi	0	1	794	good choic	1	0
740	se cocina	1	0	795	streetlif	1	0
741	oil cocktail toast	1	0	796	serrano ham	1	0
742	stove	1	0	797	marca	1	0
743	fillet	1	0	798	littl bland	0	1
744	establish	0	1	799	joke	0	1
745	girlfriend	0	1	800	bargain	1	0
746	salami	1	0	801	congrio pio pio	1	0
747	shot	1	0	802	excel pisco	1	0
748	puconIt	1	0	803	whole time	1	0
749	caesar	1	0	804	staf	1	0
750	shop spree	1	0				

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
805	roast beef sandwich	1	0	857	nice wine	1	0
806	gorgeou bar	1	0	858	key	1	0
807	southern chile	1	0	859	first night	1	0
808	invit menu	1	0	860	round	0	1
809	beauti environ	1	0	861	yogurt	1	0
810	meringu	1	0	862	friendli staff	1	0
811	good place	1	0	863	beauti decor	1	0
812	cab ride	1	0	864	nice chocol	1	0
813	fine cuisin	1	0	865	yummi daili lunch	1	0
814	disgrac	0	1	866	fresh juic	1	0
815	flouri	1	0	867	rang	1	0
816	basil	1	0	868	gnochi	0	1
817	fillet steak	1	0	869	sidewalk	1	0
818	meat menu	1	0	870	term	1	0
819	info	1	0	871	entrada	1	0
820	h ]	1	0	872	afternoon	1	0
821	excel i	1	0	873	meal cost	1	0
822	allesandro	1	0	874	jasmin tea	1	0
823	pub	1	0	875	great fish	1	0
824	tuna	1	0	876	receptionist	1	0
825	lech	1	0	877	machin	0	1
826	merlot	0	1	878	churrasco	1	0
827	creami qinoua	1	0	879	everyon	0	1
828	small room	0	1	880	inform	0	1
829	wine select	1	0	881	vagu apolog	0	1
830	powder	1	0	882	prior even	1	0
831	tart	1	0	883	chilean wine	1	0
832	wheat	1	0	884	custard	1	0
833	wine boutiqu	1	0	885	amongst	1	0
834	eco	1	0	886	loin	1	0
835	puerta	0	1	887	relax breakfast	1	0
836	differ type	1	0	888	support staff	1	0
837	degustacion	1	0	889	beaten path	0	1
838	good sushi	0	1	890	girth	1	0
839	cost	1	0	891	carrot	1	0
840	macha parmesan	1	0	892	chalet-lik atmospher	1	0
841	italian cuisin	1	0	893	hotel torremolino	1	0
842	femal	1	0	894	guy	0	1
843	da alessandro pizza	1	0	895	cosi restaur	1	0
844	anyon	1	0	896	load	0	1
845	wine cellar	1	0	897	mare	1	0
846	boar	0	1	898	enjoy experi	1	0
847	apolog	0	1	899	campfir	1	0
848	ox	1	0	900	long time	0	1
849	busi place	0	1	901	quaint catalan	1	0
850	excel steak	1	0	902	countri	1	0
851	excel music	1	0	903	appetit	1	0
852	croissant	1	0	904	vanilla ice cream	1	0
853	chalet-lik	1	0	905	imagin menu	1	0
854	lodg	1	0	906	loung corner	1	0
855	d ]	1	0	907	small town	1	0
856	qualiti servic	1	0	908	cute littl restaur	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
909	pisco sour	1	0	912	valpo	0	1
910	long afternoon	1	0	913	clase	1	0
911	problem	1	0	914	night receptionist	1	0

#### C.8. Hasil *Object-based Opinion Mining* Restoran (*Frequent Object* dengan Filter)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review restoran dengan pendekatan *frequent object* yang terfilter sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	food	48	9	24	room	4	3
2	place	37	8	25	area	7	1
3	restaur	25	10	26	dinner	3	1
4	servic	27	4	27	owner	4	0
5	menu	16	4	28	experi	5	1
6	wine	18	1	29	town	4	0
7	pizza	12	2	30	bread	4	1
8	staff	16	0	31	salmon	2	2
9	seafood	14	1	32	chees	2	2
10	dish	12	0	33	breakfast	4	0
11	meal	13	1	34	reserv	1	2
12	salad	8	2	35	pasta	5	0
13	meat	9	2	36	lunch	4	0
14	price	10	3	37	flavour	5	0
15	atmospher	12	0	38	decor	5	0
16	fish	7	1	39	tast	3	0
17	peopl	5	3	40	bar	3	2
18	view	10	0	41	chef	4	1
19	coffe	6	2	42	sandwich	3	0
20	hotel	6	2	43	locat	2	1
21	beer	8	0	44	drink	2	1
22	steak	8	0	45	sauc	2	0
23	qualiti	6	2				

#### C.9. Hasil *Object-based Opinion Mining* Restoran (*Frequent Object and Infrequent Object* dengan Filter)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review restoran dengan pendekatan *frequent object* dan *infrequent object* yang terfilter sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	food	48	9	7	pizza	11	3
2	place	37	8	8	staff	16	0
3	restaur	25	10	9	seafood	13	2
4	servic	26	5	10	meal	12	2
5	menu	15	5	11	dish	12	0
6	wine	19	1	12	salad	8	2
No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
13	price	10	3	71	book	1	0

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
14	meat	9	2	72	platter	1	1
15	atmospher	11	1	73	risotto	2	0
16	fish	7	1	74	bill	2	0
17	peopl	6	2	75	music	2	0
18	view	10	0	76	toast	2	0
19	coffe	6	2	77	deer	2	0
20	beer	8	0	78	cocktail	2	0
21	steak	8	0	79	environ	2	0
22	hotel	6	2	80	video	2	0
23	dinner	3	1	81	dine room	1	0
24	owner	4	0	82	crepe	1	0
25	qualiti	6	2	83	tv	1	1
26	room	4	3	84	fruit	1	0
27	area	7	1	85	hostel	2	0
28	town	3	1	86	pork	2	0
29	experi	5	1	87	rice	1	1
30	chees	2	2	88	wifi	1	0
31	breakfast	4	0	89	server	1	1
32	reserv	1	2	90	manag	2	0
33	salmon	2	2	91	design	2	0
34	bread	4	1	92	pate	1	0
35	sandwich	3	0	93	internet	1	0
36	lunch	4	0	94	season	0	1
37	locat	2	1	95	lake view	1	0
38	flavour	5	0	96	cappuccino	0	1
39	tast	3	0	97	main dish	1	0
40	pasta	5	0	98	spaghetti	1	0
41	decor	5	0	99	barman	1	0
42	bar	3	2	100	dine experi	1	0
43	chef	4	1	101	culinari experi	0	1
44	beef	3	0	102	merlot	0	1
45	pisco	3	1	103	lunch menu	1	0
46	cake	3	0	104	waitstaff	1	0
47	drink	2	1	105	pud	1	0
48	waiter	2	2	106	valu	1	0
49	cevich	4	0	107	wine select	1	0
50	dessert	2	0	108	cost	1	0
51	plate	2	1	109	sushi	0	1
52	soup	2	0	110	product	1	0
53	ambianc	3	0	111	beer list	1	0
54	tea	2	0	112	artefact	1	0
55	flavor	1	0	113	appet	1	0
56	juic	2	0	114	ice cream	1	0
57	mushroom	2	0	115	vine select	1	0
58	potato	2	0	116	gnochi	0	1
59	park	2	1	117	tart	1	0
60	portion	3	0	118	wine recommend	1	0
61	starter	1	0	119	ox	1	0
62	sauc	2	0	120	panna cotta	1	0
63	burger	1	2	121	breakfast place	1	0
64	smoke	1	1	122	qinoua	1	0
65	empanada	3	0	123	seat	1	0
66	waitress	2	1	124	outdoor seat	1	0
67	ale	3	0	125	loung	1	0
68	cuisin	3	0	126	nois	1	0
69	fire	1	0	127	servic co-ordin	1	0
70	outdoor	2	0	128	countri	1	0
129	host	1	0	132	ingredi	1	0

130	jazz	1	0		133	focaccia	1	0
131	stove	1	0					

#### C.10. Hasil *Object-based Opinion Mining* Restoran (*Filter Object*)

Berikut ini merupakan hasil output dari *object-based opinion mining* data review restoran dengan pendekatan *filter object* sebagai acuan ekstraksi.

No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
1	food	48	9	47	drink	2	1
2	place	36	9	48	waiter	3	1
3	restaurant	21	7	49	pisco	3	1
4	service	26	5	50	beef	3	0
5	menu	15	5	51	cake	3	0
6	wine	18	1	52	ceviche	3	1
7	pizza	12	2	53	plate	1	2
8	staff	16	0	54	dessert	2	0
9	fresh	15	1	55	potato	1	1
10	seafood	12	3	56	cuisine	2	1
11	dish	12	0	57	desserts	1	0
12	meal	13	1	58	ambiance	3	0
13	salad	8	2	59	sauce	2	0
14	price	10	3	60	parking	2	1
15	meat	9	2	61	ale	3	0
16	atmosphere	12	0	62	juice	2	0
17	fish	6	2	63	mushroom	2	0
18	restaurants	4	3	64	portion	3	0
19	view	10	0	65	starter	1	0
20	people	6	2	66	soup	2	0
21	beer	8	0	67	smoke	1	1
22	hotel	6	2	68	flavor	1	0
23	steak	8	0	69	waitress	2	1
24	coffee	6	2	70	empanada	3	0
25	area	7	1	71	tea	2	0
26	room	4	3	72	burger	1	2
27	quality	5	3	73	fruit	1	0
28	dinner	3	1	74	wifi	1	0
29	owner	4	0	75	french fries	1	1
30	experience	5	1	76	bill	2	0
31	town	4	0	77	rice	1	1
32	breakfast	4	0	78	server	1	1
33	bread	4	1	79	design	2	0
34	reservation	1	2	80	cocktail	2	0
35	salmon	2	2	81	hostel	2	0
36	cheese	2	2	82	deer	2	0
37	ice cream	5	0	83	fire	1	0
38	taste	2	1	84	dining experience	2	0
39	sandwich	3	0	85	outdoor	2	0
40	bar	3	2	86	video	2	0
41	lunch	4	0	87	tv	1	1
42	flavour	5	0	88	platter	1	1
43	location	2	1	89	crowded	1	1
44	decor	5	0	90	pork	2	0
45	chef	4	1	91	toast	2	0
46	pasta	5	0	92	crepes	1	0
No	Objek	Positif	Negatif	No	Objek	Positif	Negatif
93	book	1	0	121	lunch menu	1	0

94	music	2	0		122	sushi	0	1
95	environment	2	0		123	service co-ordination	1	0
96	dining room	1	0		124	tuna steak	1	0
97	manager	2	0		125	beer list	1	0
98	risotto	2	0		126	artefacts	1	0
99	tart	1	0		127	breakfast place	1	0
100	noise	1	0		128	country	1	0
101	internet	1	0		129	main courses	0	1
102	daily lunch	1	0		130	lounge	1	0
103	focaccia	1	0		131	vine selection	1	0
104	vegetable	0	1		132	pudding	1	0
105	cappuccino	1	0		133	cost	1	0
106	main course	0	1		134	product	1	0
107	dining area	1	0		135	spaghetti	1	0
108	music video	1	0		136	wine selection	1	0
109	parking area	1	0		137	ingredient	1	0
110	stove	1	0		138	seasoning	1	0
111	pate	1	0		139	merlot	1	0
112	panna cotta	1	0		140	culinary experience	0	1
113	lake view	1	0		141	value	1	0
114	waiters	1	0		142	main dish	1	0
115	gnochi	0	1		143	ox	1	0
116	seat	1	0		144	outdoor seating	1	0
117	jazz	1	0		145	outdoor dining	1	0
118	appetizer	1	0		146	waitstaff	1	0
119	barman	1	0		147	qinoua	1	0
120	wine recommendation	1	0		148	host	1	0

*Halaman ini sengaja dikosongkan.*

## LAMPIRAN D

Lampiran ini berisi mengenai data filter yang digunakan dalam penelitian ini. Data filter terdiri dari dua set data, yaitu data filter hotel dan data filter restoran.

### D.1. Data Filter Hotel

Berikut ini merupakan data filter hotel yang digunakan dalam *object-based opinion mining* sebagai filter saat proses ekstraksi objek.

hotel	garden	bathroom	meeting lounge
facility	golf	beauty salon	parking
bathroom	internet access	casino	playground
balance	internet	conference room	lounge
bath	broadband	cot	lounge chairs
bath robes	swimming pool	disco	restaurant
dryer	pool	game room	shoe shine
faucet	parking	hair dresser	smoking area
hygiene product	parking area	hall	square
shower	parking system	heated swimming pool	tennis square
toilet	pool bar	heating system	tennis court
towel	square	internet access	soccer square
wash basin	court	kids club	stair
water	soccer square	laundry	tobacco shop
hot water	tennis square	laundry room	wellness center
hot spring	squash court	laundry service	fitness center
bathroom size	wifi	lift	gym
closet	driveway	luggage room	hydro massage jacuzzi
driver	ground	meeting	massage
external facility	landscape	meeting room	sauna
exterior jacuzzi	internal facility	wifi	turkish bath
jacuzzi	ballroom	wifi	lobby
entrance	cafe	public room	shop
light	wall	crockery	spa
furniture	breakfast area	rec room	ping pong table

basketball hoop	pen	bar	elevator
elevator size	library	cafeteria	computer room
soap	matches	seating area	space
room facility	air condition	air-conditioning	A/C
AC	balcony	terrace	balcony size
bed	bed size	carpet	coffe machine
cuisine	curtain	hanger	iron
steam iron	iron board	kettle	electric kettle
manual kettle	lamp	room light	mattress
mini-bar	mirror	phone	pillow
plug socket	snacks	tv	tv remote
tv cable	tv guide	wardrobe	wifi
wifi signal	kitchen	kitchen equipment	dining room
dining area	dinning room	window	window size
toaster	room night light	sheet	fireplace
sitting room	sitting room size	sofa	chair
easy chair	table	coffee table	bookcase
refrigerator	fridge	back room	room size
decoration	room decoration	decor	living room
computer	heater	room ceiling	ceiling
menu	bedroom	bedroom size	microwave
bathtub	bathtub size	cookware	stove
detergent	sponge	cabin	flower
coffee maker	pot	game	drawer
shelf space	wheelchair accessible	equipment	hotel size
bike loan	bike	building	animal
location	place	neighbourhood	atmosphere
meal	food	breakfast	breakfast selection
dinner	lunch	quality	ingredient
freshness	food freshness	brunch	point of interest
spot	environment	airport	arena
beach	bus station	bus stop	church
downtown	gardens	garden	historic building
historic centre	monument	museum	park
shopping	squares	stadium	theatre
university	mall	lakeside	night activity
localization	lake	lake access	party

price	cost	rate	bar price
breakfast price	coffe price	internet price	parking price
restaurant price	room price	parking price	wifi price
broadband price	laundry price	tour price	rating
accomodation	access	design	room
deluxe room	deluxe double room	deluxe family room	double room
family room	romantic room	single room	standard room
suit	suite	triple room	twin
room quality	room temperature	heating	paint
room smell	bedroom smell	service	airport shuttle
baby sitting	bicycle rental	car rental	checkout service
reception	concierge service	currency exchange	dry cleaning service
newspaper	honeymoon suite	house keeping	in room breakfast
room service	tour desk	reservation	stay
reception service	front desk service	front-desk service	service quality
hospitality	cleanliness	room cleanliness	bathroom cleanliness
surroundings cleanliness	bedroom cleanliness	hotel cleanliness	pool cleanliness
cabin cleanliness	service speed	food flavor	food taste
menu price	food variety	food quality	wifi speed
dessert flavour	towel quality	amenity	website
site	beach service	tour	staff
animator	bar staff	cleaning staff	doorman
kitchen staff	leisure staff	management	reception
receptionist	restaurant staff	swimming pool staff	owner
host	chef	waitress	waiter
manager	timetable	indoor timetable swimming pool	outdoor swimming pool timetable
restaurant timetable	spa centre timetable	view	lake view
volcano view	mountain view	beverage	wine
wine choice	juice	coffee	espresso
tea	food	food options	hamburger
bread	cereal	cake	pasta
egg	croissant	fruit	yogurt
cheese	jam	ham	marmelade
hot food	meat	pastry	dessert

sandwich			
----------	--	--	--

## D.2. Data Filter Restoran

Berikut ini merupakan data filter restoran yang digunakan dalam *object-based opinion mining* sebagai filter saat proses ekstraksi objek.

restaurant	location	address	area
place	town	country	atmosphere
environment	ambiance	view	view lake
view montain	noise	crowded	people
beverage	drink	beer	beer list
wine	wine selection	wine list	vine selection
merlot	wine quality	wine recommendation	digestives
champagne	whiskey	softdrink	coffee
juice	cappuccino	ale	cocktail
pisco	tea	contact details	contact
date	date periode	day of week	duration
facility	access facility	metro facility	business facility
business breakfast	business dining	children and baby facility	baby chair
baby laundry	baby sitter	children game arena	children meal
disabled facility	wheel chair access	elevator	parking facility
parking	parking area	parking indoor	parking outdoor
private parking	public parking	valet parking	payment facility
payment card	pet facility	only dog accepted	pet accepted
pet food	pet refused	smoking facility	no smoking
smoking	smoke	smoking area	smoking room
wifi	internet access	entertainment facility	video
movie	tv	music	jazz
band	music video	integrated facility	hotel
hostel	office	other facility	air conditioner
ac	fireplace	fire	dining room
breakfast place	lounge	dining area	dining area size
room	plate	server	bill

internet	stove	bar	meal
food	dish	cuisine	pizza
steak	salad	ice cream	seafood
burger	hamburger	bread	sushi
sandwich	cake	tart	pastry
cookie	pasta	pudding	cheese
cheese selection	fruit	toast	rice
french fries	risotto	crepes	night food
muffin	soup	qinoua	panna cotta
spaghetti	tuna steak	gnochi	sanguche
empanada	focaccia	ceviche	pate
opening period	price	cost	rate
book price	rating	restaurant service	delivery
take away	reservation	book	outdoor dining
restaurant group party	birthday party	christmas party	cocktail party
convention party	new year evening party	office party	wedding party
restaurant seating	seat	air conditioning	indoor seating
indoor	outdoor seating	outdoor	restaurant specific features
dinner	breakfast	lunch	daily lunch
brunch	show	staff	chef
manager	hostess	cleaning	barman
waiters	waiter	waitress	owner
host	waitstaff	menu	menu choice
drinks menu	alcoholic drinks	non alcoholic drinks	food menu
desserts	sides	main courses	appetizers
day menu	daily menu	lunch menu	food component
meat	beef	fish	salmon
pork	deer	ox	sauce
vegetable	mushroom	potato	egg
fruit	quality	ingredient	freshness
fresh	service	service speed	meat amount
taste	flavor	flavour	food quality
platter	portion	service co- ordination	presentation

seasoning	value	texture	experience
vegetarian experience	culinary experience	dining experience	product
building	decor	design	artefacts

## LAMPIRAN E

Lampiran ini berisi mengenai data sebagian script program prototype yang digunakan dalam penelitian ini.

### E.1. Script Praproses Data

Berikut ini merupakan script yang digunakan untuk tahap praproses data. Script ini ditulis menggunakan bahasa pemrograman Python.

```
import nltk
import MySQLdb
from nltk.corpus import state_union
from nltk.tokenize import PunktSentenceTokenizer
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.chunk import RegexpParser
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
def getid(num):
    if num<10:
        number = "0000000" + str(num)
    elif num<100:
        number = "000000" + str(num)
    elif num<1000:
        number = "00000" + str(num)
    elif num<10000:
        number = "0000" + str(num)
    elif num<100000:
        number = "000" + str(num)
    elif num<1000000:
        number = "00" + str(num)
    elif num<10000000:
        number = "0" + str(num)
    else:
        number = str(num)
    return str(number)
review_id = 'R.2016.05.27.00001'
conn = MySQLdb.connect("localhost","root","","thesis")
c = conn.cursor()
no3 = 0
stop_words = set(stopwords.words('english'))
ps = PorterStemmer()
c.execute("SELECT * FROM review_detail WHERE review_id = %s ORDER BY
review_detail_id", (review_id,))
rows = c.fetchall()
for eachRow in rows:
    sentence = eachRow[2]
    sentence = sentence.replace("/", " or ")
    word_tokens = nltk.word_tokenize(sentence)
    tagged = nltk.pos_tag(word_tokens)
    patterns = """
    NP: {<NN>+<NN>}
    {<JJ>+<NN>}
    """
    NPChunker = nltk.RegexpParser(patterns)
    result = NPChunker.parse(tagged)
    no2 = 0
```

```

for subtree in result.subtrees(filter=lambda t: t.label() == 'NP'):
    no3 += 1
    np_tag = str(subtree.leaves())
    np_tag = np_tag.replace("'", "")
    np_tag = np_tag.replace(" ", "")
    np_tag = np_tag.replace("[", "")
    np_tag = np_tag.replace("]", "")
    np_real = ""
    np_stem = ""
    no = 0
    for leave in subtree.leaves():
        no += 1
        if no == 1:
            np_real += leave[0]
            np_stem += ps.stem(leave[0])
        else:
            np_real += " "+leave[0]
            np_stem += " "+ps.stem(leave[0])
    np_stem = np_stem.replace("'", "")
    np_stem = np_stem.replace(" ", "")
    np_stem = np_stem.replace("", "")
    np_stem = np_stem.replace("?", "")
    np_stem = np_stem.replace("!", "")
    preprocessing_np_id = review_id + ".PNP." + getid(no3)
    c.execute("INSERT INTO preprocessing_np (preprocessing_np_id, review_id, review_detail_id,
preprocessing_np_detail, preprocessing_np_original, preprocessing_np_word) VALUES
(%s,%s,%s,%s,%s,%s",
           ,(preprocessing_np_id, review_id, eachRow[0], np_tag, np_real, np_stem))
    conn.commit()
    no2 = 0
    for tag in tagged:
        no1 += 1
        preprocessing_data_id = review_id + ".PD." + getid(no1)
        no2 += 1
        preprocessing_data_word = tag[0]
        preprocessing_data_tag = tag[1]
        preprocessing_data_word_processed = ""
        if preprocessing_data_word not in stop_words:
            preprocessing_data_word_processed = ps.stem(preprocessing_data_word)
            preprocessing_data_word_processed = preprocessing_data_word_processed.replace(".", "")
            preprocessing_data_word_processed = preprocessing_data_word_processed.replace(" ", "")
            preprocessing_data_word_processed = preprocessing_data_word_processed.replace("", "")
            preprocessing_data_word_processed = preprocessing_data_word_processed.replace("?", "")
            preprocessing_data_word_processed = preprocessing_data_word_processed.replace("!", "")
            c.execute("INSERT INTO preprocessing_data (preprocessing_data_id, review_id, review_detail_id,
preprocessing_data_word_number, preprocessing_data_word, preprocessing_data_tag,
preprocessing_data_word_processed, preprocessing_data_word_orientation,
preprocessing_data_word_feature) VALUES (%s,%s,%s,%s,%s,%s,%s,%s,%s)",(preprocessing_data_id,
review_id, eachRow[0], no2, preprocessing_data_word, preprocessing_data_tag,
preprocessing_data_word_processed, "", ""))
            conn.commit()

```

## E.2. Script Ekstraksi *Frequent Object*

Berikut ini merupakan script yang digunakan untuk tahap ekstraksi *frequent object*. Script ini ditulis menggunakan bahasa pemrograman PHP.

```

mysql_query("UPDATE review SET start_object_extraction_frequent = NOW() WHERE review_id =
'$r[review_id]'");
mysql_query("UPDATE review SET start_object_extraction_filtered_frequent = NOW() WHERE review_id =
'$r[review_id]'");

```

```

$stampil2=mysql_query("SELECT preprocessing_np_word,count(preprocessing_np_word) as jml FROM
preprocessing_np where review_id='$r[review_id]' and preprocessing_np_word not in(") group by
preprocessing_np_word ORDER BY count(preprocessing_np_word) desc");
$no = 1;
while ($r2=mysql_fetch_array($stampil2)){
if($r2['jml']>=$min){
$frequent_object_id=generate_id('FA','frequent_object','frequent_object_id');
mysql_query("INSERT INTO frequent_object(frequent_object_id, review_id, frequent_object_word,
frequent_object_number_appears, frequent_object_status) VALUES('$frequent_object_id', '$r[review_id]', '$r2[preprocessing_np_word]', '$r2[jml]', 'NP')");
}
}
$stampil3=mysql_query("SELECT preprocessing_data_word_processed,
count(preprocessing_data_word_processed) as jml FROM preprocessing_data where
review_id='$r[review_id]' and preprocessing_data_tag in ('NNS','NN') and
preprocessing_data_word_processed not in(") and preprocessing_data_word_processed not in (SELECT
preprocessing_np_word FROM preprocessing_np where review_id='$r[review_id]' and
preprocessing_np_word not in(") group by preprocessing_np_word ORDER BY
count(preprocessing_np_word) desc) group by preprocessing_data_word_processed ORDER BY
count(preprocessing_data_word_processed) desc");
$no = 1;
while ($r3=mysql_fetch_array($stampil3)){
if($r3['jml']>=$min and strlen($r3['preprocessing_data_word_processed'])>1){
$frequent_object_id=generate_id('FA','frequent_object','frequent_object_id');
mysql_query("INSERT INTO
frequent_object(frequent_object_id,review_id,frequent_object_word,frequent_object_number_appears,
frequent_object_status) VALUES('$frequent_object_id', '$r[review_id]', '$r3[preprocessing_data_word_processed]', '$r3[jml]', 'NN')");
}
}
mysql_query("UPDATE review SET finish_object_extraction_frequent = NOW() WHERE review_id =
'$r[review_id]'");
mysql_query("UPDATE review SET finish_object_extraction_filtered_frequent = NOW() WHERE
review_id = '$r[review_id]'");

```

### E.3. Script Ekstraksi *Infrequent Object*

Berikut ini merupakan script yang digunakan untuk tahap ekstraksi *infrequent object*. Script ini ditulis menggunakan bahasa pemrograman PHP.

```

mysql_query("UPDATE review SET start_object_extraction_infrequent = NOW() WHERE review_id =
'$r[review_id]'");
mysql_query("UPDATE review SET start_object_extraction_filtered_infrequent = NOW() WHERE
review_id = '$r[review_id]'");
$stampil2=mysql_query("SELECT preprocessing_np_word,count(preprocessing_np_word) as jml FROM
preprocessing_np where review_id='$r[review_id]' and preprocessing_np_word not in(") group by
preprocessing_np_word ORDER BY count(preprocessing_np_word) desc");
$no = 1;
while ($r2=mysql_fetch_array($stampil2)){
if($r2['jml']<$min ){
$infrequent_object_id=generate_id('IA','infrequent_object','infrequent_object_id');
mysql_query("INSERT INTO infrequent_object(infrequent_object_id, review_id,infrequent_object_word,
infrequent_object_number_appears, infrequent_object_status) VALUES('$infrequent_object_id',
'$r[review_id]', '$r2[preprocessing_np_word]', '$r2[jml]', 'NP')");
}
}
$stampil3=mysql_query("SELECT
preprocessing_data_word_processed, count(preprocessing_data_word_processed) as jml FROM
preprocessing_data where review_id='$r[review_id]' and preprocessing_data_tag in ('NNS','NN') and
preprocessing_data_word_processed not in(") and preprocessing_data_word_processed not in (SELECT
preprocessing_np_word FROM preprocessing_np where review_id='$r[review_id]' and
preprocessing_np_word not in(") group by preprocessing_np_word ORDER BY
count(preprocessing_np_word) desc)

```

```

preprocessing_np_word    not    in(")    group    by    preprocessing_np_word    ORDER    BY
count(preprocessing_np_word)    desc)    group    by    preprocessing_data_word_processed    ORDER    BY
count(preprocessing_data_word_processed)    desc");
$no = 1;
while ($r3=mysql_fetch_array($stampil3)){
    if($r3['jml']<$min and strlen($r3['preprocessing_data_word_processed'])>1){
        $infrequent_object_id=generate_id('IA','infrequent_object','infrequent_object_id');
        mysql_query("INSERT INTO infrequent_object(infrequent_object_id, review_id, infrequent_object_word,
infrequent_object_number_appears, infrequent_object_status) VALUES('$infrequent_object_id',
'$r[review_id]', '$r3[preprocessing_data_word_processed]', '$r3[jml]', 'NN')");
    }
}
mysql_query("UPDATE review SET finish_object_extraction_infrequent = NOW() WHERE review_id =
'$r[review_id]'");
mysql_query("UPDATE review SET finish_object_extraction_filtered_infrequent = NOW() WHERE
review_id = '$r[review_id]'");

```

#### E.4. Script Ekstraksi *Filter Object*

Berikut ini merupakan script yang digunakan untuk tahap ekstraksi *filter object*. Script ini ditulis menggunakan bahasa pemrograman PHP.

```

mysql_query("UPDATE review SET start_object_extraction_filter = NOW() WHERE review_id =
'$r[review_id]'");
$no = 1;
for($h=1;$h<=$nox2;$h++){
    for($i=1;$i<=$nox;$i++){
        if(sizeof(explode(" ",$hotel_filter_word[$i]))>1){
            if(strpos($review_sentence[$h], $hotel_filter_word[$i]) !== false) {

                $jmldata4 = mysql_num_rows(mysql_query("SELECT * FROM filter_object WHERE
review_id='$r[review_id]' and filter_object_word='".$hotel_filter_word[$i]."' ORDER BY filter_object_id"));
                if($jmldata4<1){
                    $filter_object_id=generate_id('OA','filter_object','filter_object_id');
                    mysql_query("INSERT INTO filter_object(filter_object_id, review_id, filter_object_word)
VALUES('$filter_object_id', '$r[review_id]', '".$hotel_filter_word[$i]."')");
                }
            }else{
                $jmldata4 = mysql_num_rows(mysql_query("SELECT * FROM preprocessing_data WHERE
review_detail_id='".$review_detail_id[$h]."' and preprocessing_data_word not in () and
preprocessing_data_word_processed not in () and (preprocessing_data_word='".$hotel_filter_word[$i]."' or
preprocessing_data_word_processed='".$hotel_filter_word[$i]."' ) ORDER BY preprocessing_data_id"));
                if($jmldata4>0){
                    $jmldata5 = mysql_num_rows(mysql_query("SELECT * FROM filter_object WHERE
review_id='$r[review_id]' and filter_object_word='".$hotel_filter_word[$i]."' ORDER BY filter_object_id"));
                    if($jmldata5<1{
                        $filter_object_id=generate_id('OA','filter_object','filter_object_id');
                        mysql_query("INSERT INTO filter_object(filter_object_id, review_id, filter_object_word)
VALUES('$filter_object_id', '$r[review_id]', '".$hotel_filter_word[$i]."')");
                    }
                }
            }
        }
    }
}
mysql_query("UPDATE review SET finish_object_extraction_filter = NOW() WHERE review_id =
'$r[review_id]'");

```

#### E.5. Script Proses Klasifikasi Sentimen Objek

Berikut ini merupakan script yang digunakan untuk proses klasifikasi sentimen objek. Script ini ditulis menggunakan bahasa pemrograman PHP.

```
mysql_query("UPDATE review SET start_object_orientation_frequent = NOW() WHERE review_id = '$r[review_id]'");
mysql_query("UPDATE review SET start_object_orientation_filtered_frequent = NOW() WHERE review_id = '$r[review_id]'");
$stampil1=mysql_query("SELECT * FROM review_detail where review_id='$r[review_id]' and review_detail_id in (SELECT review_detail_id FROM preprocessing_data where review_id='$r[review_id]' and preprocessing_data_word_orientation not in () ORDER BY preprocessing_data_id) ORDER BY review_detail_id limit $r[valid_opinion_mining_frequent],1");
while ($r1=mysql_fetch_array($stampil1)){
    mysql_query("DELETE FROM opinion_mining_result WHERE review_detail_id='$r1[review_detail_id]' and object_type='frequent'");
}
$stampil1=mysql_query("SELECT * FROM review_detail where review_id='$r[review_id]' and review_detail_id in (SELECT review_detail_id FROM preprocessing_data where review_id='$r[review_id]' and preprocessing_data_word_orientation not in () ORDER BY preprocessing_data_id) ORDER BY review_detail_id");
$no1 = 0;
while ($r1=mysql_fetch_array($stampil1)){
$no1++;
if($no1>$r['valid_opinion_mining_frequent']){
    unset($object);
    unset($object_real);
    unset($object_orientation);
    unset($as_pos);
    $nof=0;
    $batas="";
    $stampil2=mysql_query("SELECT * FROM preprocessing_np where review_detail_id='$r1[review_detail_id]' and preprocessing_np_word in (select frequent_object_word from frequent_object where review_id='$r[review_id]' ) ORDER BY preprocessing_np_id");
    //echo "SELECT * FROM preprocessing_np where review_detail_id='$r1[review_detail_id]' and preprocessing_np_word in (select frequent_object_word from frequent_object where review_id='$r[review_id]' and frequent_object_status='NP') ORDER BY preprocessing_np_id <br />";
    while ($r2=mysql_fetch_array($stampil2)){
        $nof++;
        $object[$nof]=$r2['preprocessing_np_word'];
        $object_real[$nof]=$r2['preprocessing_np_original'];
        $batas.=",$object[$nof].";
    }
    $stampil2=mysql_query("SELECT * FROM preprocessing_data where review_detail_id='$r1[review_detail_id]' and preprocessing_data_word_processed in (select frequent_object_word from frequent_object where review_id='$r[review_id]' ) and preprocessing_data_word_processed not in(\".$batas.\") ORDER BY preprocessing_data_id");
    //echo "SELECT * FROM preprocessing_data where review_detail_id='$r1[review_detail_id]' and preprocessing_data_word_processed in (select frequent_object_word from frequent_object where review_id='$r[review_id]' and frequent_object_status='NN') ORDER BY preprocessing_data_id <br />";
    while ($r2=mysql_fetch_array($stampil2)){
        $nof++;
        $object[$nof]=$r2['preprocessing_data_word_processed'];
        $object_real[$nof]=$r2['preprocessing_data_word'];
    }
    $batas1="";
    $batas2="";
    $batas3="";
    for($i=1;$i<=$nof;$i++){
        $as=explode(" ",$object[$i]);
        if(sizeof($as)==1){
            $stampil3=mysql_query("SELECT * FROM preprocessing_data where review_detail_id='$r1[review_detail_id]' and preprocessing_data_word_processed=\"$object[$i].\" and preprocessing_data_word_number not in (\".$batas1.\") ORDER BY preprocessing_data_id");
        }
    }
}
```

```

$rl=mysql_fetch_array($tampil3);
$object_pos=$rl['preprocessing_data_word_number'];
$batas1.="$object_pos";
}elseif(sizeof($as)==2){
$find=0;
do{
    $tampil3=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id=$rl[review_detail_id] and preprocessing_data_word_processed='".$as[0]." and
preprocessing_data_word_number not in (".$batas2.") ORDER BY preprocessing_data_id");
    $rl=mysql_fetch_array($tampil3);
    $object_pos1=$rl['preprocessing_data_word_number'];
    $batas2.="$object_pos1";
    $tampil4=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id=$rl[review_detail_id] and preprocessing_data_word_processed='".$as[1]." and
preprocessing_data_word_number='".$( $object_pos1+1)." and preprocessing_data_word_number not in
("$. $batas2.") ORDER BY preprocessing_data_id");
    $j4=mysql_num_rows($tampil4);
    $rl=mysql_fetch_array($tampil4);
    $object_pos2=$rl['preprocessing_data_word_number'];
    $batas2.="$object_pos2";
    if($j4>0){
        $find=1;
        $object_pos=$object_pos1;
    }
}while($find==0);
}
elseif(sizeof($as)==3){
$find=0;
do{
    $tampil3=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id=$rl[review_detail_id] and preprocessing_data_word_processed='".$as[0]." and
preprocessing_data_word_number not in (".$batas3.") ORDER BY preprocessing_data_id");
    $rl=mysql_fetch_array($tampil3);
    $object_pos1=$rl['preprocessing_data_word_number'];
    $batas3.="$object_pos1";
    $tampil4=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id=$rl[review_detail_id] and preprocessing_data_word_processed='".$as[1]." and
preprocessing_data_word_number='".$( $object_pos1+1)." and preprocessing_data_word_number not in
("$. $batas3.") ORDER BY preprocessing_data_id");
    $j4=mysql_num_rows($tampil4);
    $rl=mysql_fetch_array($tampil4);
    $object_pos2=$rl['preprocessing_data_word_number'];
    $batas3.="$object_pos2";
    if($j4>0){
        $tampil5=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id=$rl[review_detail_id] and preprocessing_data_word_processed='".$as[2]." and
preprocessing_data_word_number='".$( $object_pos2+1)." and preprocessing_data_word_number not in
("$. $batas3.") ORDER BY preprocessing_data_id");
        $j5=mysql_num_rows($tampil5);
        $rl=mysql_fetch_array($tampil5);
        $object_pos3=$rl['preprocessing_data_word_number'];
        $batas3.="$object_pos3";
        if($j5>0){
            $find=1;
            $object_pos=$object_pos1;
        }
    }
}while($find==0);
}
$as_pos[$i]=$object_pos;
}
$object_view="";
for($i=1;$i<=$nof;$i++){

```

```

if($i==1){
    $object_view.=$object[$i]['$.as_pos[$i].'];
}else{
    $object_view.= ' '.$object[$i]['$.as_pos[$i].'];
}
}
unset($mark_word);
unset($mark_word_real);
unset($mark_word_pos);
unset($mark_word_orientation);
$nomw=0;
$tampil6=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id='".$r1[review_detail_id]' and preprocessing_data_word_orientation not in () ORDER BY
processing_data_id");
$j6=mysql_num_rows($tampil6);
while($r6=mysql_fetch_array($tampil6)){
    $nomw++;
    $mark_word[$nomw]=$r6['processing_data_word_processed'];
    $mark_word_real[$nomw]=$r6['processing_data_word'];
    $mark_word_pos[$nomw]=$r6['processing_data_word_number'];
    $mark_word_orientation[$nomw]=$r6['processing_data_word_orientation'];
}
$mark_word_list="";
for($i=1;$i<=$nomw;$i++){
    if($i==1){
        $mark_word_list.=$mark_word[$i]['$.mark_word_orientation[$i].'];
    }else{
        $mark_word_list.= ' '.$mark_word[$i]['$.mark_word_orientation[$i].'];
    }
}
$but='no';
$although='no';
$tampil7=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id='".$r1[review_detail_id]' and preprocessing_data_word in ('but','except','exception') ORDER
BY processing_data_id");
$j7=mysql_num_rows($tampil7);
$tampil72=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id='".$r1[review_detail_id]' and preprocessing_data_word in ('although','though','even
though','even if','whereas') ORDER BY processing_data_id");
$j72=mysql_num_rows($tampil72);
$tampil74=mysql_query("SELECT * FROM review_detail where review_detail_id='".$r1[review_detail_id]'
and (review_sentence like '%even though%' or review_sentence like '%even if%')");
$j74=mysql_num_rows($tampil74);
if($j7>0){
    $r7=mysql_fetch_array($tampil7);
    $but='yes';
    $but_pos=$r7['processing_data_word_number'];
}elseif($j74>0){
    $tampil75=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id='".$r1[review_detail_id]' and preprocessing_data_word in ('even') ORDER BY
processing_data_id");
    while($r75=mysql_fetch_array($tampil75)){
        $tampil76=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id='".$r1[review_detail_id]' and
processing_data_word_number=".($r75['processing_data_word_number']+1)." ORDER BY
processing_data_id");
        $r76=mysql_fetch_array($tampil76);
        if($r76['processing_data_word']=='though' or $r76['processing_data_word']=='if'){
            $although='yes';
            $although_pos=$r75['processing_data_word_number'];
            break;
        }
    }
}

```

```

}elseif($j72>0){
    $r72=mysql_fetch_array($stampil72);
    $although='yes';
    $although_pos=$r72['preprocessing_data_word_number'];
}
if($nomw>0 and $nof>0){
    if($but=='yes'){
        $or_but=0;
        for($i=1;$i<=$nof;$i++){
            $object_orientation[$i]=0;
            if($as_pos[$i]>$but_pos){
                //but clause
                $fo='no';
                for($j=1;$j<=$nomw;$j++){
                    if($mark_word_pos[$j]>$but_pos and $mark_word_orientation[$j]!=0 and $fo=='no'){
                        $object_orientation[$i]=$mark_word_orientation[$j];
                        $or_but=$mark_word_orientation[$j];
                        $fo='yes';
                    }
                }
            }
        }
        $or_not_but=0;
        for($i=1;$i<=$nof;$i++){
            // $object_orientation[$i]=0;
            if($as_pos[$i]<$but_pos){
                //but clause
                $fo='no';
                for($j=1;$j<=$nomw;$j++){
                    if($mark_word_pos[$j]<$but_pos and $mark_word_orientation[$j]!=0 and $fo=='no'){
                        $object_orientation[$i]=$mark_word_orientation[$j];
                        $or_not_but=$mark_word_orientation[$j];
                        $fo='yes';
                    }
                }
            }
        }
        for($i=1;$i<=$nof;$i++){
            if($object_orientation[$i]==0){
                if($as_pos[$i]<$but_pos and $or_but!=0){
                    $object_orientation[$i]=$or_but*(-1);
                }elseif($as_pos[$i]>$but_pos and $or_not_but!=0){
                    $object_orientation[$i]=$or_not_but*(-1);
                }
            }
        }
        $or_object='no';
        for($i=1;$i<=$nof;$i++){
            if($object_orientation[$i]!=0){
                $or_object='yes';
            }
        }
        $or_mw='no';
        for($j=1;$j<=$nomw;$j++){
            if($mark_word_orientation[$j]!=0){
                $or_mw='yes';
            }
        }
        if($or_object=='no' and $or_mw=='yes'){
/////////////////
            for($i=1;$i<=$nof;$i++){
                // $object_orientation[$i]=0;
                if($object_orientation[$i]==0){

```

```

//but clause
$fo='no';
for($j=1;$j<=$nomw;$j++){
    if($mark_word_orientation[$j]!=0 and $fo=='no'){
        $object_orientation[$i]=$mark_word_orientation[$j];
        $fo='yes';
    }
}
}

}elseif($although=='yes'){
$pecah_kalimat=0;
if($although_pos==1){
    $stampil73=mysql_query("SELECT * FROM preprocessing_data where
review_detail_id='".$r1['review_detail_id']."' and preprocessing_data_word in (',') ORDER BY
preprocessing_data_id");
    $r73=mysql_fetch_array($stampil73);
    $pecah_kalimat=$r73['preprocessing_data_word_number'];
} else{
    $pecah_kalimat=$although_pos;
}
$or_although=0;
for($i=1;$i<=$nof;$i++){
    $object_orientation[$i]=0;
    if($as_pos[$i]>$pecah_kalimat){
        //although clause
        $fo='no';
        for($j=1;$j<=$nomw;$j++){
            if($mark_word_pos[$j]>$pecah_kalimat and $mark_word_orientation[$j]!=0 and $fo=='no'){
                $object_orientation[$i]=$mark_word_orientation[$j];
                $or_although=$mark_word_orientation[$j];
                $fo='yes';
            }
        }
    }
}
$or_not_although=0;
for($i=1;$i<=$nof;$i++){
    //object_orientation[$i]=0;
    if($as_pos[$i]<$pecah_kalimat){
        //although clause
        $fo='no';
        for($j=1;$j<=$nomw;$j++){
            if($mark_word_pos[$j]<$pecah_kalimat and $mark_word_orientation[$j]!=0 and $fo=='no'){
                $object_orientation[$i]=$mark_word_orientation[$j];
                $or_not_although=$mark_word_orientation[$j];
                $fo='yes';
            }
        }
    }
}
for($i=1;$i<=$nof;$i++){
    if($object_orientation[$i]==0){
        if($as_pos[$i]<$pecah_kalimat and $or_although!=0){
            $object_orientation[$i]=$or_although*(-1);
        } elseif($as_pos[$i]>$pecah_kalimat and $or_not_although!=0){
            $object_orientation[$i]=$or_not_although*(-1);
        }
    }
}
$or_object='no';

```

```

for($i=1;$i<=$nof;$i++){
    if($object_orientation[$i]!=0){
        $or_object='yes';
    }
}
$or_mw='no';
for($j=1;$j<=$nomw;$j++){
    if($mark_word_orientation[$j]!=0){
        $or_mw='yes';
    }
}
if($or_object=='no' and $or_mw=='yes'){
///////////////////////////////
for($i=1;$i<=$nof;$i++){
    // $object_orientation[$i]=0;
    if($object_orientation[$i]==0){
        //although clause
        $fo='no';
        for($j=1;$j<=$nomw;$j++){
            if($mark_word_orientation[$j]!=0 and $fo=='no'){
                $object_orientation[$i]=$mark_word_orientation[$j];
                $fo='yes';
            }
        }
    }
}
}else{
    $final_orientation=0;
    for($i=1;$i<=$nof;$i++){
        $orientation=0;
        $as=explode(" ",$object[$i]);
        for($j=1;$j<=sizeof($as);$j++){
            $suborientation=0;
            for($k=1;$k<=$nomw;$k++){
                if(abs($mark_word_pos[$k]-($as_pos[$i]+($j-1)))!=0)){
                    $suborientation+=round($mark_word_orientation[$k]/abs($mark_word_pos[$k]-($as_pos[$i]+($j-1)))),4);
                }
            }
            $orientation+=$suborientation;
        }
        $final_orientation+=$orientation;
    }
    if($final_orientation>0){
        $orientation_sentence=1;
    }elseif($final_orientation<0){
        $orientation_sentence=-1;
    }else{
        $orientation_sentence=0;
    }
}

for($i=1;$i<=$nof;$i++){
    $object_orientation[$i]=$orientation_sentence;
}
}
$output_analisis="";
for($i=1;$i<=$nof;$i++){
    if($object_orientation[$i]==1){
        $object_orientation_symbol='+';
    }elseif($object_orientation[$i]==(-1)){
        $object_orientation_symbol='-' ;
    }
}

```

```

}else{
    $object_orientation_symbol="";
}
if($object_orientation_symbol!=""){
    if($i==1){
        $output_analisis.= $object[$i]."[.$object_orientation_symbol.]";
    }else{
        $output_analisis.= '.'.$object[$i]."[.$object_orientation_symbol.]";
    }
}
$tampil8=mysql_query("SELECT * FROM hotel_filter where
hotel_filter_word="" .str_replace("", "", $object[$i])."" or
hotel_filter_word="" .str_replace("", "", $object_real[$i])."" ORDER BY hotel_filter_id");
$j8=mysql_num_rows($tampil8);
if($j8>0){
    $filter_filter='pass';
}else{
    $filter_filter='failed';
}
$opinion_mining_result_id=generate_id('OMR','opinion_mining_result','opinion_mining_result_id');
mysql_query("INSERT INTO opinion_mining_result(opinion_mining_result_id, review_id,
review_detail_id, object_word, object_real_word, object_position, object_orientation, object_type,
filter_filter) VALUES('$opinion_mining_result_id', '$r[review_id]', '$r1[review_detail_id]', ".$object[$i].",
".$object_real[$i].", ".$as_pos[$i].", '$object_orientation_symbol', 'frequent', '$filter_filter')");
}elseif($object_orientation_symbol==""){
    if($i==1){
        $output_analisis.= $object[$i]."[.$object_orientation_symbol.]";
    }else{
        $output_analisis.= '.'.$object[$i]."[.$object_orientation_symbol.]";
    }
//echo "SELECT * FROM hotel_filter where hotel_filter_word="" .str_replace("", "", $object[$i])."" or
hotel_filter_word="" .str_replace("", "", $object_real[$i])."" ORDER BY hotel_filter_id";
$tampil8=mysql_query("SELECT * FROM hotel_filter where
hotel_filter_word="" .str_replace("", "", $object[$i])."" or
hotel_filter_word="" .str_replace("", "", $object_real[$i])."" ORDER BY hotel_filter_id");
$j8=mysql_num_rows($tampil8);
if($j8>0){
    $filter_filter='pass';
}else{
    $filter_filter='failed';
}
$opinion_mining_result_id=generate_id('OMR','opinion_mining_result','opinion_mining_result_id');
mysql_query("INSERT INTO opinion_mining_result(opinion_mining_result_id, review_id,
review_detail_id, object_word, object_real_word, object_position, object_orientation, object_type,
filter_filter) VALUES('$opinion_mining_result_id', '$r[review_id]', '$r1[review_detail_id]', ".$object[$i].",
".$object_real[$i].", ".$as_pos[$i].", 'N', 'frequent', '$filter_filter')");
}
}
mysql_query("UPDATE review SET valid_opinion_mining_frequent = '$no1' WHERE review_id =
'$r[review_id]'");
mysql_query("UPDATE review SET valid_opinion_mining_filtered_frequent = '$no1' WHERE review_id =
'$r[review_id]'");
}
}
mysql_query("UPDATE review SET finish_object_orientation_frequent = NOW() WHERE review_id =
'$r[review_id]');
mysql_query("UPDATE review SET finish_object_orientation_filtered_frequent = NOW() WHERE
review_id = '$r[review_id]'");

```

## E.6. Script Proses Deteksi Kalimat Opini

Berikut ini merupakan script yang digunakan untuk proses deteksi kalimat opini. Script ini ditulis menggunakan bahasa pemrograman PHP.

```

mysql_query("UPDATE review SET start_opinion_sentence_detection_frequent = NOW() WHERE
review_id = '$r[review_id]'");
mysql_query("UPDATE review SET start_opinion_sentence_detection_filtered_frequent = NOW() WHERE
review_id = '$r[review_id]'");
$tp5=0;
$fn5=0;
$fp5=0;
$tn5=0;
$tp52=0;
$fn52=0;
$fp52=0;
$tn52=0;
$stampil2=mysql_query("SELECT * FROM review_detail WHERE review_id='$r[review_id]' ORDER BY
review_detail_id");
$no = 0;
while ($r2=mysql_fetch_array($stampil2)){
    $no++;
    $data=$no-1;
    $batas="";
    $stampil4=mysql_query("SELECT * FROM hotel_corpus ORDER BY hc_id LIMIT $data,1");
    $r4=mysql_fetch_array($stampil4);
    $stampil5=mysql_query("SELECT * FROM hotel_corpus_object where hc_id='$r4[hc_id]' ORDER BY
hc_id");
    $j5=mysql_num_rows($stampil5);
    $stampil6=mysql_query("SELECT * FROM opinion_mining_result where
review_detail_id='$r2[review_detail_id]' and object_type='frequent' and object_orientation not in ('N')
ORDER BY opinion_mining_result_id");
    $j6=mysql_num_rows($stampil6);
    $stampil62=mysql_query("SELECT * FROM opinion_mining_result where
review_detail_id='$r2[review_detail_id]' and object_type='frequent' and object_orientation not in ('N') and
filter_filter='pass' ORDER BY opinion_mining_result_id");
    $j62=mysql_num_rows($stampil62);
    if($j5>0 and $j6>0){
        $tp5++;
    }elseif($j5<1 and $j6<1){
        $tn5++;
    }elseif($j5>0 and $j6<1){
        $fn5++;
    }elseif($j5<1 and $j6>0){
        $fp5++;
    }
    if($j5>0 and $j62>0){
        $tp52++;
    }elseif($j5<1 and $j62<1){
        $tn52++;
    }elseif($j5>0 and $j62<1){
        $fn52++;
    }elseif($j5<1 and $j62>0){
        $fp52++;
    }
}
mysql_query("UPDATE review SET finish_opinion_sentence_detection_frequent = NOW() WHERE
review_id = '$r[review_id]'");
mysql_query("UPDATE review SET finish_opinion_sentence_detection_filtered_frequent = NOW() WHERE
review_id = '$r[review_id]'");
mysql_query("UPDATE review SET start_opinion_sentence_detection_infrequent = NOW() WHERE
review_id = '$r[review_id]'");
mysql_query("UPDATE review SET start_opinion_sentence_detection_filtered_infrequent = NOW() WHERE
review_id = '$r[review_id]'");

```

```

$tp6=0;
$fn6=0;
$fp6=0;
$tn6=0;
$tp62=0;
$fn62=0;
$fp62=0;
$tn62=0;
$stampil2=mysql_query("SELECT * FROM review_detail WHERE review_id='".$r[review_id]' ORDER BY review_detail_id");
$no = 0;
while ($r2=mysql_fetch_array($stampil2)){
    $no++;
    $data=$no-1;
    $batas="";
    $stampil4=mysql_query("SELECT * FROM hotel_corpus ORDER BY hc_id LIMIT $data,1");
    $r4=mysql_fetch_array($stampil4);
    $stampil5=mysql_query("SELECT * FROM hotel_corpus_object where hc_id='".$r4[hc_id]' ORDER BY hca_id");
    $j5=mysql_num_rows($stampil5);
    $stampil6=mysql_query("SELECT * FROM opinion_mining_result where review_detail_id='".$r2[review_detail_id]' and object_type='infrequent' and object_orientation not in ('N') ORDER BY opinion_mining_result_id");
    $j6=mysql_num_rows($stampil6);
    $stampil62=mysql_query("SELECT * FROM opinion_mining_result where review_detail_id='".$r2[review_detail_id]' and object_type='infrequent' and object_orientation not in ('N') and filter_filter='pass' ORDER BY opinion_mining_result_id");
    $j62=mysql_num_rows($stampil62);
    if($j5>0 and $j6>0){
        $tp6++;
    }elseif($j5<1 and $j6<1){
        $tn6++;
    }elseif($j5>0 and $j6<1){
        $fn6++;
    }elseif($j5<1 and $j6>0){
        $fp6++;
    }
    if($j5>0 and $j62>0){
        $tp62++;
    }elseif($j5<1 and $j62<1){
        $tn62++;
    }elseif($j5>0 and $j62<1){
        $fn62++;
    }elseif($j5<1 and $j62>0){
        $fp62++;
    }
}
mysql_query("UPDATE review SET finish_opinion_sentence_detection_infrequent = NOW() WHERE review_id = '".$r[review_id]"');
mysql_query("UPDATE review SET finish_opinion_sentence_detection_filtered_infrequent = NOW() WHERE review_id = '".$r[review_id]"');
$tp53=0;
$accuracy5=round(((($tp5+$tn5)/($tp5+$tn5+$fp5+$fn5))*100,2);
$precision5=round((($tp5/($tp5+$fp5))*100,2);
$recall5=round((($tp5/($tp5+$fn5))*100,2);
$measure5=round(((2*$precision5*$recall5)/($precision5+$recall5)),2);
$accuracy6=round(((($tp6+$tn6)/($tp6+$tn6+$fp6+$fn6))*100,2);
$precision6=round((($tp6/($tp6+$fp6))*100,2);
$recall6=round((($tp6/($tp6+$fn6))*100,2);
$measure6=round(((2*$precision6*$recall6)/($precision6+$recall6)),2);
$accuracy52=round(((($tp52+$tn52)/($tp52+$tn52+$fp52+$fn52))*100,2);
$precision52=round((($tp52/($tp52+$fp52))*100,2);
$recall52=round((($tp52/($tp52+$fn52))*100,2);

```

```

$fmmeasure52=round(((2*$precision52*$recall52)/($precision52+$recall52)),2);
$accuracy62=round(((tp62+$tn62)/(tp62+$tn62+$fp62+$fn62))*100,2);
$precision62=round((tp62/(tp62+$fp62))*100,2);
$recall62=round((tp62/(tp62+$fn62))*100,2);
$fmmeasure62=round(((2*$precision62*$recall62)/($precision62+$recall62)),2);
$accuracy53=round(((tp53+$tn53)/(tp53+$tn53+$fp53+$fn53))*100,2);
$precision53=round((tp53/(tp53+$fp53))*100,2);
$recall53=round((tp53/(tp53+$fn53))*100,2);
$fmmeasure53=round(((2*$precision53*$recall53)/($precision53+$recall53)),2);
$fn53=0;
$fp53=0;
$tn53=0;
$stampil2=mysql_query("SELECT * FROM review_detail WHERE review_id='".$r[review_id]" ORDER BY review_detail_id");
$no = 0;
while ($r2=mysql_fetch_array($stampil2)){
    $no++;
    $data=$no-1;
    $batas="''";
    $stampil4=mysql_query("SELECT * FROM hotel_corpus ORDER BY hc_id LIMIT $data,1");
    $r4=mysql_fetch_array($stampil4);
    $stampil5=mysql_query("SELECT * FROM hotel_corpus_object where hc_id='".$r4[hc_id]" ORDER BY hc_id");
    $j5=mysql_num_rows($stampil5);
    $stampil6=mysql_query("SELECT * FROM opinion_mining_result where review_detail_id='".$r2[review_detail_id]" and object_type='filter' and object_orientation not in ('N') ORDER BY opinion_mining_result_id");
    $j6=mysql_num_rows($stampil6);
    if($j5>0 and $j6>0){
        $tp53++;
    }elseif($j5<1 and $j6<1){
        $tn53++;
    }elseif($j5>0 and $j6<1){
        $fn53++;
    }elseif($j5<1 and $j6>0){
        $fp53++;
    }
}
mysql_query("UPDATE review SET finish_opinion_sentence_detection_filter = NOW() WHERE review_id = '".$r[review_id]"");

```

## E.7. Script Ekspor Data Hasil *Object-based Opinion Mining*

Berikut ini merupakan script yang digunakan untuk tahap ekstraksi *filter object*. Script ini ditulis menggunakan bahasa pemrograman PHP.

```

<?php
function convert_string($string){
    $string2=str_replace('[u]', " ", $string);
    $string2=str_replace('[p]', " ", $string2);
    $string2=str_replace(' ','(*)', $string2);
    $string2=str_replace(' ', " ", $string2);
    $string2=trim($string2);
    $string2=str_replace('(*)', ' ', $string2);
    $string2=str_replace('[+]', 'positive', $string2);
    $string2=str_replace(['-'], 'negative', $string2);
    return $string2;
}
function array2csv(array &$array)
{
    if(count($array) == 0) {

```

```

        return null;
    }
    ob_start();
    $df = fopen("php://output", 'w');
    fputcsv($df, array_keys(reset($array)));
    foreach ($array as $row) {
        fputcsv($df, $row);
    }
    fclose($df);
    return ob_get_clean();
}
function download_send_headers($filename) {
    // disable caching
    $now = gmdate("D, d M Y H:i:s");
    header("Expires: Tue, 03 Jul 2001 06:00:00 GMT");
    header("Cache-Control: max-age=0, no-cache, must-revalidate, proxy-revalidate");
    header("Last-Modified: {$now} GMT");

    // force download
    header("Content-Type: application/force-download");
    header("Content-Type: application/octet-stream");
    header("Content-Type: application/download");

    // disposition / encoding on response body
    header("Content-Disposition: attachment;filename={$filename}");
    header("Content-Transfer-Encoding: binary");
}
include "../config/koneksi.php";
if($_GET['module']=='hotel_summary'){
    $edit=mysql_query("SELECT * FROM review where review_type='hotel' and review_id='".$_GET[id]"");
    $r=mysql_fetch_array($edit);
    $tampil2=mysql_query("SELECT * FROM review_detail WHERE review_id='".$_GET[id]' ORDER BY review_detail_id");
    $no = 0;
    while ($r2=mysql_fetch_array($tampil2)){
        $no++;
        $orientation="";
        $filtered_orientation="";
        $no2=0;
        $no3=0;
        $tampil3=mysql_query("SELECT * FROM opinion_mining_result WHERE
review_detail_id='".$r2[review_detail_id']."' and object_type='frequent' and object_orientation not in ('N')
ORDER BY opinion_mining_result_id");
        while ($r3=mysql_fetch_array($tampil3)){
            $no2++;
            if($no2==1){
                $orientation.=$r3[object_word] [".{$r3[object_orientation]}."];
            }else{
                $orientation.=, {$r3[object_word]} [".{$r3[object_orientation]}."];
            }
            if($no3==0 and $r3['filter_filter']=='pass'){
                $no3++;
                $filtered_orientation.=$r3[object_word] [ $r3[object_orientation] ];
            }elseif($r3['filter_filter']=='pass'){
                $no3++;
                $filtered_orientation.=, {$r3[object_word]} [ $r3[object_orientation] ];
            }
        }
        if($no2==0){
            $orientation='statement';
        }
        if($no3==0){
            $filtered_orientation='statement';
        }
    }
}

```

```

}
$orientation2="";
$filtered_orientation2="";
$no2=0;
$no3=0;
$tampil3=mysql_query("SELECT * FROM opinion_mining_result WHERE
review_detail_id=$r2[review_detail_id] and object_type='infrequent' and object_orientation not in ('N')
ORDER BY opinion_mining_result_id");
while ($r3=mysql_fetch_array($tampil3)){
$no2++;
if($no2==1){
$orientation2.=$r3[object_word] [".{$r3['object_orientation']}."];
}else{
$orientation2.=, $r3[object_word] [".{$r3['object_orientation']}."];
}
if($no3==0 and $r3['filter_filter']=='pass'){
$no3++;
$filtered_orientation2.=$r3[object_word] [ $r3[object_orientation] ];
}elseif($r3['filter_filter']=='pass'){
$no3++;
$filtered_orientation2.=, $r3[object_word] [ $r3[object_orientation] ];
}
}
if($no2==0){
$orientation2='statement';
}
if($no3==0){
$filtered_orientation2='statement';
}
$orientation3="";
$no2=0;
$no3=0;
$tampil3=mysql_query("SELECT * FROM opinion_mining_result WHERE
review_detail_id=$r2[review_detail_id] and object_type='filter' and object_orientation not in ('N') ORDER
BY opinion_mining_result_id");
while ($r3=mysql_fetch_array($tampil3)){
$no2++;
if($no2==1){
$orientation3.=$r3[object_word] [".{$r3['object_orientation']}."];
}else{
$orientation3.=, $r3[object_word] [".{$r3['object_orientation']}."];
}
}
if($no2==0){
$orientation3='statement';
}
if($r['review_name']=='hotel corpus'){
$data=$no-1;
$corpus_data="";
$no2=0;
$tampil4=mysql_query("SELECT * FROM hotel_corpus ORDER BY hc_id LIMIT $data,1");
while ($r4=mysql_fetch_array($tampil4)){
$nox=0;
$tampil5=mysql_query("SELECT * FROM hotel_corpus_object where hc_id='".$r4[hc_id]' ORDER BY
hca_id");
while ($r5=mysql_fetch_array($tampil5)){
$nox++;
$no2++;
if($nox==1){
$corpus_data.=$r5["hca_object"].['!'.$r5["hca_orientation"].'];
}else{
$corpus_data.=, '$r5["hca_object"].['!'.$r5["hca_orientation"].'];
}
}
}
}

```

```

if($r5["hca_symbol"]!=""){
    $corpus_data.=[$r5["hca_symbol"].];
}
}
$corpus_data="";
if($no2==0){
    $corpus_data='statement';
}
}
$review_detail_id[$data] = $r2['review_detail_id'];
$review_sentence[$data] = $r2['review_sentence'];
$corpus_object[$data] = convert_string($corpus_data);
$frequent_object[$data] = convert_string($orientation);
$infrequent_object[$data] = convert_string($orientation2);
$filtered_frequent_object[$data] = convert_string($filtered_orientation);
$filtered_infrequent_object[$data] = convert_string($filtered_orientation2);
$filter_object[$data] = convert_string($orientation3);
}
$j_m = $no;
$xlsRow = 1;
header('Content-Type: text/csv; charset=utf-8');
header('Content-Disposition: attachment; filename=hotel_result.csv');
$output = fopen('php://output', 'w');
fputcsv($output, array('review detail id', 'review sentence', 'corpus object', 'frequent object', 'infrequent object', 'filtered frequent object', 'filtered infrequent object', 'filter object'));
$j_m = $no;
$xlsRow = 1;
for ($y=0; $y<$j_m; $y++) {
    ++
    $i;
    fputcsv($output, array($review_detail_id[$y], trim($review_sentence[$y]), $corpus_object[$y], $frequent_object[$y], $infrequent_object[$y], $filtered_frequent_object[$y], $filtered_infrequent_object[$y], $filter_object[$y]));
    $xlsRow++;
}
}

```

## **BAB 5**

### **KESIMPULAN DAN SARAN**

#### **5.1. Kesimpulan**

Setelah penelitian mengenai pemanfaatan filter pada *object-based opinion mining* yang telah selesai dilaksanakan, maka berikut ini merupakan kesimpulan dari penelitian ini.

- a. Penerapan filter dalam proses ekstraksi objek jelas dapat meningkatkan presisi dengan cukup meyakinkan yaitu pada pendekatan *frequent object* yang semula 45.7% menjadi 64.49% pada data *review hotel* dan dari 44.82% menjadi 64.61% pada data *review restoran*, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 44.35% menjadi 43.52% pada data *review hotel* dan dari 45.4% menjadi 45.11% pada data *review restoran*. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan *frequent object* dan *infrequent object* dimana pada kedua data *review* terjadi peningkatan hasil presisi yang cukup meyakinkan yaitu yang semula 22.33% menjadi 63.02% pada data *review hotel* dan dari 21.6% menjadi 65.4% pada data *review restoran*, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 59.97% menjadi 55.95% pada data *review hotel* dan dari 61.06% menjadi 59.2% pada data *review restoran*.
- b. Penggunaan filter dalam *object-based opinion mining* mendapatkan hasil yang dominan secara keseluruhan dibanding dengan tanpa filter pada proses klasifikasi sentimen objek. Hal ini terbukti saat dilakukan pengukuran performa menggunakan semua objek yang telah berhasil diekstrak pada semua pendekatan yang telah dilakukan. Hasil tertinggi didapatkan pada pendekatan *frequent object* dengan filter, yaitu 56.85% akurasi, 60.91% presisi, 79.93% recall, 69.14% f-measure pada data *review hotel*, serta mendapatkan hasil 58.85% akurasi, 63.26% presisi, 84.14% recall, 72.22% f-measure pada data *review restoran*.

- c. Penggunaan data filter secara langsung dalam *object-based opinion mining* sebenarnya dapat mendapatkan hasil yang cukup bagus, akan tetapi efek negatif yang ditimbulkan adalah proses analisis menghabiskan waktu yang jauh lebih lama dibandingkan dengan pendekatan yang lain contohnya yaitu pada proses ekstraksi objek waktu yang dibutuhkan mencapai 3.051 detik pada data hotel dan 1.559 detik pada data restoran, padahal pendekatan lain hanya memakan waktu tidak lebih dari 4 detik untuk melakukan ekstraksi objek.
- d. Pada proses klasifikasi sentimen kalimat dapat ditarik kesimpulan bahwa penggunaan objek terekstraksi sebagai atribut klasifikasi untuk menentukan orientasi kalimat mendapatkan hasil yang lebih baik dibanding dengan hanya menggunakan kalimat *review* secara langsung sebagai atribut klasifikasi. Walaupun hasil objek dari *object-based opinion mining* masih kurang dibandingkan dengan penggunaan aspek corpus sebagai atribut, akan tetapi hasil pemanfaatan *object-based opinion mining* dalam klasifikasi sentimen kalimat menghasilkan lebih baik dibanding menggunakan kalimat *review*. Selain itu yang paling penting adalah hasil dari proses klasifikasi menggunakan objek dari *object-based opinion mining* dengan filter mendapatkan hasil yang lebih baik dibanding dari tanpa menggunakan data filter sama sekali. Secara keseluruhan tidak ada teknik klasifikasi yang mendapatkan hasil yang dominan lebih baik dibanding dengan yang lain, tetapi teknik klasifikasi yang lebih banyak mendapatkan hasil yang lebih baik adalah teknik klasifikasi *Random Forest*, dimana pada hasil klasifikasi dengan atribut data *frequent object* dan *infrequent object* dengan filter, Random Forest mendominasi seluruh hasil dibanding dengan Naïve Bayes dan SVM. Pada data *review* hotel Random Forest mendapatkan hasil 61.34% akurasi, 62.2% presisi, 61.3% recall, 57.1% f-measure dan pada data *review* restoran mendapatkan hasil 64.68% akurasi, 65.5% presisi, 64.7% recall, 61.7% f-measure.

## **5.2. Saran**

Pada penelitian ini memang peran dari data filter dapat meningkatkan hasil *object-based opinion mining*, akan tetapi dalam penelitian ini masih banyak yang bisa digali yaitu seperti belum adanya proses untuk menganalisa objek yang secara implisit terdapat dalam data *review*, karena dalam konteks data *review* pariwisata cukup banyak opini terhadap objek yang disampaikan secara implisit.

Belum adanya konsep dan landasan ilmiah yang pasti dalam menyusun dan menentukan data filter yang sesuai, sehingga proses penyusunan data filter yang digunakan dalam penelitian ini dilakukan secara manual oleh peneliti, yaitu berdasarkan data ontologi yang didapat dari internet dan juga berdasarkan data aspek yang ada pada data corpus. Oleh karena itu apabila ada peneliti lain mencoba melakukan penelitian serupa maka kemungkinan besar data filter yang dihasilkan akan berbeda. Hal ini termasuk merupakan kelemahan yang ada dalam penelitian ini dan diharapkan pada penelitian selanjutnya dapat diperbaiki dengan adanya landasan ilmiah yang pasti.

Selain itu permasalahan lain adalah belum adanya proses koreksi untuk menangani penulisan kata yang salah. Permasalahan yang menjadi kekurangan pada penelitian ini diharapkan dapat menjadi masukan bagi penelitian mendatang sehingga perbaikan dari proses analisis *object-based opinion mining* akan menghasilkan performa yang lebih baik. Tentu saja meningkatnya performa *object-based opinion mining* akan berdampak meningkat pula hasil klasifikasi sentimen kalimat *review* untuk mendapatkan orientasi *review* secara global.

## DAFTAR PUSTAKA

- Archak, N., Ghose, A., & Ipeirotis, P. (2007). *Show me the money!: Deriving the pricing power of product features by mining consumer reviews*. In Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 56–65). ACM.
- Bucur, C. (2014). *Opinion Mining Platform for Intelligence in Business, Economic Insights-Trends and Challenges*, Volume 3, No 3/ p. 99-108, ISSN 2284-8576
- Bucur, C. (2015). *Using Opinion Mining Techniques in Tourism*. Procedia Economics and Finance 23, 1666 – 1673.
- Decker, R., & Trusov, M. (2010). *Estimating aggregate consumer preferences from online product reviews*. International Journal of Research in Marketing, 27(4), 293–307.
- Ding, X., Liu, B., & Yu, P. (2008). *A holistic lexicon-based approach to opinion mining*. In Proceedings of the international conference on Web search and web data mining (pp. 231–240). ACM.
- Domingos, P., Pazzani, M. J. 1997. *On the optimality of the simple bayesian classifier under zeroone loss*. *Machine Learning*, 29(2/3):103–130.
- Dragut, E. F. (2009). Stop word and related problems in web interface integration. *Proceedings of the VLDB Endowment*, 2(1), 349-360.
- Gaia.fdi.ucm.es. *restaurant*. Retrieved June 1, 2016, from gaia.fdi.ucm.es: <http://gaia.fdi.ucm.es/ontologies/restaurant.owl>
- Haddi, E. L. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17, 26-32.
- Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 168–177). ACM.
- Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. In Proceedings of the national conference on artificial intelligence (pp. 755–

- 760). Menlo Park, CA; Cambridge, MA; London: AAAI Press. MIT Press; 1999.
- Hu, M., & Liu, B. (2006). Opinion extraction and summarization on the web. Proceedings of the national conference on artificial intelligence (Vol. 21, pp. 1621). Menlo Park, CA; Cambridge, MA; London: AAAI Press. MIT Press; 1999. Menlo Park, CA; Cambridge, MA; London: AAAI Press. MIT Press; 1999.
- James, G. e. (2014). *An Introduction to Statistical Learning*. London, United Kingdom: Springer.
- Jian Sheu, J. (2008). An Efficient Two-Phase Spam Filtering Method. Sciendirect.
- Kim, H., Ganesan, K., Sondhi, P., & Zhai, C. (2011). Comprehensive review of opinion summarization.
- Ku, L.-W., Liang, Y.-T., & Chen, H.-H. (2006). *Opinion extraction, summarization and tracking in news and blog corpora*. In AAAI spring symposium: Computational approaches to analyzing weblogs (pp. 100–107).
- Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer Verlag. 107
- Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. Proceedings of the 14th international conference on World Wide Web (pp. 342–351). ACM.
- Lu, Y., Zhai, C., & Sundaresan, N. (2009). *Rated aspect summarization of short comments*. In Proceedings of the 18th international conference on World Wide Web (pp. 131–140). ACM.
- Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F., & Matsuo, Y. (2013). *Identifying customer preferences about tourism products using an aspect-based opinion mining approach*. In Proceedings of the knowledge-based and intelligent information and engineering systems (pp. 182–191). KES.
- Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F. (2014). *A novel deterministic approach for aspect-based opinion mining in tourism products reviews*. Expert Systems with Applications.
- Ma'ady, M. D. (2014). Implementation of Text Classification for Twitter Data Mapping of Telkomsel Customers in The Form of Heat Map. *International*

- Research Conference on Business, Economics, and Social Science*, (pp. 66-76).
- Naradhipa, A. R. (2012). Sentiment classification for Indonesian message in social media. In *Cloud Computing and Social Networking (ICCCSN)* (pp. 1-5). IEEE.
- Park, D., & Kim, S. (2009). *The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews*. Electronic Commerce Research and Applications, 7(4), 399–410.
- Popescu, A., & Etzioni, O. (2005). *Extracting product features and opinions from reviews*. In Proceedings of the conference on human language technology and empirical methods in natural language processing (pp. 339–346). Association for Computational Linguistics.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In Encyclopedia of database systems (pp. 532-538). USA: Springer.
- Rong Jia, Li Gang, Chen Yi-Ping P. (2009). *Acoustic Feature Selection For Automatic Emotion Recognition From Speech, Information Processing and Management* 45, 315–328.
- Semanticweb.org. *accommodation*. Retrieved June 1, 2016, from semanticweb.org: <http://www.semanticweb.org/ontologies/2012/4/accommodation-17042012.owl>
- Tala, a. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Netherlands: Master of Logic Project. Institute for Logic, Language and Computation, Universiteit van Amsterdam.
- Titov, I., & McDonald, R. (2008). *A joint model of text and aspect ratings for sentiment summarization*. Urbana, 51, 61801.
- Weiss, S. M. (2010). *Text mining: predictive methods for analyzing unstructured information*. New York: Springer.
- Zhao, L., & Li, C. (2009). *Ontology based opinion mining for movie reviews*. Knowledge Science, Engineering and Management, 204–214. 1117
- Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). *Movie review mining and summarization*. In Proceedings of the 15th ACM international conference on information and knowledge management (pp. 43–50). ACM.

Zhu, F., & Zhang, X. (2010). *Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics*. Journal of Marketing, 74(2), 133–148

## BIOGRAFI PENULIS



**Bonda Sisepaputra**, lahir di Kota Tuban pada tanggal 10 Maret 1988. Penulis merupakan anak kedua dari tiga bersaudara. Memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi Stikom Surabaya, lulus pada tahun 2011. Pada tahun 2013 penulis diterima menjadi Mahasiswa Program Pasca Sarjana Magister Teknik Informatika bidang Keahlian Sistem Informasi di Institut Teknologi Sepuluh Nopember. Penulis lulus dan memperoleh gelar Magister Komputer (M.Kom) pada bulan September tahun 2016.

Penulis dapat dihubungi melalui *e-mail* di:

*bonda.sisepaputra@gmail.com*