



TESIS BM185407

**ANALISIS OPINI ULASAN PELANGGAN HOTEL DENGAN  
MEMANFAATKAN ANALISIS SENTIMEN MENGGUNAKAN  
METODE SUPPORT VECTOR MACHINE**

**SITI AZZA AMIRA  
09211750053019**

Dosen Pembimbing:  
Prof. Dr. Ing, Drs. M. Isa Irawan, MT

Departemen Manajemen Teknologi  
Fakultas Bisnis dan Manajemen Teknologi  
Institut Teknologi Sepuluh Nopember  
2019

LEMBAR  
PENGESAHAN

## LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar

**Magister Manajemen Teknologi (M.MT)**

di

**Institut Teknologi Sepuluh Nopember**

Oleh:

**Siti Azza Amira**

**NRP: 09211750053019**

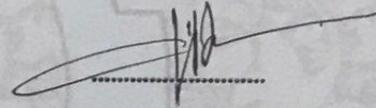
**Tanggal Ujian: 11 Juli 2019**

**Periode Wisuda: September 2019**

Disetujui oleh:

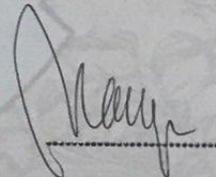
**Pembimbing:**

1. Prof. Dr. Ing, Drs. M. Isa Irawan, MT  
NIP: 196312251989031001



**Penguji:**

1. Dr. Tech. Ir. R. V. Hari Ginardi, M.Sc  
NIP: 196505181992031003



2. Dr. Sutikno, S.Si, M.Si  
NIP: 197103131997021001



**Kepala Departemen Manajemen Teknologi**

**Fakultas Bisnis dan Manajemen Teknologi**



**Prof. Ir. Nyoman Pujawan, M.Eng, Ph.D, CSCP**  
NIP: 196912311994121076

# ABSTRAK

# **ANALISIS OPINI ULASAN PELANGGAN HOTEL DENGAN MEMANFAATKAN ANALISIS SENTIMEN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE**

Nama : Siti Azza Amira  
NRP : 09211750053019  
Dosen Pembimbing : Prof. Dr.Ing, Drs. M. Isa Irawan, MT

## **ABSTRAK**

Perkembangan teknologi sekarang ini membuat banyak orang semakin mudah dalam mengakses informasi, salah satunya untuk mencari informasi terkait sebuah tempat. Banyak calon pengunjung yang akan membaca ulasan dari orang yang pernah mengunjungi suatu tempat untuk mengetahui bagaimana penilaian mereka terhadap tempat tersebut. Opini pada ulasan orang lain sangat berpengaruh di dalam mempengaruhi keputusan orang lain dalam menilai sebuah tempat yang ingin dikunjungi. Analisis opini dapat dilakukan dengan melakukan analisis sentimen terhadap ulasan pelanggan hotel. Data ulasan diambil dengan melakukan crawling pada situs TripAdvisor dan selanjutnya data di pre-processing serta dilakukan pembobotan kata menggunakan metode TF-IDF. Klasifikasi dilakukan dengan menggunakan metode *support vector machine* untuk mengetahui opini dari ulasan yaitu positif atau negatif. Berdasarkan hasil klasifikasi, Hotel dengan ulasan sentimen positif terbanyak di kota Surabaya adalah Harris Hotel Gubeng dan Pop! Hotel Gubeng dengan jumlah ulasan sama yaitu 252 ulasan. Sedangkan ulasan sentimen positif terbanyak di kota Malang adalah Harris Hotel Malang dengan 311 ulasan. Berdasarkan hasil perhitungan rating, hotel yang mendapatkan peringkat pertama di kota Surabaya adalah Harris Hotel Gubeng Surabaya dengan nilai rating 4,65 (luar biasa) dan peringkat pertama di kota Malang adalah The 101 OJ Hotel Malang dengan nilai rating 4,9 (luar biasa).

*Kata Kunci : Analisis Sentimen, TF-IDF, Klasifikasi, Support Vector Machine*

# ABSTRACT

# **OPINION ANALYSIS OF HOTEL CUSTOMER REVIEW UTILIZE SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE METHOD**

Name : Siti Azza Amira  
NRP : 09211750053019  
Supervisor : Prof. Dr. Ing, Drs. M. Isa Irawan, MT

## **ABSTRACT**

Technology development nowadays make it easier for people to access information, one of them is to find information regarding a place. Many prospective visitors will read reviews from people who have visited a place to find out how they they rate a place. Opinion on other people's reviews is very influential in influencing the decisions of others in assessing a place they want to visit. Opinion analysis can be done by conducting a sentiment analysis towards hotel customer reviews. Review data was taken by crawling on TripAdvisor site and then data will be pre-processing and term weighted using the TF-IDF method. Classification process using support vector machine method to find out opinions from reviews that are positive or negative. Based on the results of the classification, Hotels with the most positive sentiment reviews in Surabaya are Harris Hotel Gubeng and Pop! Hotel Gubeng with the same number of reviews, 252 reviews. While the most positive sentiment review in Malang is Harris Hotel Malang with 311 reviews. Based on the results of rating calculation, hotels that get the first rank in Surabaya are Harris Hotel Gubeng Surabaya with a rating of 4.65 (excellent) and in Malang is The 101 OJ Malang Hotel with a rating of 4.9 (excellent).

*Keywords : Sentiment Analysis, TF-IDF, Classification, Support Vector Machine*

# KATA PENGANTAR

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT, karena berkat rahmat dan ridho-Nya penulis dapat menyelesaikan tesis yang merupakan persyaratan dalam menyelesaikan Magister pada Manajemen Teknologi Informasi di MMT-ITS. Tesis ini merupakan implemetasi dari teori-teori tentang perancangan dan analisis yang didapat oleh penulis diperkuliahan maupun secara mandiri. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada berbagai pihak :

1. Allah SWT yang telah memberikan kelancaran, kemudahan serta petunjuk kepada penulis.
2. Kedua orang tua penulis, adik Ahmad Indy Azrian dan keluarga besar yang selalu memberikan semangat, dukungan, serta doa.
3. Prof. Dr.Ing, Drs. M. Isa Irawan, MT selaku pembimbing penulis pada penelitian ini yang telah banyak membimbing dan memberikan arahan.
4. Dosen-dosen di MMT-ITS Surabaya yang sudah berkenan untuk memberikan ilmunya.
5. Teman-teman dekat Penulis I Wayan Vendy Wiranatha, Anindya Pramitha Harviyanti, Eka Novita Shandra, dan Rangga Dinata yang selalu menemani, membantu proses pengerjaan tesis ini dan terus memberikan semangat.
6. Teman-teman seperjuangan MTI 2017 dan terkhusus kepada Septi, Amrina, Afrianda, Reza, Giri, Desi dan Dike atas segala bantuannya dalam menjalani masa kuliah. Juga untuk Dewi dan Aya yang membantu dalam pemahaman teori-teori yang berkaitan dengan penelitian ini.
7. Serta semua pihak yang telah membantu dalam penyelesaian tesis ini.

Tesis ini mungkin masih sangat jauh dari kata sempurna. Sehingga penulis akan menerima dengan senang hati jika ada pihak yang memberikan kritik dan saran demi pengembangan kedepannya. Akhir kata, penulis berharap bahwa tesis ini dapat bermanfaat bagi seluruh pihak.

Surabaya, Juli 2019

Penulis

DAFTAR ISI, DAFTAR  
GAMBAR & DAFTAR  
TABEL

## DAFTAR ISI

<b>ABSTRAK .....</b>	<b>V</b>
<b>ABSTRACT .....</b>	<b>VII</b>
<b>KATA PENGANTAR.....</b>	<b>IX</b>
<b>DAFTAR ISI.....</b>	<b>XI</b>
<b>DAFTAR GAMBAR.....</b>	<b>XV</b>
<b>DAFTAR TABEL .....</b>	<b>XVII</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	4
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	4
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan .....	5
<b>BAB 2 KAJIAN PUSTAKA .....</b>	<b>8</b>
2.1 Studi Literatur.....	8
2.2 Opini.....	11
2.2.1 Komponen Opini.....	12
2.3 Kepuasan Pelanggan.....	12
2.4 TripAdvisor .....	13
2.5 Text Mining .....	15
2.5.1 Case Folding .....	17
2.5.2 Tokenizing.....	17
2.5.3 Filtering .....	18

2.5.4 Stemming.....	18
2.6 Pembobotan TF-IDF.....	18
2.7 Support Vector Machine.....	20
2.7.1 Support Vector Machine Non-Linear.....	21
2.7.2 Sequential Training SVM.....	22
<b>BAB 3 METODOLOGI PENELITIAN .....</b>	<b>25</b>
3.1 Tahap Awal .....	27
3.1.1 Studi Literatur.....	27
3.1.2 Pengumpulan data .....	27
3.2 Tahap Pengembangan.....	29
3.2.1 Pre-processing Data.....	29
3.2.2 Pembobotan Kata.....	29
3.2.3 Klasifikasi Sentimen.....	30
3.3 Tahap Akhir.....	31
3.3.1 Analisis Hasil dan Pembahasan.....	31
3.3.2 Kesimpulan dan Saran.....	32
<b>BAB 4 HASIL PENELITIAN.....</b>	<b>34</b>
4.1 Data Ulasan .....	34
4.2 Pre-processing Data.....	38
4.2.1 Case Folding.....	39
4.2.2 Tokenizing.....	40
4.2.3 Filtering .....	40
4.2.4 Stemming.....	41
4.3 Proses Pembobotan Kata .....	42
4.3.1 Menghitung nilai TF.....	42
4.3.2 Menghitung nilai IDF.....	43

4.3.3 Menghitung nilai TF*IDF .....	44
4.4 Proses Perhitungan Support Vector Machine .....	45
4.4.1 Perhitungan Kernel.....	45
4.4.2 Melakukan Inisialisasi Parameter .....	46
4.4.3 Menghitung Matriks Hessian .....	46
4.4.3 Menghitung iterasi, $\delta ai$ dan $ai$ .....	46
4.4.4 Pemilihan Support Vector .....	48
4.4.5 Perhitungan Nilai Bias .....	48
4.4.6 Perhitungan Testing .....	49
4.5 Hasil Klasifikasi Support Vector Machine .....	49
4.5.1 Hasil Klasifikasi pada Ulasan di Kota Surabaya.....	50
4.5.2 Hasil Klasifikasi pada Ulasan di Kota Malang .....	50
4.6 Hasil Klasifikasi Support Vector Machine Berdasarkan Aspek .....	51
4.6.1 Analisis Hasil Klasifikasi pada Ulasan Berdasarkan Aspek di Kota Surabaya.....	51
4.6.2 Analisis Hasil Klasifikasi pada Ulasan Berdasarkan Aspek di Kota Malang.....	53
4.7 Hasil Perbandingan Rating Klasifikasi Sentimen dengan Rating pada Situs TripAdvisor.....	55
4.7.1 Analisis Hasil Perbandingan Rating Klasifikasi Sentimen dengan Rating pada Situs TripAdvisor di Kota Surabaya.....	56
4.7.1 Analisis Hasil Perbandingan Rating Klasifikasi Sentimen dengan Rating pada Situs TripAdvisor di Kota Malang .....	56
4.8 Hasil Evaluasi.....	64
4.8.1 Hasil Evaluasi Data Ulasan Hotel di Kota Surabaya .....	64
4.8.2 Hasil Evaluasi Data Ulasan Hotel di Kota Malang.....	65
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>	<b>68</b>

5.1	Kesimpulan .....	68
5.2	Saran .....	68
	<b>DAFTAR PUSTAKA .....</b>	<b>71</b>
	<b>LAMPIRAN 1 .....</b>	<b>75</b>
	<b>LAMPIRAN 2 .....</b>	<b>81</b>
	<b>LAMPIRAN 3 .....</b>	<b>87</b>
	<b>LAMPIRAN 4 .....</b>	<b>91</b>

## DAFTAR GAMBAR

Gambar 1.1 Aspek pada Situs Tripadvisor (www.tripadvisor.co.id).....	3
Gambar 2.1 Logo Tripadvisor.....	13
Gambar 2.2 Situs Tripadvisor .....	14
Gambar 2.3 Ulasan pada situs TripAdvisor Analisis Sentimen.....	14
Gambar 2.4 Diagram Venn Interaksi Text Mining dengan Bidang Lain.....	15
Gambar 2.5 Hubungan Antara Berbagai Teknik Dalam Text Mining.....	16
Gambar 2.6 Tahapan Pada Text Mining .....	17
Gambar 2.7 <i>Hyperlane Support Vector Machine</i> .....	20
Gambar 3.1 Tahapan Penelitian .....	26
Gambar 3.2 Alur Proses Klasifikasi Menggunakan <i>Support Vector Machine</i> .....	30
Gambar 4.1 Grafik Hasil Klasifikasi SVM pada Ulasan Hotel di Kota Surabaya	50
Gambar 4.2 Grafik Hasil Klasifikasi SVM pada Ulasan Hotel di Kota Malang ..	51
Gambar 4.3 Penilaian Rating Pada TripAdvisor.....	56
Gambar 4.4 Grafik Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Surabaya .	65
Gambar 4.5 Grafik Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Malang ....	66

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

Tabel 2.1 Kajian Penelitian Terdahulu.....	8
Tabel 3.1 Kriteria Aspek Pelayanan yang Digunakan pada Penelitian.....	28
Tabel 3.2 <i>Confusion Matrix</i> .....	31
Tabel 4.1 <i>Coding</i> Scrapy pada Python untuk Crawling Data .....	34
Tabel 4.2 Data Ulasan Mentah Dari Situs TripAdvisor.....	36
Tabel 4.3 Data Ulasan yang Sudah di <i>Split</i> dan Diberikan ID.....	37
Tabel 4.4 Data Ulasan yang Sudah Dilabeli Menurut Aspeknya. ....	38
Tabel 4.5 Data Ulasan Sesudah Melalui Tahap <i>Case Folding</i> .....	39
Tabel 4.6 Data Ulasan Sesudah Melalui Tahap <i>Tokenizing</i> .....	40
Tabel 4.7 Data Ulasan Sesudah Melalui Tahap <i>Filtering</i> .....	41
Tabel 4.8 Data Ulasan Sesudah Melalui Tahap <i>Stemming</i> .....	41
Tabel 4.9 Hasil Perhitungan Nilai <i>Term Frequency</i> (TF).....	42
Tabel 4.10 Hasil Perhitungan Nilai <i>Inverse Document Frequency</i> (IDF) .....	43
Tabel 4.11 Hasil Perhitungan Nilai Bobot TF-IDF.....	44
Tabel 4.12 Hasil Perhitungan Kernel <i>polynomial</i> .....	45
Tabel 4.13 Nilai Inisialisasi Parameter yang Digunakan .....	46
Tabel 4.14 Hasil Perhitungan Matriks Hessian.....	46
Tabel 4.15 Hasil Perhitungan Nilai E pada Iterasi ke-0.....	47
Tabel 4.16 Hasil Perhitungan Nilai $\delta\alpha$ pada Iterasi ke-0.....	47
Tabel 4.17 Hasil Perhitungan Nilai $\alpha$ pada Iterasi ke-0.....	47
Tabel 4.18 Hasil Perhitungan Nilai $\alpha$ pada Iterasi ke-3.....	48
Tabel 4.19 Hasil Dari Pemilihan $\alpha_i$ Terbesar dari Masing-Masing Kelas .....	48
Tabel 4.20 Hasil Perhitungan Nilai Bias.....	48
Tabel 4.21 Hasil Data Testing.....	49
Tabel 4.22 Hasil Klasifikasi Sentimen Berdasarkan Aspek Di Kota Surabaya....	52
Tabel 4.23 Hasil Klasifikasi Sentimen Berdasarkan Aspek Di Kota Malang.....	54
Tabel 4.24 Range Nilai Pada Situs TripAdvisor.....	56
Tabel 4.25 Hasil Perbandingan Rating Klasifikasi Sentimen dan rating TripAdvisor di 5 hotel kota Surabaya.....	56

Tabel 4.26 Peringkat Berdasarkan Hasil Rating Klasifikasi Sentimen di Kota Surabaya .....	59
Tabel 4.27 Hasil Perbandingan Rating Klasifikasi Sentimen dan rating TripAdvisor pada 5 hotel di kota Malang .....	60
Tabel 4.28 Peringkat Berdasarkan Hasil Rating Klasifikasi Sentimen di Kota Malang .....	64
Tabel 4.29 Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Surabaya .....	65
Tabel 4.30 Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Malang.....	66

# BAB I

## PENDAHULUAN



# **BAB 1**

## **PENDAHULUAN**

Bab ini menjelaskan beberapa hal dasar dalam penelitian yang dilakukan meliputi latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

### **1.1 Latar Belakang**

Perkembangan internet sekarang ini sangat pesat. Pemanfaatan internet sudah mencakup di dalam banyak bidang seperti sebagai media untuk mendapat berbagai macam informasi, media untuk menjalankan bisnis (berjualan secara online maupun berbelanja online), media untuk mendapatkan hiburan (musik, film, majalah), hingga untuk mengakses media sosial seperti instagram, twitter dan facebook. Sekarang ini banyak orang yang menggunakan internet untuk mempermudah mereka dalam memperoleh informasi terkait hotel. Semakin banyaknya pilihan hotel yang dapat dituju saat berlibur akan membuat orang-orang kebingungan untuk menentukan hotel mana yang cocok untuk mereka pesan. Ada banyak situs yang menyajikan informasi terkait hotel diberbagai negara, salah satunya adalah Tripadvisor. Berdasarkan data yang dirilis Skift pada tahun 2013 dengan melihat statistik dari similiarweb dalam melihat traffic pengunjung selama bulan oktober 2013, Tripadvisor berada diperingkat dua dari 10 situs perjalanan online di dunia yang paling banyak diakses dan dimanfaatkan oleh banyak orang dengan total 48,5 juta pengunjung<sup>1</sup>. Pada laman situs ini dapat ditemukan banyak informasi seputar hotel, restoran hingga tempat wisata di berbagai negara beserta ulasan dari orang-orang yang sudah pernah berkunjung ke tempat tersebut. Pada fitur ulasan inilah orang-orang yang pernah mengunjungi suatu hotel, restoran atau tempat wisata dapat menuliskan opininya tentang tempat tersebut. Ulasan ini tentunya sangat berpengaruh dalam pengambilan keputusan calon pengunjung selanjutnya dalam menentukan destinasi tempat yang ingin mereka kunjungi.

Indonesia adalah salah satu negara yang terkenal dengan keindahan alam dan budayanya. Banyak sekali tempat wisata di seluruh indonesia yang dapat dikunjungi wisatawan lokal maupun wisatawan asing. Salah satu provinsi yang banyak dikunjungi

---

<sup>1</sup> <https://travel.kompas.com/read/2013/11/14/1450559/10.Situs.Perjalanan.Online.Paling.Populer>

untuk berlibur adalah Jawa Timur. Ada banyak kota/kabupaten di Jawa Timur yang dapat dijadikan tujuan berwisata dikarenakan banyaknya keindahan alamnya seperti pantai, pegunungan, dan air terjun yang banyak terdapat di beberapa wilayahnya. Selain itu juga banyak wisata buatan dan wisata budaya yang ada di Jawa Timur. Bahkan menurut kepala Dinas Kebudayaan dan Pariwisata Jawa Timur, Sinarito S.Kar, MM. Pada tahun 2017 jumlah kunjungan wisatawan lokal ke Jawa Timur sebanyak 65.623.535, angka ini mengalami kenaikan sebesar 13,01% dari tahun 2016 dimana jumlah kunjungan wisatawan lokal ke Jawa Timur sebanyak 58.068.493. Sedangkan jumlah kunjungan wisatawan asing ke Jawa Timur pada tahun 2017 sebanyak 690.509 meningkat sebanyak 11,62% dibandingkan tahun 2016 yang jumlah kunjungan wisatawan asingnya sebanyak 618.651<sup>2</sup>. Dilihat dari banyaknya jumlah wisatawan yang tertarik untuk berlibur di Jawa Timur tentunya juga semakin membuka peluang bagi pengelola hotel dalam menarik wisatawan untuk menginap di hotel yang mereka kelola. Dikarenakan hal tersebut maka penting bagi pengelola hotel untuk lebih memperhatikan bagaimana ulasan dari orang-orang yang pernah berkunjung ke hotel yang mereka miliki dan kelola agar dapat mempertahankan maupun meningkatkan jumlah kunjungan.

Berdasarkan hasil penelitian Rimba Nuzulul Chory, Muhammad Nasrun dan Casi Setianingsih (Chory, R., N., et al, 2018) perlu adanya analisis terhadap komentar dari pengguna untuk melihat bagaimana kepuasan publik dalam menggunakan layanan. Penelitian juga dilakukan oleh Dewi Ayu dan Riyanarto Sarno (Khotimah D., A., K., Riyanarto Sarno. 2018) pada penelitian dengan judul “Sentimen Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis” untuk melakukan analisis terhadap ulasan data pelanggan pada situs booking.com dengan menggunakan metode PLSA.

Pada fitur ulasan diberbagai situs biasanya tidak dapat dilakukan pemilihan untuk menampilkan mana ulasan yang bersifat positif (ulasan dengan opini baik) atau yang bersifat negatif (ulasan dengan opini buruk), Sehingga dapat dilakukan analisis sentimen untuk mengetahui bagaimana ulasan pelanggan hotel pada situs tersebut. Ulasan yang dituliskan oleh pelanggan tentunya ditulis berdasarkan apa yang mereka rasakan disaat menulis ulasan tersebut sehingga ulasan tersebut dapat dikatakan jujur tanpa adanya

---

<sup>2</sup> <http://disbudpar.jatimprov.go.id/read/umum/anugerah-wisata-jawa-timur-2018>

rekayasa. Analisis sentimen adalah sebuah ilmu yang mempelajari bagaimana untuk melakukan analisa terhadap opini, sentimen, evaluasi, penilaian, sikap dan emosi dari sebuah entitas dimana hal tersebut dapat berupa produk, pelayanan, individu, isu-isu, peristiwa, organisasi, dan topik (Liu, 2012). Analisis sentimen dapat diterapkan pada beberapa jenis tingkatan yang berbeda entah itu teks yang berupa dokumen ataupun kalimat. Ulasan dari wisatawan dipengaruhi oleh emosi (sentimen) sehingga dapat dikelompokkan atau diklasifikasi agar dapat ditentukan kepolarisasiannya, yaitu positif atau negatif (Indriati, Ridok, A, 2016).

Terdapat banyak metode yang dapat digunakan untuk melakukan klasifikasi teks, Salah satu metode yang dapat digunakan adalah *Support Vector Machine (SVM)*. Metode *Support Vector Machine* adalah salah satu metode yang dapat digunakan untuk melakukan klasifikasi dimana metode ini telah banyak diterapkan untuk menyelesaikan berbagai macam permasalahan dalam banyak bidang, baik dalam masalah gene expression analysis, finansial, cuaca hingga di bidang kedokteran. Pada bidang analisis sentimen sendiri implementasi metode *support vector machine* banyak digunakan dikarenakan lebih dapat memberikan hasil yang lebih baik dari pada metode klasifikasi sejenis seperti *Artificial Neural Network (ANN)* terutama di dalam menemukan solusi dikarenakan SVM dapat menemukan solusi yang global optimal (Santosa, 2015).

Berdasarkan latar belakang yang telah dipaparkan tersebut maka pada penelitian ini dilakukan analisis sentimen untuk mengetahui opini dari ulasan yang di tulis oleh pelanggan hotel pada situs TripAdvisor. Selain itu ulasan tersebut juga akan di analisa berdasarkan aspek mana ulasan tersebut dapat dikategorikan. Pada Gambar 1.1 dapat dilihat tiga aspek yang digunakan dalam melakukan pengkategorian ulasan, Aspek tersebut berasal dari penilaian aspek pada situs TripAdvisor yang meliputi lokasi, kebersihan, dan layanan.



Gambar 1.1 Aspek pada Situs Tripadvisor ([www.tripadvisor.co.id](http://www.tripadvisor.co.id))

Data penelitian yang digunakan adalah data ulasan dari Tripadvisor yang diperoleh dengan melakukan proses *crawling* data menggunakan library scrapy pada

python. Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan sebagai metode pembobotannya dan Metode *Support Vector Machine* digunakan untuk menentukan sentiment dari opini.

## **1.2 Perumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan sebelumnya maka penulis merumuskan masalah dalam penelitian ini yaitu sebagai berikut :

1. Bagaimana hasil sentimen pada ulasan di situs tripadvisor?
2. Bagaimana proporsi kelompok ulasan positif dan negatif pada situs Tripadvisor?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian tesis ini adalah melakukan analisis sentimen untuk mengetahui opini yang bersifat positif atau negatif dari ulasan yang di tulis oleh wisatawan di situs tripadvisor serta menganalisis aspek mana yang paling banyak memiliki proporsi sentimen yang positif dan negatif.

## **1.4 Manfaat Penelitian**

Manfaat bagi peneliti

1. Dapat menerapkan ilmu yang dipelajari dibangku kuliah
2. Sebagai bahan pembelajaran bagi penulis untuk mempelajari penerapan metode TF-IDF dan *Support Vector Machine* dalam klasifikasi teks.

Manfaat bagi Masyarakat :

1. Dapat menyajikan informasi terkait opini dari wisatawan terhadap tempat wisata yang ada di Jawa Timur.
2. Dapat menyajikan informasi yang dapat digunakan sebagai acuan bagi pemerintah daerah Jawa Timur serta pengelola tempat wisata tersebut untuk melakukan evaluasi dan pengembangan.

## **1.5 Batasan Masalah**

Untuk memfokuskan permasalahan penelitian ini, batasan masalah yang ditentukan adalah sebagai berikut :

- a) Data yang digunakan hanya data ulasan yang terdapat di situs Tripadvisor pada data hotel.

- b) Data ulasan yang digunakan hanya data hotel yang terdapat di wilayah Jawa Timur.
- c) Data ulasan yang digunakan adalah ulasan hotel dalam 3 tahun terakhir (2017-2019).
- d) Metode yang digunakan untuk melakukan klasifikasi adalah *Support Vector Machine*.
- e) Klasifikasi sentimen pada ulasan hanya dikategorikan menjadi 2 yaitu ulasan positif (ulasan yang baik) dan ulasan negatif (ulasan yang buruk).
- f) Pengkategorian aspek ditinjau dari tiga aspek hotel pada situs tripadvisor meliputi lokasi, kebersihan dan layanan.

## **1.6 Sistematika Penulisan**

Berikut ini adalah sistematika penulisan yang diterapkan pada proses penelitian ini :

### **Bab I Pendahuluan**

Bab ini menjelaskan beberapa hal dasar dalam penelitian yang dilakukan meliputi : latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

### **Bab II Kajian Pustaka**

Bab ini menjelaskan tentang studi literatur dan teori-teori pendukung yang terkait dengan penelitian. Teori yang dijelaskan meliputi Opini, Kepuasan Pelanggan, Tripadvisor, Analisis Sentimen, *Text Mining*, *Text Pre-Processing*, Pembobotan TF-IDF, dan *Support Vector Machine*.

### **Bab III Metode Penelitian**

Bab ini menguraikan deskripsi tentang bagaimana penelitian dilakukan dan menjelaskan tahapan-tahapan yang akan dilakukan meliputi proses tahap awal, tahap pengembangan hingga tahap akhir dari penelitian.

### **Bab IV Hasil Penelitian dan Pembahasan**

Bab ini akan menampilkan hasil dari penelitian dan dilakukan pembahasan atas penelitian yang telah dilakukan.

## **Bab V Kesimpulan dan Saran**

Bab ini menampilkan kesimpulan dari penelitian serta akan dituliskan saran bagi penelitian selanjutnya.

BAB II  
KAJIAN  
PUSTAKA

## BAB 2 KAJIAN PUSTAKA

Bab ini menjelaskan tentang studi literatur dan teori-teori pendukung yang terkait dengan penelitian. Teori yang dijelaskan meliputi Opini, Kepuasan Pelanggan, Tripadvisor, Analisis Sentimen, *Text Mining*, *Text Pre-Processing*, Pembobotan TF-IDF, dan *Support Vector Machine*.

### 2.1 Studi Literatur

Tabel 2.1 Kajian Penelitian Terdahulu

No.	Nama Penulis	Judul	Ringkasan
1.	<ul style="list-style-type: none"> <li>• Dewi Ayu Khusnul Khotimah</li> <li>• Riyanarto Sarno</li> </ul>	Sentimen Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis <ul style="list-style-type: none"> <li>• 6th International Conference on Information and Communication Technology (ICoICT))</li> </ul>	Pada jurnal ini dilakukan analisis sentimen pada situs booking.com dengan data ulasan pelanggan hotel di kota Ponorogo menggunakan metode Probabilistic Latent Semantic Analysis (PLSA). Klasifikasi menggunakan teks berbahasa inggris untuk menentukan kepuasan dan ketidakpuasan pelanggan berdasarkan testimoni yang ditinggalkan pelanggan. proses pengambilan data dengan cara crawling data menggunakan webHarvy. Hasil penelitian ini metode

No.	Nama Penulis	Judul	Ringkasan
			PLSA lebih unggul dibandingkan penelitian sebelumnya yang menggunakan metode LSA dengan hasil akurasi data sebesar 76%.
3.	<ul style="list-style-type: none"> <li>• Rimba Nuzulul Chory</li> <li>• Muhammad Nasrun</li> <li>• Casi Setianingsih</li> </ul>	<p>Sentiment Analysis on User Satisfaction Level of Mobile Data Services Using Support Vector Machine (SVM) Algorithm</p> <ul style="list-style-type: none"> <li>• The 2018 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)</li> </ul>	<p>Pada jurnal ini dilakukan analisis sentimen tentang kepuasan publik dalam menggunakan layanan data operator telekomunikasi di Indonesia, data diambil dari twitter pada akun resmi masing-masing operator seluler atau menggunakan kata kunci terkait dengan operator seluler. Metode yang digunakan adalah Support Vector Machine dengan pembobotan TF-IDF. Hasil dalam penelitian ini adalah rata-rata tingkat akurasi sebesar 99,01%, precision sebesar 92,45%, recall sebesar</p>

No.	Nama Penulis	Judul	Ringkasan
			93,90%, dan f1-score sebesar 95,43%
	<ul style="list-style-type: none"> <li>• Shamsul Arafin Mahtab</li> <li>• Nazmul Islam</li> <li>• Md Mahfuzur Rahaman</li> </ul>	<p>Sentiment Analysis on Bangladesh Cricket with Support Vector Machine</p> <ul style="list-style-type: none"> <li>• International Conference on Bangla Speech and Language Processing(ICBSLP), 21-22 September, 2018</li> </ul>	<p>Pada jurnal ini dilakukan analisis sentimen terhadap opini masyarakat bangladesh terhadap olahraga kriket di sosial media. Data yang digunakan adalah data komentar di twitter dalam bahasa bengali. Metode pembobotan yang digunakan adalah TF-IDF dan klasifikasi dilakukan dengan menggunakan metode Support Vector Machine. Hasil penelitian ini adalah nilai akurasi yang didapatkan sebesar 64,596%, precision sebesar 66%, recall sebesar 65%, dan f1-score sebesar 63%</p>
4.	<ul style="list-style-type: none"> <li>• Indriati</li> <li>• Achmad Ridok</li> </ul>	<p>Sentiment Analysis For Review Mobile Application Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN)</p>	<p>Pada jurnal ini dilakukan analisis sentimen terhadap ulasan pada aplikasi selular. Data yang digunakan diambil</p>

No.	Nama Penulis	Judul	Ringkasan
		<ul style="list-style-type: none"> <li>Journal of Environmental Engineering &amp; Sustainable Technology Vol. 03 No.01, July 2016)</li> </ul>	<p>dari kolom komentar aplikasi lazada yang ada di google play. Metode yang digunakan adalah Neighbor Method Weighted K-Nearest Neighbor (NWKNN). Hasil dari penelitian ini adalah dengan membandingkan metode NWKNN dan KNN maka metode NWKNN lebih baik 0,27 dalam melakukan klasifikasi.</p>

## 2.2 Opini

Opini dapat diartikan suatu kesimpulan yang belum dikeluarkan dan ada di dalam pemikiran dimana hal ini dapat dijadikan sebuah bahan untuk dapat diperdebatkan. Sentimen adalah suatu opini yang sudah bersifat tetap dan dianggap keyakinan yang dapat dipegang secara teguh, namun jika dipengaruhi oleh kecenderungan maka dapat dikatakan sebagai pandangan (Moore, 1988). Opini juga dapat dikatakan sebagai jawaban terhadap suatu persoalan, isu maupun jawaban yang telah diajukan baik secara lisan ataupun tertulis. Opini biasanya dianggap sebagai sebuah jawaban lisan untuk memberikan respon kepada individu atau sebuah tanggapan kepada situasi dimana adanya sebuah pertanyaan yang diajukan (Soenarjo, 1997).

Jadi berdasarkan apa yang dijelaskan maka opini dapat dikatakan sebuah pernyataan dari seseorang atau sebuah pendapat yang diberikan seseorang baik itu disampaikan melalui lisan maupun dengan cara tertulis melalui tulisan, kata-kata, atau cara apapun yang dapat mengandung arti.

### **2.2.1 Komponen Opini**

Ada 3 komponen pada dasarnya di dalam opini (Rivers, 1994), yaitu :

a. Kepercayaan

Kepercayaan adalah sesuatu yang dapat diterima oleh khalayak, benar atau tidaknya sesuatu tersebut yang didasarkan pengalaman di masa lalu, serta pengetahuan dan informasi di masa sekarang serta persepsi yang berkesinambungan.

b. Nilai

Nilai adalah sesuatu yang melibatkan cinta dan kebencian, hasrat dan ketakutan, perasaan suka atau ketidaksukaan, bagaimana penilaian seseorang akan suatu hal, serta intensitas untuk menilai sesuatu netral, lemah atau kuat.

c. Pengharapan

Pengharapan adalah sesuatu yang mengandung citra dari seseorang tentang bagaimana keadaan yang akan ditimbulkan setelah terjadinya tindakan. Selain itu juga dapat dikatakan pertimbangan akan suatu hal yang dapat terjadi di masa lampau, suatu hal di keadaan sekarang ini serta suatu hal yang sekiranya dapat terjadi jika dilakukannya suatu tindakan.

### **2.3 Kepuasan Pelanggan**

Kepuasan pelanggan adalah keadaan dimana apa yang diharapkan konsumen terhadap suatu pelayanan sudah sesuai dengan apa yang mereka harapkan. Jadi dapat dikatakan jika suatu pelayanan yang diharapkan sudah sesuai dengan kenyataan yang pelanggan terima maka otomatis hal tersebut akan membuat pelanggan senang dikarenakan karena sudah sesuai harapan mereka. Sebaliknya, jika suatu layanan tersebut tidak sesuai dengan harapan pelanggan maka akan membuat pelanggan kecewa. Untuk mengetahui bagaimana harapan konsumen terhadap suatu pelayanan maka dapat diketahui dari pengalaman mereka sendiri saat menggunakan pelayanan suatu perusahaan jasa maupun dari omongan orang lain.

Menurut Kotler dan Keller (2009), Kepuasan adalah suatu perasaan senang atau kecewa dari seseorang yang berasal dari suatu perbandingan terhadap kinerja atau hasil suatu produk dan harapannya. Suatu perusahaan dapat mendapatkan pelanggan dengan jumlah banyak jika dapat dinilai memberikan kepuasan bagi pelanggannya. Jika

pelanggan puas terhadap suatu layanan yang telah diberikan maka dapat tercipta suatu hubungan yang harmonis antara perusahaan dan pelanggan serta dapat menciptakan dasar yang baik bagi pelanggan untuk dapat menggunakan jasa atau produk tersebut lagi. Selain itu jika pelanggan puas maka mereka dapat merekomendasikan jasa atau produk tersebut dari mulut ke mulut sehingga hal ini tentunya dapat menguntungkan suatu perusahaan.

## 2.4 TripAdvisor

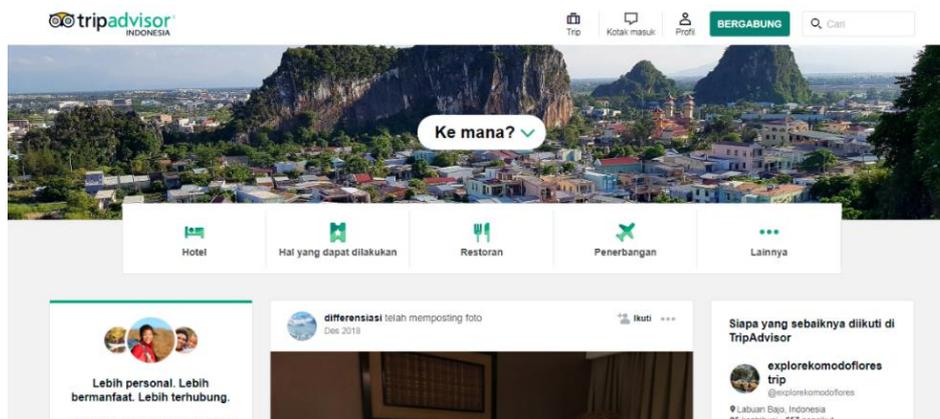
TripAdvisor adalah salah satu situs perjalanan wisata terbesar di dunia yang berkantor pusat di Needham, Massachusetts. TripAdvisor didirikan oleh Stephen Kaufer dan Langlely Steinert. Steve pada bulan Februari 2000 dan menjabat sebagai Chief Executive Officer serta Presiden Direktur sejak didirikan. Stephen Kaufer dan Langlely Steinert telah membangun Media Group TripAdvisor yang terdiri dari 19 brand penyedia layanan perjalanan terkenal, TripAdvisor juga telah memperluas merek TripAdvisor ke berbagai negara. TripAdvisor sudah berkembang di kurang lebih 30 negara, termasuk Inggris, Perancis, Jerman, Jepang, Yunani, Italia, Rusia, Spanyol, Argentina, Indonesia, Taiwan, Malaysia, India, dan Mesir. Situs ini menyediakan informasi terkait tempat wisata, hotel, restoran, dan penerbangan serta dapat membantu wisatawan dalam memesan perjalanan wisata mereka. Situs ini menawarkan fitur komentar dimana wisatawan dapat memberikan ulasan terhadap tempat wisata, hotel, dan restoran yang pernah dikunjungi sehingga dapat berbagi pengalaman dengan berbagai orang di seluruh dunia dengan jumlah reviewers aktif dan tidak aktif sebanyak 315 juta orang<sup>3</sup>. Pada situs ini juga terdapat fitur untuk dapat membandingkan harga penerbangan, menyediakan link bagi wisatawan untuk dapat memesan paket perjalanan wisata, serta untuk mencari harga hotel.



Gambar 2.1 Logo Tripadvisor  
([www.tripadvisor.co.id](http://www.tripadvisor.co.id))

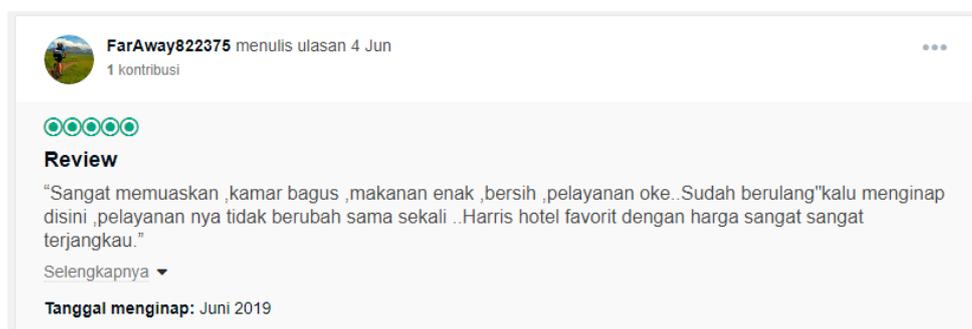
---

<sup>3</sup> <https://en.wikipedia.org/wiki/TripAdvisor>



Gambar 2.2 Situs Tripadvisor  
(www.tripadvisor.co.id)

Pada situs TripAdvisor, wisatawan dapat memberikan ulasan dalam berbagai bahasa. Pada Gambar 2.3, dapat dilihat contoh ulasan berbahasa Indonesia yang terdapat pada ulasan hotel di situs TripAdvisor.



Gambar 2.3 Ulasan pada situs TripAdvisor Analisis Sentimen  
(www.tripadvisor.co.id)

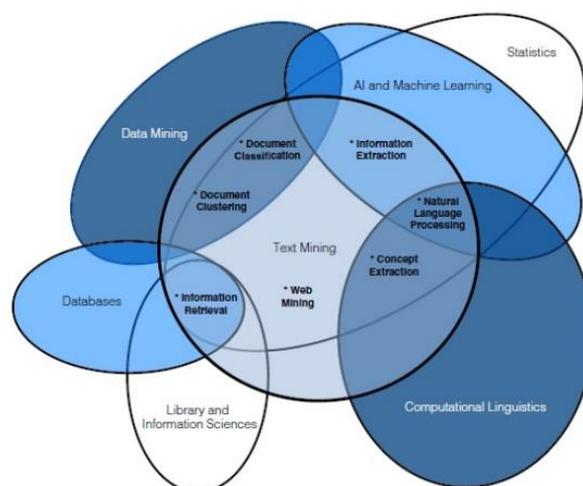
Analisis sentimen adalah sebuah cabang ilmu dari *text mining*, *natural language program*, dan *artificial intelligence* yang mempelajari bagaimana untuk menganalisis opini, sentimen, evaluasi, sikap, penilaian, dan emosi dari sebuah entitas yang dapat berupa sebuah produk, pelayanan, organisasi, individu, isu-isu, peristiwa, dan topik (Liu, 2012). Analisis sentimen disebut juga *opinion mining* yang berguna didalam pengelolaan bahasa alami, *text mining*, dan komputasi linguistik dan bertujuan untuk menentukan opini pada suatu topik tertentu, dimana perilaku tersebut dapat mengindikasikan penilaian serta alasan serta kondisi kecenderungan (Basari, 2013).

Analisis sentimen kebanyakan digunakan untuk melakukan analisis atau untuk dapat melakukan penilaian terhadap opini masyarakat baik itu opini yang merujuk

kesukaan atau ketidaksukaan terhadap suatu barang maupun jasa. Sentimen tersebut merupakan informasi yang bersifat subjektif dan memiliki nilai polaritas yang positif dan negatif dimana nilai polaritas ini dapat digunakan sebagai parameter untuk dapat menentukan sebuah keputusan (Indriati, Ridok, A, 2016).

## 2.5 Text Mining

Text mining adalah sebuah proses untuk menemukan pola yang berupa informasi atau pengetahuan pada sebuah dokumen atau sumber yang sebelumnya tidak terlihat agar dapat menjadi sebuah pola yang diinginkan untuk tujuan tertentu (Mustafa, 2009). Teknik text mining dapat diterapkan pada data dengan tujuan untuk dapat mengekstraksi pengetahuan dan text mining dapat disimpan dalam format format semi-terstruktur dan tidak terstruktur (Liao, S. H., 2012). Diagram venn dari text mining dan interaksinya dengan bidang lain, dapat dilihat pada Gambar 2.4.

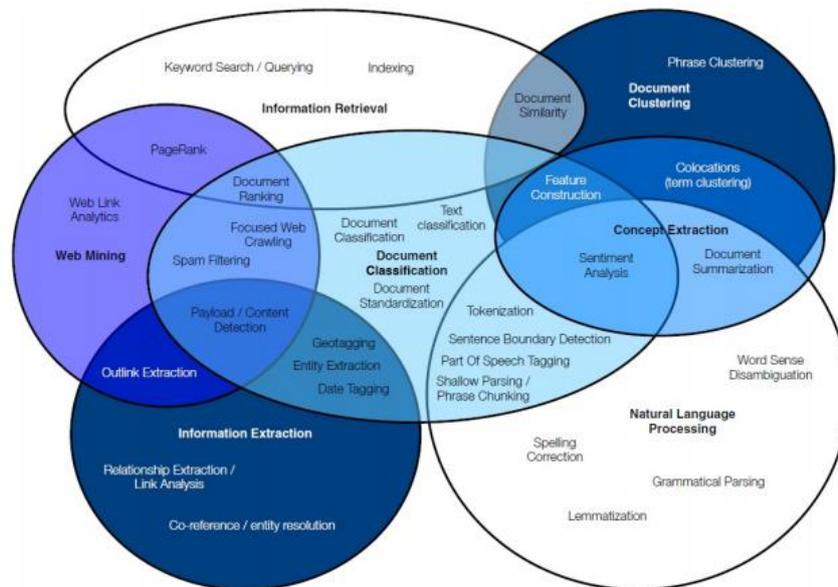


Gambar 2.4 Diagram Venn Interaksi Text Mining dengan Bidang Lain

(Liao, S. H., 2012)

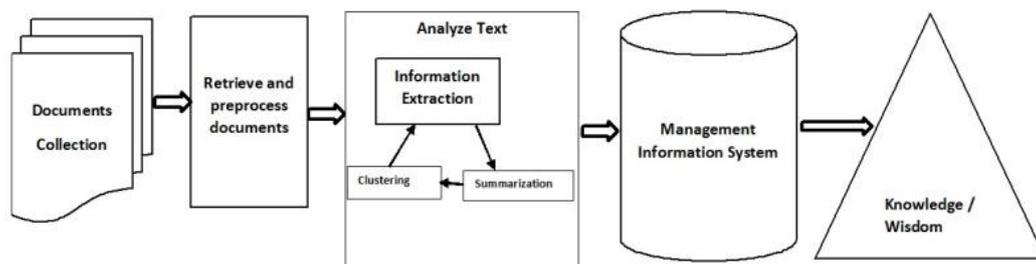
Banyak terdapat teknik data mining yang tersedia untuk dapat diterapkan untuk menganalisis data teks (Gupta, V., Lehal G. S. 2009). Pada gambar 2.5 dapat dilihat bagaimana diagram venn dari hubungan antara teknik *text mining* dan fungsionalistas intinya. Klasifikasi dokumen (*text classification, document standardization*), Pengambilan informasi (*keyword search/querying* dan *indexing*), pengelompokan dokumen (*phrase clustering*), Pemrosesan bahasa alami (*spelling correction, lemmatization, grammatical parsing* dan *word sense disambiguation*), *Information*

extraction (relationship extraction, link analysis) dan Web mining (web link analysis) (W. He. 2013). Hubungan antara berbagai teknik dalam text mining yang dapat dilihat pada Gambar 2.5.



Gambar 2.5 Hubungan Antara Berbagai Teknik Dalam Text Mining  
(Liao, S. H., 2012)

*Text mining* sering kali digunakan untuk membantu dalam menganalisis informasi, untuk membantu proses pengambilan keputusan, dan untuk manajemen informasi yang berbentuk teks dalam jumlah yang besar. Data tersebut nantinya akan diolah dengan berbagai metode seperti klasifikasi, *clustering*, analisis sentimen, dll. *Text mining* ada di dalam data mining namun mempunyai tahapan proses yang berbeda dan tahapan yang lebih banyak dari pada *data mining*. Hal ini dikarenakan pada *text mining* mengolah data berbentuk teks yang karakteristiknya lebih kompleks dari pada data berbentuk biasa atau yang sudah terstruktur. Oleh karena itu maka pada *text mining* diperlukan beberapa tahap awal untuk mempersiapkan agar teks dapat diubah hingga menjadi terstruktur (Indriati, Ridok, A, 2016). Tahapan dari proses *text mining* yang dapat dilihat pada Gambar 2.6.



Gambar 2.6 Tahapan Pada Text Mining

(Talib, R., 2016)

Pada *text mining* terdapat tahap awal sebelum melakukan pengolahan data yaitu *pre-processing*. Proses *Pre-processing* bertujuan untuk dapat mengolah data yang awalnya masih berupa teks diubah terlebih dahulu agar sesuai dengan format yang dibutuhkan. Pada *Text preprocessing* terdapat beberapa tahapan yang dapat dilakukan, yaitu *Case folding*, *tokenizing*, *filtering*, dan *stemming*.

### 2.5.1 Case Folding

Di dalam sebuah teks atau dokumen, biasanya terjadinya ketidak konsisten di dalam penggunaan huruf kapital, selain itu di dalam sebuah teks atau dokumen juga masih terdapat karakter selain huruf seperti angka dan tanda baca sehingga dibutuhkan proses *casefolding*. Pada proses ini akan dilakukan perubahan pada dokumen atau teks yang tidak konsisten tadi menjadi huruf kecil selain itu juga dilakukan proses untuk menghilangkan karakter selain huruf, seperti angka dan tanda baca. Nantinya hasil keluaran dari proses ini adalah teks atau dokumen yang kata-kata yang hurufnya sudah kecil semua dan tidak memiliki karakter selain huruf (Vijayarani, 2015).

### 2.5.2 Tokenizing

*Tokenizing* adalah salah satu proses dalam *pre-processing* untuk melakukan pemecahan kata dari kumpulan teks yang banyak dan setelah itu teks tersebut akan dipisahkan menjadi beberapa bagian. Dalam tokenisasi sebuah kalimat tersebut kemudian akan dipisahkan menjadi kata-kata (Vijayarani, 2015).

### 2.5.3 Filtering

*Filtering* atau bisa juga disebut *stop-word removal* adalah sebuah proses yang mana dilakukan pengeliminasian atau *filtering* terhadap kata-kata yang dianggap tidak penting dari proses tokenisasi yang telah dilakukan sebelumnya. Dengan melakukan *filtering* maka akan mengurangi *space* pada *tabel term index* hingga 40% atau lebih. Tahap *filtering* dilakukan dengan menggunakan algoritme *stoplist* atau *wordlist*. Untuk menghilangkan kata yang tidak penting digunakan *stopword* yang nantinya setiap kata dari hasil *case folding* di periksa dengan *stopword* dan nantinya apabila kata dari hasil *casefolding* terdapat pada daftar *stopword* maka kata tersebut di hilangkan (Vijayarani, 2015).

### 2.5.4 Stemming

*Stemming* adalah proses untuk melakukan pemangkasan kata dari bentuk aslinya yang mana masih terdapat awalan, akhiran atau sisipan menjadi suatu kata dasar dan tentunya sesuai dengan morfologi bahasa Indonesia yang benar (Indriati, Ridok A., 2016). Proses *stemming* dilakukan dengan cara merubah kata dari hasil proses *filtering* sebelumnya untuk menjadi kata dasar dengan cara menghilangkan semua imbuhan. Imbuhan yang harus di hilangkan adalah awalan, akhiran, sisipan, sehingga akan dihasilkan kata yang sudah tidak terdapat imbuhan (kata dasar).

## 2.6 Pembobotan TF-IDF

*Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode yang digunakan untuk melakukan perhitungan bobot dari setiap kata yang sudah diekstrak. TF-IDF digunakan untuk menghitung kata umum yang ada pada *information retrieval*. Pada model pembobotan TF-IDF metode ini mengintegrasikan model *term frequency* (TF) dan *inverse document frequency* (IDF), yang mana *term frequency* (TF) adalah proses untuk menghitung banyaknya jumlah kemunculan dari suatu kata pada suatu dokumen/teks dan *inverse document frequency* (IDF) untuk menghitung term pada berbagai dokumen/teks yang dianggap sebagai term umum dan dinilai tidak penting (Akbari, et al., 2012).

Pada pembobotan TF-IDF, pertama dihitung terlebih dahulu *term frequency*  $tf_{t,d}$ . Dimana  $t$  adalah term dalam dokumen  $d$  yang berfungsi untuk menunjukkan

kemunculan term  $t$  pada dokumen  $d$ . Hal ini akan berpengaruh pada bobot term yang akan semakin tinggi ketika banyak term yang muncul dalam satu dokumen (Luqyana, et al., 2018). Nilai dari  $tf$  akan dihitung bobotnya dengan menggunakan *weighting term frequency* ( $W_{tf}$ ), dengan rumus pada persamaan 2.1.

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{if } tf_{t,d} = 0 \end{cases} \quad (2.1)$$

Banyak kata yang muncul di dokumen pada umumnya merupakan nilai term frequency dari kata yang tidak penting. Sehingga untuk menghindari pembobotan pada kata yang tidak penting digunakan pembobotan *document frequency* dengan maksud untuk menghitung jumlah dokumen yang mengandung nilai term.

Di dalam suatu dokumen, munculnya suatu term yang ada di sebagian besar dokumen dapat mengakibatkan proses pencarian term unik terganggu. *Inverse Document Frequency (IDF)* berguna untuk mengurangi bobot dari suatu term jika kemunculan dari term tersebut tersebar di seluruh dokumen yang ada. Rumus dari *inverse document frequency* ditunjukkan pada persamaan 2.2.

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2.2)$$

Selanjutnya pembobotan TF-IDF dilakukan dengan mengkalikan hasil dari *document frequency* dengan *inverse document frequency*. Hal ini ditunjukkan pada persamaan 2.3.

$$W_{t,d} = W_{tf_{t,d}} \times idf_t \quad (2.3)$$

Keterangan :

$W_{tf_{t,d}}$  = bobot kata dalam setiap dokumen

$tf_{t,d}$  = jumlah kemunculan term pada dokumen

$N$  = jumlah dokumen pada kumpulan dokumen

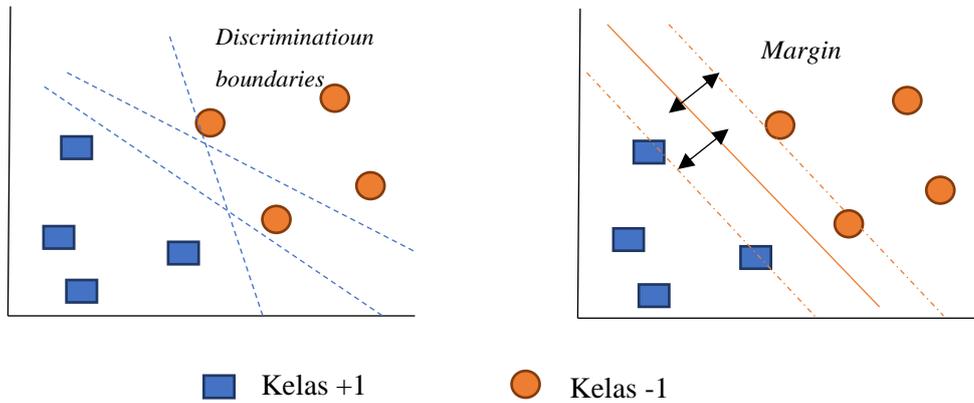
$Df$  = jumlah dokumen yang mengandung term

$idf_t$  = bobot inverse dari nilai  $df$

$W_{t,d}$  = pembobotan TF-IDF

## 2.7 Support Vector Machine

*Support Vector Machine* merupakan salah satu metode untuk melakukan klasifikasi dimana metode pembelajaran yang digunakan pada metode ini adalah terbimbing. Algoritme ini diciptakan oleh Vladimir Vapnik, berfungsi dalam menganalisis data dan mengenali pola agar mendapatkan hasil klasifikasi. Algoritma *support vector machine* mempunyai konsep yang sederhana dimana akan dicari *hyperplane* atau garis pembatas terbaik yang berfungsi untuk memisahkan dua buah kelas, apakah data tersebut masuk ke dalam suatu kategori atau kategori yang lain. (Lidya, et. al.,2015). Contoh pemisahan kelas menggunakan *hyperplane* pada metode *support vector machine* ditunjukkan pada Gambar 2.7.



Gambar 2.7 *Hyperlane Support Vector Machine*

(Shandra, E. N, 2019)

Pada Gambar 2.7, telah ditunjukkan terdapat dua buah kelas yang berbeda yaitu +1, yang menunjukkan kelas positif dan -1, menunjukkan kelas negatif. Agar mendapatkan garis *hyperplane* yang terbaik untuk memisahkan data ke dua buah kelas tersebut maka digunakan perhitungan *margin hyperplane* dan mencari titik maksimal. Dalam membagi dua buah kelas dengan *hyperplane* terbaik tersebut, dapat didefinisikan pada Persamaan 2.4.

$$(w \cdot x_i) + b = 0 \tag{2.4}$$

Di dalam pattern  $x_i$ , yang termasuk pada *class* -1 maka dapat dirumuskan seperti pada Pertidaksamaan 2.5.

$$(w \cdot x_i + b) \leq 1, y_i = -1 \quad (2.5)$$

Sedangkan pattern  $x_i$  yang termasuk pada *class* +1 maka dapat dirumuskan seperti pada Pertidaksamaan 2.6.

$$w \cdot x_i + b \geq 1, y_i = 1 \quad (2.6)$$

Untuk dapat menemukan nilai margin terbesar dilakukan dengan cara memaksimalkan nilai jarak antara hyperlane dan titik terdekatnya. Hal ini di dapatkan dengan rumus :

$$1/\|w\| \quad (2.7)$$

Keterangan :

$x_i$  : data ke- $i$ .

$y_i$  : kelas data ke- $i$ .

$w$  : vektor tegak lurus terhadap *hyperplane*.

$b$  : nilai bias.

### 2.7.1 Support Vector Machine Non-Linear

Di dalam proses untuk mengklasifikasikan data biasanya terdapat sebuah masalah yaitu dimana di dalam proses klasifikasi adalah kebanyakan sampel data tidak terpisah secara linier sehingga jika digunakan *support vector machine* linier maka hasil yang diperoleh tidak akan optimal dan mengakibatkan hasil klasifikasi yang tidak baik (buruk). *Support vector machine* linier dapat diubah menjadi *support vector machine* non-linier dengan menggunakan menambahkan fungsi *kernel*. Dengan menggunakan *kernel*, Metode ini bekerja dengan cara memetakan data input ke ruang *feature* yang dimensinya lebih tinggi. Sehingga diharapkan data input hasil pemetaan ke ruang *feature* akan terpisah secara linier sehingga dapat dicari *hyperplane* yang optimal (Shandra, 2019). Berikut adalah beberapa fungsi *kernel* yang sering digunakan.

Kernel Linier,

$$K(x_i, x_d) = X_i^T \cdot X_j \quad (2.8)$$

Kernel Polinomial,

$$K(x_i, x_d) = (X_i^T X_j + 1)^d, y > 0 \quad (2.9)$$

Kernel Radial Basis Function (RBF),

$$K(x_i, x_d) = \exp(-\gamma \|X_i - X_j\|^2), y > 0 \quad (2.10)$$

Kernel Sigmoid,

$$K(x_i, x_d) = \tan y X_i^T X_j + r \quad (2.11)$$

### 2.7.2 Sequential Training SVM

Dalam memproses data latih *support vector machine* dapat digunakan tiga algoritme yaitu *Quadratic Programming*, *Sequential Minimal Optimization*, dan *Sequential Training*. Di dalam penggunaannya harus diperhatikan kekurangan dan kelebihan dari masing-masing algoritme. *Quadratic Programming* merupakan proses perumusan yang dapat memberikan hasil berupa analisa numerik dengan algoritme yang kompleks dan memakan waktu cukup lama. *Sequential Minimal Optimization* adalah pengembangan dari *Quadratic Programming* dimana algoritme ini hanya mampu memberikan optimasi yang kecil. Sedangkan *Sequential Training* adalah algoritme sederhana yang tidak memakan banyak waktu. Langkah-langkah *sequential learning* adalah sebagai berikut (Vijayakumar, 1999) :

- a. Melakukan inisialisasi terhadap  $\alpha_i=0$  dan parameter lainnya, seperti nilai  $\lambda, \gamma, C$  dan  $\epsilon$ .

Keterangan :

$\alpha_i$  = alfa ke-i, digunakan untuk mencari support vector.

$\gamma$  = konstanta *gamma* untuk mengontrol kecepatan.

$C$  = variabel slack

$\epsilon$  = epsilon digunakan untuk mencari nilai error

- b. Menghitung matriks Hessian yang didapat dari perkalian antar kernel polynomial dan  $y$  dimana merupakan vector bernilai 1 dan -1. Persamaannya adalah sebagai berikut :

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (2.12)$$

Dengan  $i, j = 1, 2, \dots, n$ .

Keterangan :

$x_i$  = data ke- $i$

$x_j$  = data ke- $j$

$y_i$  = kelas data ke- $i$

$y_j$  = kelas data ke- $j$

$n$  = jumlah data

$\lambda$  = nilai *Lambda*

$K(x_i, x_j)$  = fungsi *Kernel* yang digunakan

- c. Melakukan iterasi untuk setiap iterasi yang telah di inisialisasi di awal, lalu menghitung nilai  $E_i$  yang dapat dihitung menggunakan persamaan 2.13.

$$E_i = \sum_{j=1}^n \alpha_j D_{ij} \quad (2.13)$$

Keterangan :

$\alpha_i$  = alfa ke- $i$

$E_i$  = *Error rate*

$D_{ij}$  = Matriks *Hessian*

Menghitung nilai  $\delta\alpha_i$  dapat dihitung menggunakan Persamaan 2.14.

$$\delta\alpha_i = \min(\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i) \quad (2.14)$$

Keterangan :

$\alpha_i$  = alfa ke- $i$

$E_i$  = *Error rate*

$C$  = Konstanta  $C$

$\gamma$  = konstanta *gamma*

Memperbarui nilai  $\alpha_i$  dapat dihitung menggunakan Persamaan 2.15.

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (2.15)$$

Keterangan :

$\alpha_i$  = alfa ke- $i$

$\delta\alpha_i$  = delta alfa ke- $i$

- d. Kembali ke langkah ketiga dan lakukan secara berulang-ulang sampai mendapatkan iterasi maksimum atau  $\max(|\delta\alpha_i|) < \varepsilon$ .

- e. Dari proses di atas akan didapatkan nilai support vector (SV), dimana nilai  $SV = (\alpha > thresholdSV)$ . Setelah itu dilakukan perhitungan nilai bias  $b$  dengan menggunakan persamaan 2.16.

$$b = -\frac{1}{2} \left( \sum_{i=1}^N \alpha_i y_i K(x_i, x^-) \right) + \sum_{i=1}^N \alpha_i y_i K(x_i, x^+) \quad (2.16)$$

- f. Melakukan perhitungan fungsi  $f(x)$  agar dapat mengetahui hasil klasifikasi pada kelas sentimen tertentu dengan persamaan 2.17

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (2.17)$$

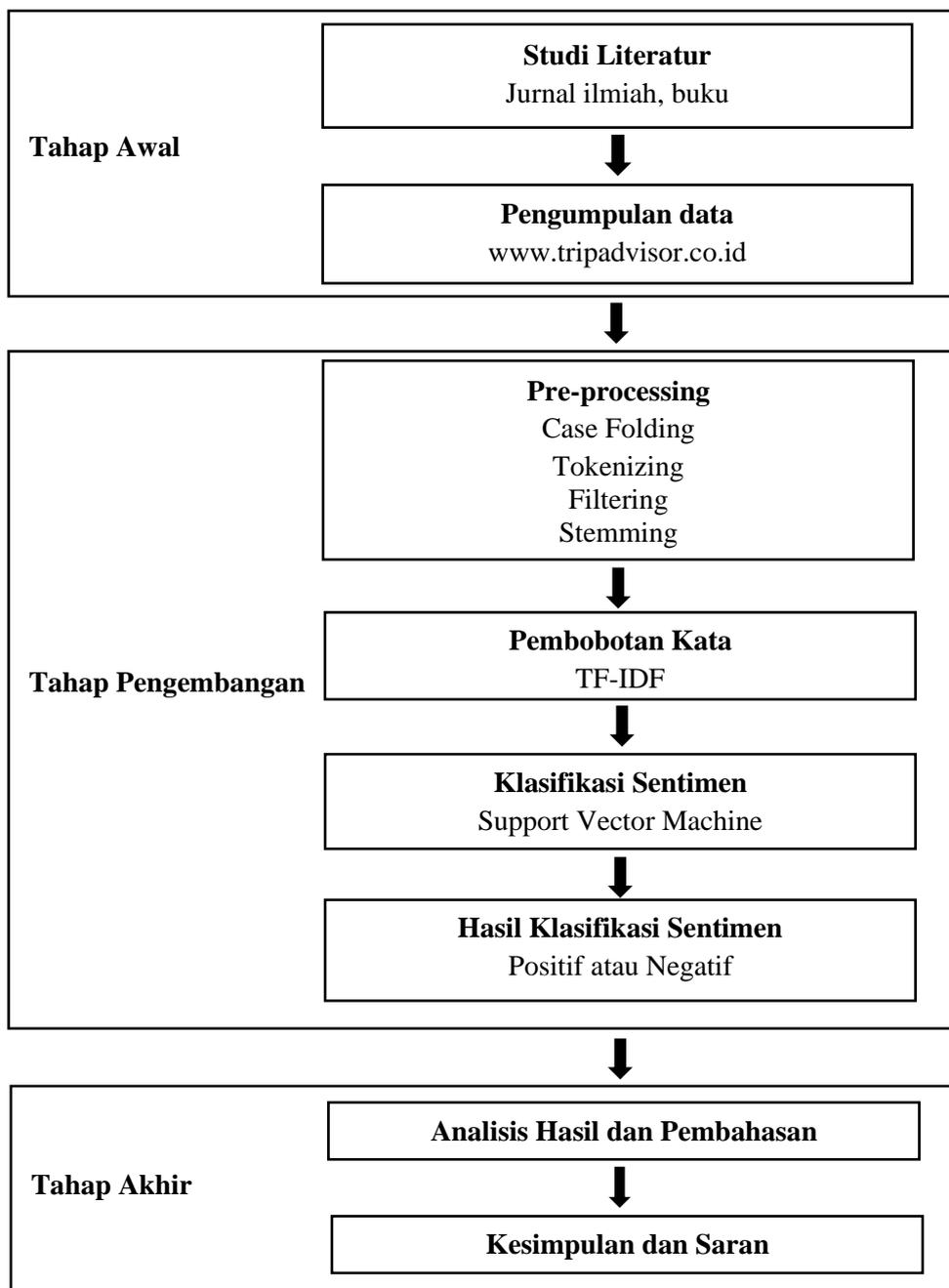
Jika fungsi bernilai positif maka dokumen akan diklasifikasikan pada kelas sentimen positif, sebaliknya jika fungsi bernilai negatif maka akan diklasifikasikan pada kelas sentimen negatif.

BAB III  
METODOLOGI  
PENELITIAN

### BAB 3

## METODOLOGI PENELITIAN

Bab ini menguraikan deskripsi tentang bagaimana penelitian dilakukan dan menjelaskan tahapan-tahapan yang akan dilakukan meliputi proses tahap awal, tahap pengembangan hingga tahap akhir dari penelitian. Tahap yang dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

### 3.1 Tahap Awal

Tahap ini adalah tahapan awal dari penelitian yang dilakukan. Pada tahap ini terdapat 2 tahapan yang akan dilakukan yaitu melakukan studi literatur terhadap topik-topik yang berkaitan dengan penelitian yang dilakukan serta pengumpulan data yang dibutuhkan untuk penelitian.

#### 3.1.1 Studi Literatur

Studi literatur dilakukan dengan cara mencari sumber literatur yang berhubungan dengan analisis sentimen agar dapat memperoleh pengetahuan dan dapat menunjang penelitian yang dilakukan. Literatur yang digunakan dapat bersumber dari buku, jurnal, thesis terdahulu, maupun sumber pustaka internet. Hasil dari tahapan ini sudah di jelaskan pada Bab 2 seperti Studi literatur dari penelitian terdahulu serta teori-teori dari :

- a. Opini
- b. Kepuasan Pelanggan
- c. Tripadvisor
- d. Analisis Sentimen
- e. *Text Mining*
- f. *Text Pre-Processing*
- g. Pembobotan TF-IDF
- h. *Support Vector Machine*

#### 3.1.2 Pengumpulan data

Pada tahap ini dilakukan pengumpulan data untuk digunakan dalam melakukan penelitian. Data yang digunakan adalah data yang terkait dengan ulasan hotel pada situs tripadvisor dengan batasan hanya data hotel yang ada di Jawa Timur dalam 3 tahun terakhir (2017-2019). Proses pengambilan data dengan menggunakan library scrapy pada python dan ulasan akan disimpan dengan bentuk file CSV.

Data yang sudah didapatkan selanjutnya dipisahkan lagi berdasarkan tanda titik (.). Jadi jika dalam suatu ulasan terdapat dua kalimat, maka data pada satu dokumen tersebut akan di *split* dan dihitung sebagai dua ulasan. Setelah itu data yang sudah didapatkan akan dilakukan pelabelan pada aspeknya secara manual oleh anatator untuk

mengkategorikan ulasan ke aspek lokasi, kebersihan, atau layanan. Setelah dilakukan pelabelan selanjutnya data akan dibagi menjadi data training dan data testing.

Kriteria dari pelabelan aspek secara manual yang digunakan dapat dilihat pada Tabel 3.1.

Tabel 3.1 Kriteria Aspek Pelayanan yang Digunakan pada Penelitian

No.	Nama Aspek	Rincian
1.	Lokasi	Aspek lokasi berisi ulasan yang membahas tentang lokasi, kemudahan akses transportasi dan jalan, serta kedekatan tempat dengan beberapa tempat umum seperti mall, bandara, dll.
2.	Kebersihan	Aspek kebersihan berisi ulasan yang membahas bagaimana kebersihan baik dari kebersihan tempat secara umum ataupun bagian di dalamnya seperti ruangan kamar, kebersihan fasilitas maupun kamar mandi.
3.	Layanan	Aspek layanan berisi ulasan yang membahas bagaimana pelayanan yang diberikan baik berupa layanan dari pegawai hotel (jasa), harga yang diberikan, penyediaan makanan beserta fasilitas yang ada seperti wifi, kolam renang, gym, tempat meeting, dsb yang dapat dikatakan sebagai bentuk pelayanan lain dari mereka.

## 3.2 Tahap Pengembangan

Tahap pengembangan dilakukan setelah melakukan tahapan awal. Terdapat tiga tahapan yang dilakukan yaitu melakukan pre-processing data, melakukan pembobotan kata, dan melakukan klasifikasi sentimen.

### 3.2.1 Pre-processing Data

Setelah mendapatkan data ulasan dari orang-orang di situs tripadvisor dari hasil crawling maka proses pengolahan data dilanjutkan ke tahap pre-processing. Pada tahap ini terdapat beberapa tahapan yang dilakukan yaitu, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

#### a. *Case folding*

Pada tahapan ini data ulasan yang sudah didapatkan akan diubah teks yang hurufnya tidak konsisten tadi menjadi huruf kecil semua dan akan dilakukan penghilangan karakter selain huruf, seperti angka dan tanda baca sehingga hasil akhirnya dokumen/teks tersebut sudah kecil semua hurufnya dan tanda bacanya sudah hilang.

#### b. *Tokenizing*

Pada tahap ini ulasan yang sudah hurufnya sudah kecil semua dan sudah hilang tanda bacanya tersebut akan di tokenizing untuk dipecah katanya menjadi beberapa bagian sehingga akan terpisah menjadi kata-kata.

#### c. *Filtering*

Pada tahap *Filtering* ini hasil dari tahap sebelumnya akan dieliminasi kata-kata yang dianggap tidak penting seperti kata penghubung (di, ke, dan, dll).

#### d. *Stemming*

Setelah di filter maka kata-kata tersebut akan di pangkas jika masih terdapat awalan, akhiran atau sisipan agar menjadi suatu kata dasar. Dari kata dasar ini lah nantinya yang akan dilakukan pembobotan katanya sebelum masuk ke tahap klasifikasi.

### 3.2.2 Pembobotan Kata

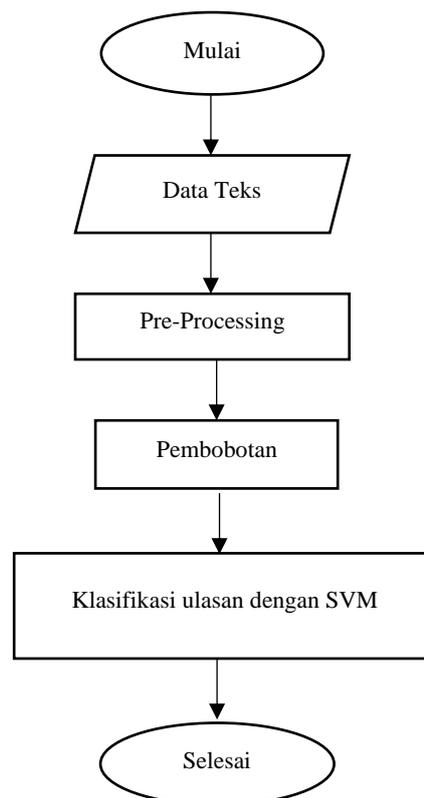
Tahap selanjutnya adalah melakukan pembobotan kata. Pada tahapan ini dilakukan pembobotan terhadap kata dasar yang sudah di dapatkan dari hasil proses pre-processing. Pembobotan kata dilakukan dengan menggunakan metode *Term Frequency-*

*Inverse Document Frequency* (TF-IDF). Pembobotan dilakukan dengan memanfaatkan module yang ada pada python, yaitu dengan menggunakan module scikit-learn untuk melakukan pembobotan TF-IDF.

### 3.2.3 Klasifikasi Sentimen

Tahap selanjutnya adalah melakukan klasifikasi sentimen. Setiap ulasan diklasifikasi menurut jenis sentimennya untuk diketahui opininya yaitu bersifat positif atau negatif. Proses klasifikasi sentimen pada penelitian ini dilakukan dengan menggunakan metode *support vector machine*.

Cara kerja dari metode *support vector machine* yaitu akan menemukan hyperlane atau garis pembatas terbaik yang akan membagi dua kelas dan akan mengklasifikasikan data berdasarkan kedekatan kata pada suatu sisi garis. Pada gambar 3.2, ditunjukkan Alur proses klasifikasi menggunakan *support vector machine*.



Gambar 3.2 Alur Proses Klasifikasi Menggunakan *Support Vector Machine*

### 3.3 Tahap Akhir

Tahapan ini adalah tahapan akhir dari penelitian yang telah dilakukan. Pada tahapan ini dilakukan analisis dan pembahasan dari hasil klasifikasi serta kesimpulan dan saran. Pada tahap ini juga dilakukan penyusunan dokumen dari penelitian yang telah dilakukan dari pendahuluan, kajian pustaka, metodologi penelitian, hasil penelitian serta kesimpulan dan saran.

#### 3.3.1 Analisis Hasil dan Pembahasan

Setelah didapatkan hasil sentimen maka akan dilakukan analisis terhadap hasil yang sudah didapatkan. Dari hasil analisis dapat diketahui bagaimana sentimen dari ulasan pelanggan hotel di Jawa timur pada situs TripAdvisor, tempat mana yang memiliki banyak ulasan dengan opini yang positif dan tempat mana yang memiliki banyak ulasan dengan opini negatif. Berdasarkan hasil klasifikasi sentimen berdasarkan pengkategorian aspek juga dapat diketahui aspek mana yang paling banyak mendapat sentimen negatif dan positif, juga bagaimana rating yang diberikan TripAdvisor pada setiap aspek jika dibandingkan dengan rating yang didapatkan dari hasil penelitian.

Setelah itu dilakukan pengukuran evaluasi untuk mengukur tingkat keberhasilan suatu *classifier*. Pada umumnya untuk mengukur evaluasi menggunakan *confusion matrix*. Pengukuran evaluasi dengan berdasarkan *confusion matrix* dapat dilihat pada Tabel 3.2.

Tabel 3.2 *Confusion Matrix*  
(Han et al, 2012)

		Kelas Prediksi		Total
		TRUE	FALSE	
Kelas Sebenarnya	TRUE	$TP$ ( <i>True Positive</i> )	$FN$ ( <i>False Negative</i> )	$P$
	FALSE	$FP$ ( <i>False Positive</i> )	$TN$ ( <i>True Negative</i> )	$N$
Total		$P'$	$N'$	$P+N$

Keterangan :

*True Positive* (TP) = Data positif yang dapat dilabeli secara benar oleh *classifier*

*True Negative* (TN) = Data negatif yang dapat dilabeli secara benar oleh *classifier*

*False Positive* (FP) = Data negatif yang dilabeli salah menjadi positif oleh *classifier*

*False Negative* (FN) = Data positif yang dilabeli salah menjadi negatif oleh *classifier*

Perhitungan yang dilakukan adalah Accuracy, Precision, Recall dan F1 Score yang didefinisikan pada persamaan berikut.

Akurasi adalah keberhasilan *classifier* di dalam menentukan label data. Berikut adalah persamaan untuk menghitung akurasi.

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN} \quad (3.1)$$

*Precision* adalah pengukuran untuk dapat mengukur ketepatan (*Exactness*). Sedangkan *Recall* adalah pengukuran untuk mendapatkan *completeness*. Berikut adalah persamaan untuk menghitung *Precision* dan *Recall*.

$$Precision = \frac{TP}{TP+FP} \quad (3.2)$$

$$Recall = \frac{TP}{TP+FN} = \frac{TP}{P} \quad (3.3)$$

Nilai *Precision* dan *Recall* dapat digunakan untuk membentuk matrix pengukuran lain yaitu *F1 Score* atau *F-Measure*. *F1 Score*, adalah perhitungan yang merupakan harmonic mean dari *Precision* dan *Recall*. Berikut adalah persamaan untuk menghitung *F1 Score*.

$$F1\ Score = 2 \times \frac{Precision \times recall}{Precision + recall} \quad (3.4)$$

### 3.3.2 Kesimpulan dan Saran

Tahapan terakhir yang dilakukan adalah menarik kesimpulan dari penelitian yang telah dilakukan serta memberikan saran untuk perbaikan pada penelitian berikutnya yang akan dilakukan.

**BAB IV**  
**HASIL**  
**PENELITIAN**

## BAB 4 HASIL PENELITIAN

Bab ini akan menampilkan hasil dari penelitian dan dilakukan pembahasan atas penelitian yang telah dilakukan.

### 4.1 Data Ulasan

Data ulasan pada penelitian ini di dapatkan dari hasil crawling dengan menggunakan library scrapy pada python. Data yang diambil adalah ulasan pelanggan hotel daerah Jawa Timur pada situs TripAdvisor. Data yang didapatkan tersebut selanjutnya dibagi menjadi data training dan data testing. Untuk data training digunakan 300 ulasan dengan rincian 150 ulasan berlabel positif dan 150 ulasan berlabel negatif dari berbagai hotel di Jawa Timur. Sedangkan untuk data testing menggunakan data ulasan dari 5 hotel di kota Surabaya dan 5 hotel di kota Malang yang paling populer menurut rating pada situs TripAdvisor. Pada Tabel 4.1 dapat dilihat *coding* python yang digunakan untuk melakukan crawling.

Tabel 4.1 *Coding* Scrapy pada Python untuk Crawling Data

```
import sys
import scrapy
from scrapy import Selector
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as ec

class CrawlingHotelJatimSpider(scrapy.Spider):
    index = 0
    name = 'crawling_hotel_jatim'
    allowed_domains = ['tripadvisor.co.id']
    start_urls = [' https://www.tripadvisor.co.id/Hotel_Review-g297710-
d5080545-Reviews-The_101_Malang_OJ-Malang_East_Java_Java.html ' ]

    def __init__(self):
        self.driver = webdriver.Firefox()
    def parse(self, response):
        self.driver.get(response.url)
```

```

while True:
    try:
        button_next = self.driver.find_element_by_xpath(
            ("//a[contains(@class,'nav next')]")
        )
        self.driver.execute_script("document.querySelector('a.next')
            .scrollIntoView();")

        # if self.index == 1:
        #     break
        self.index += 1
        expandable_review = self.driver\
            .find_element_by_xpath("//span[text()='Selengkapnya' and
            contains(@class,'ExpandableReview')]")
        expandable_review.click()

        sel = Selector(text=self.driver.page_source)
        print('selector ', sel)
        for index, review in enumerate(
            sel.xpath("//div[contains(@class,'SingleReview') and
            contains(@class,'reviewContainer')]")):
            print("data test", review.xpath("//q[contains
            (@class,'hotels-review')]/span[1]/text()")
                .getall()[index].strip())
            yield {
                'author':
                    review.xpath("//div[contains(@class,'SingleReview'
                    ) and contains(@class,'reviewContainer')]")
                    .xpath("//a[contains(@class,'member')]/text()").getall()[
                    index].strip(),
                'event_date':
                    review.xpath("//div[contains(@class,'EventDate')]/
                    span/text()").getall()[index].strip(),
                'review':
                    ' '.join(review.xpath("//q[contains(@class,'hotels-
                    review')]/span[1]")
                        [index]
                        .xpath("text()").getall()).strip(),
            }
        button_next.click()

```

```

WebDriverWait(self.driver, 10).until
(ec.presence_of_element_located((
By.XPATH, "///div[contains(@class, 'SingleReview')
and contains (@class, 'reviewContainer')]"))))
except:
    print(sys.exc_info())
    break
self.driver.close()

```

Setelah dilakukan proses crawling maka akan didapatkan data ulasan mentah dari situs TripAdvisor. Contoh hasil crawling data mentah dari situs TripAdvisor dapat dilihat pada Tabel 4.2.

Tabel 4.2 Data Ulasan Mentah Dari Situs TripAdvisor

<b>Nama Pengulas</b>	<b>Ulasan</b>
Harry P	Staff hotel ramah dan sigap membantu saat pertama kali tiba di hotel, Sarapan variatif dan rasanya enak, Harganya relatif lebih murah. Tapi akan lebih bagus lagi seandainya disekitar hotel ada minimarket sehingga tidak menyulitkan jika ingin belanja.
Skhalid	Hotel bisnis yang sangat recommended disini staffnya sangat ramah membantu dari mulai masuk sudah disambut makanan di restorannya juga enak enak kamar nya nyaman
Lutfi	Pelayanannya menyenangkan, Ketika masuk disambut dengan hangat dan langsung diberikan minuman.
Erwin	Mungkin perlu ditambah semacam mini market di area resto dan kolam renang supaya fasilitas lebih seru dan tidak membuat bosan.
Nurhadi	lokasinya strategis di tengah kota.

Data mentah yang didapatkan dari hasil crawling selanjutnya diolah terlebih dahulu, dimana data ulasan yang mengandung lebih dari satu kalimat pada setiap

ulasannya di *split* dan diberikan id pada masing-masing dokumen. Proses untuk *split* data dilakukan ketika terdapat tanda baca (.) pada ulasan, hal ini dikarenakan adanya kemungkinan perbedaan sentimen pada setiap kalimat dalam satu ulasan tersebut. Setiap ulasan yang memiliki lebih dari satu kalimat di *split* dan diberikan id baru. Contoh data yang sudah dilakukan *split* pada setiap kalimatnya dapat dilihat pada Tabel 4.3.

Tabel 4.3 Data Ulasan yang Sudah di *Split* dan Diberikan ID

ID_Dokumen	ID_Ulasan	Ulasan
1	1	Staff hotel ramah dan sigap membantu saat pertama kali tiba dihotel, Sarapan variatif dan rasanya enak , Harganya relatif lebih murah.
1	2	Tapi akan lebih bagus lagi seandainya disekitar hotel ada minimarket sehingga tidak menyulitkan jika ingin belanja.
2	3	Hotel bisnis yang sangat rekomended disini staffnya sangat ramah membantu dari mulai masuk sudah disambut makanan di restoranya juga enak enak kamar nya nyaman
3	4	Pelayanannya menyenangkan, Ketika masuk disambut dengan hangat dan langsung diberikan minuman.
4	5	Mungkin perlu ditambah semacam mini market di area resto dan kolam renang supaya fasilitas lebih seru dan tidak membuat bosan.
5	6	lokasinya strategis di tengah kota.

Pada Tabel 4.3 dapat dilihat, ID\_Dokumen menyatakan id dari setiap ulasan wisatawan yang diambil pada situs TripAdvisor dan ID\_Ulasan yang menyatakan id dari setiap ulasan yang sudah di *split* per-kalimatnya dari setiap dokumen.

Selanjutnya dilakukan pelabelan aspek untuk mengkategorikan ulasan ke dalam tiga aspek yaitu lokasi, kebersihan dan layanan. Proses pelabelan aspek dilakukan secara manual oleh anatator. Contoh data ulasan yang sudah dilabeli menurut aspeknya dapat dilihat pada Tabel 4.4.

Tabel 4.4 Data Ulasan yang Sudah Dilabeli Menurut Aspeknya.

<b>ID_Ulasan</b>	<b>Ulasan</b>	<b>Aspek</b>
1	Staff hotel ramah dan sigap membantu saat pertama kali tiba dihotel, Sarapan variatif dan rasanya enak , Harganya relatif lebih murah.	Layanan
2	Tapi akan lebih bagus lagi seandainya disekitar hotel ada minimarket sehingga tidak menyulitkan jika ingin belanja.	Lokasi
3	Hotel bisnis yang sangat rekomended disini staffnya sangat ramah membantu dari mulai masuk sudah disambut makanan di restoranya juga enak enak kamar nya nyaman	Layanan
4	Pelayanannya menyenangkan, Ketika masuk disambut dengan hangat dan langsung diberikan minuman.	Layanan
5	Mungkin perlu ditambah semacam mini market di area resto dan kolam renang supaya fasilitas lebih seru dan tidak membuat bosan.	Layanan
6	lokasinya strategis di tengah kota.	Lokasi

#### 4.2 Pre-processing Data

Pre-processing terhadap data ulasan dilakukan dengan melakukan beberapa tahap yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Tahap pre-processing terhadap

ulasan dilakukan dengan menggunakan *Natural Language Toolkit* (NLTK) yang terdapat pada python.

#### 4.2.1 Case Folding

Proses proses *case folding* dilakukan dengan mengubah teks yang hurufnya tidak konsisten menjadi huruf kecil semua serta dilakukan penghilangan karakter selain huruf, seperti angka dan tanda baca. Hasil akhir dari proses *case folding* adalah ulasan yang sudah kecil semua hurufnya dan tanda bacanya sudah hilang. Hasil data ulasan yang telah melalui tahap *case folding* dapat dilihat pada Tabel 4.5.

Tabel 4.5 Data Ulasan Sesudah Melalui Tahap *Case Folding*

<b>ID_Ulasan</b>	<b>Setelah Case Folding</b>
1	staff hotel ramah dan sigap membantu saat pertama kali tiba di hotel sarapan variatif dan rasanya enak harganya relatif lebih murah
2	tapi akan lebih bagus lagi seandainya disekitar hotel ada minimarket sehingga tidak menyulitkan jika ingin belanja
3	hotel bisnis yang sangat oke disini staffnya sangat ramah membantu dari mulai masuk sudah disambut makanan di restoran juga enak enak kamar nya nyaman
4	pelayanan menyenangkan Ketika masuk disambut dengan hangat dan langsung diberikan minuman
5	mungkin perlu di tambah semacam minimarket di area restoran dan kolam renang supaya fasilitas lebih seru dan tidak membuat bosan
6	lokasinya strategis di tengah kota

#### 4.2.2 Tokenizing

Pada proses *tokenizing* ulasan yang sudah hurufnya sudah kecil semua dan sudah hilang tanda bacanya akan dipecah katanya menjadi beberapa bagian sehingga akan terpisah menjadi kata-kata. Hasil data ulasan yang telah melalui tahap *tokenizing* dapat dilihat pada Tabel 4.6.

Tabel 4.6 Data Ulasan Sesudah Melalui Tahap *Tokenizing*

ID_Ulasan	Setelah <i>Tokenizing</i>
1	['staff', 'hotel', 'ramah', 'dan', 'sigap', 'membantu', 'saat', 'pertama', 'kali', 'tiba', 'di', 'hotel', 'sarapan', 'variatif', 'dan', 'rasanya', 'enak', 'harganya', 'relatif', 'lebih', 'murah']
2	['tapi', 'akan', 'lebih', 'bagus', 'lagi', 'seandainya', 'disekitar', 'hotel', 'ada', 'minimarket', 'sehingga', 'tidak', 'menyulitkan', 'jika', 'ingin', 'belanja']
3	['hotel', 'bisnis', 'yang', 'sangat', 'oke', 'disini', 'staffnya', 'sangat', 'ramah', 'membantu', 'dari', 'mulai', 'masuk', 'sudah', 'disambut', 'makanan', 'di', 'restoran', 'juga', 'enak', 'enak', 'kamar', 'nya', 'nyaman']
4	['pelayanan', 'menyenangkan', 'Ketika', 'masuk', 'disambut', 'dengan', 'hangat', 'dan', 'langsung', 'diberikan', 'minuman']
5	['mungkin', 'perlu', 'di', 'tambah', 'semacam', 'minimarket', 'di', 'area', 'restoran', 'dan', 'kolam', 'renang', 'supaya', 'fasilitas', 'lebih', 'seru', 'dan', 'tidak', 'membuat', 'bosan']
6	['lokasinya', 'strategis', 'di', 'tengah', 'kota']

#### 4.2.3 Filtering

Pada proses *Filtering* akan dieliminasi kata-kata yang dianggap tidak penting atau tidak memiliki makna seperti kata penghubung. Hasil data ulasan yang telah melalui tahap *Filtering* dapat dilihat pada Tabel 4.7.

Tabel 4.7 Data Ulasan Sesudah Melalui Tahap *Filtering*

ID_Ulasan	Setelah Filtering
1	['hotel', 'ramah', 'sigap', 'membantu', 'pertama', 'hotel', 'sarapan', 'variatif', 'rasanya', 'enak', 'harganya', 'relatif', 'murah']
2	['lebih', 'bagus', 'hotel', 'minimarket', 'menyulitkan', 'belanja']
3	['hotel', 'hotel', 'bisnis', 'ramah', 'membantu', 'makanan', 'restoran', 'enak', 'enak', 'kamar', 'nyaman']
4	['pelayanan', 'menyenangkan', 'hangat', 'minuman']
5	['tambah', 'semacam', 'minimarket', 'area', 'restoran', 'kolam', 'renang', 'fasilitas', 'seru', 'membuat', 'bosan']
6	['lokasinya', 'strategis', 'tengah', 'kota']

#### 4.2.4 Stemming

Pada proses *stemming*, kata-kata tersebut akan di pangkas jika masih terdapat awalan, akhiran atau sisipan agar menjadi suatu kata dasar. Dari kata dasar ini lah nantinya yang akan dilakukan pembobotan katanya sebelum masuk ke tahap klasifikasi. Hasil data ulasan yang telah dilakukan *stemming* dapat dilihat pada Tabel 4.8.

Tabel 4.8 Data Ulasan Sesudah Melalui Tahap *Stemming*

ID_Ulasan	Setelah Stemming
1	['hotel', 'ramah', 'sigap', 'bantu', 'pertama', 'hotel', 'sarapan', 'variatif', 'rasa', 'enak', 'harga', 'relatif', 'murah']
2	['lebih', 'bagus', 'hotel', 'minimarket', 'sulit', 'belanja']
3	['hotel', 'bisnis', 'ramah', 'bantu', 'makanan', 'restoran', 'enak', 'enak', 'kamar', 'nyaman']
4	['layan', 'senang', 'hangat', 'minuman']
5	['tambah', 'macam', 'minimarket', 'area', 'restoran', 'kolam', 'renang', 'fasilitas', 'seru', 'buat', 'bosan']
6	['lokasi', 'strategis', 'tengah', 'kota']

### 4.3 Proses Pembobotan Kata

Setelah dilakukan pre-processing maka *term list* yang didapat dari proses tersebut akan di hitung bobot angka sesuai dengan jumlah kemunculannya di setiap dokumen. Proses pembobotan kata dilakukan dengan menggunakan TF-IDF. Proses perhitungan menggunakan TF-IDF digunakan untuk menghitung kata umum yang ada pada *information retrieval*. Banyaknya jumlah kemunculan dari setiap kata akan berpengaruh pada nilai bobot dimana jika kata semakin banyak muncul di dalam suatu dokumen maka nilai bobotnya akan semakin meningkat. Proses pembobotan kata menggunakan modul Scikit-learn pada python.

#### 4.3.1 Menghitung nilai TF

Tahap pertama dalam metode TF-IDF adalah dengan menghitung nilai *term frequency* (TF) dari setiap kata. TF adalah jumlah kemunculan kata yang tersebar di dokumen yang digunakan. Sebagai contoh perhitungan nilai TF pada kata ‘hotel’ bernilai 2 dikarenakan kata ‘hotel’ muncul sebanyak 2 kali pada dokumen D1. Pada Tabel 4.9 ditunjukkan hasil dari perhitungan nilai *term frequency* (TF) untuk setiap *term* pada setiap dokumen.

Tabel 4.9 Hasil Perhitungan Nilai *Term Frequency* (TF)

Term	TF					
	D1	D2	D3	D4	D5	D6
hotel	2	1	2	0	0	0
ramah	1	0	1	0	0	0
sigap	1	0	0	0	0	0
bantu	1	0	1	0	0	0
pertama	1	0	0	0	0	0
sarapan	1	0	0	0	0	0
variatif	1	0	0	0	0	0
rasa	1	0	0	0	0	0
enak	1	0	2	0	0	0
harga	1	0	0	0	0	0
relatif	1	0	0	0	0	0

Term	TF					
	D1	D2	D3	D4	D5	D6
murah	1	0	0	0	0	0
lebih	0	1	0	0	0	0
bagus	0	1	0	0	0	1
minimarket	0	1	0	0	1	0
...	...	...	...	...	...	...
Kota	0	0	0	0	0	1

### 4.3.2 Menghitung nilai IDF

Setelah mendapatkan nilai *Term Frequency* (TF), selanjutnya akan dihitung nilai *Inverse Document Frequency* (IDF) dari setiap kata. Nilai *Inverse Document Frequency* (IDF) didapatkan dengan menghitung berdasarkan rumus pada persamaan 2.2. Berikut adalah contoh perhitungan nilai IDF pada kata ‘hotel’.

$$idf_t = \log_{10} \frac{6}{3} = 0,3010$$

Pada Tabel 4.10, ditunjukkan hasil dari perhitungan nilai IDF untuk setiap *term* pada setiap dokumen.

Tabel 4.10 Hasil Perhitungan Nilai *Inverse Document Frequency* (IDF)

Term	DF	IDF
hotel	3	0,3010
ramah	2	0,4771
sigap	1	0,7782
bantu	2	0,4771
pertama	1	0,7782
sarapan	1	0,7782
variatif	1	0,7782
rasa	1	0,7782
enak	2	0,4771
harga	1	0,7782
relatif	1	0,7782

<b>Term</b>	<b>DF</b>	<b>IDF</b>
murah	1	0,7782
lebih	1	0,7782
bagus	2	0,4771
minimarket	2	0,4771
...	...	...
Kota	1	0,7782

### 4.3.3 Menghitung nilai TF\*IDF

Selanjutnya dilakukan perhitungan nilai TF-IDF. Nilai TF-IDF didapatkan dengan mengkalikan hasil dari nilai *document frequency* (TF) dengan *inverse document frequency* (IDF). Berikut adalah contoh perhitungan bobot TF-IDF untuk kata 'hotel' pada D1 dan kata 'hotel' pada D2 :

$$Tf-idf_{t,d} = TF * IDF = 2 * 0,3010 = 0,60206$$

$$Tf-idf_{t,d} = TF * IDF = 1 * 0,3010 = 0,3010$$

Tabel 4.11, ditunjukkan hasil dari perhitungan bobot TF-IDF untuk setiap *term* pada setiap dokumen.

Tabel 4.11 Hasil Perhitungan Nilai Bobot TF-IDF

<b>Term</b>	<b>TF*IDF</b>					
	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>	<b>D5</b>	<b>D6</b>
hotel	0,60206	0,3010	0,602059	0	0	0
ramah	0,47712	0	0,477121	0	0	0
sigap	0,77815	0	0	0	0	0
bantu	0,47712	0	0,477121	0	0	0
pertama	0,77815	0	0	0	0	0
sarapan	0,77815	0	0	0	0	0
variatif	0,77815	0	0	0	0	0
rasa	0,77815	0	0	0	0	0
enak	0,47712	0	0,954242	0	0	0
harga	0,77815	0	0	0	0	0

Term	TF*IDF					
	D1	D2	D3	D4	D5	D6
relatif	0,77815	0	0	0	0	0
murah	0,77815	0	0	0	0	0
lebih	0	0,77815	0	0	0	0
bagus	0	0,47712	0	0	0	0,47712
minimarket	0	0,47712	0	0	0,47712	0
...	...	...	...	...	...	...
Kota	0	0	0	0	0	0,653

#### 4.4 Proses Perhitungan Support Vector Machine

Proses klasifikasi ulasan dilakukan dengan menggunakan metode *support vector machine* yang menggunakan dua kelas yaitu kelas positif (ulasan dengan opini baik) dan kelas negatif (ulasan dengan opini buruk). Tahapan pertama dalam perhitungan *support vector machine* adalah melakukan perhitungan kernel. Tahap selanjutnya adalah melakukan perhitungan *training* dengan metode *Sequential Training* SVM. Tahap selanjutnya adalah melakukan perhitungan *Testing* untuk mendapatkan hasil klasifikasi kelas dari data *testing* yang digunakan.

##### 4.4.1 Perhitungan Kernel

Tahap pertama yang dilakukan sebelum menghitung *training* adalah dengan melakukan perhitungan kernel. Fungsi kernel yang digunakan adalah kernel *polynomial* (*polynomial degree 2*). Perhitungan kernel *polynomial* menggunakan persamaan 2.9. Pada Tabel 4.12 ditunjukkan hasil dari perhitungan kernel.

Tabel 4.12 Hasil Perhitungan Kernel *polynomial*

K(xi,xi)	D1	D2	D3	D4	D5	D6
D1	87,611	1,558	4,838	1,000	1,000	1,000
D2	1,558	20,341	1,558	1,000	2,035	1,507
D3	4,838	1,558	36,393	1,000	2,035	1,000
D4	1,000	1,000	1,000	21,551	1,000	1,000
D5	1,000	2,035	2,035	1,000	100,974	1,000
D6	1,000	1,507	1,000	1,000	1,000	1,000

#### 4.4.2 Melakukan Inisialisasi Parameter

Tahap selanjutnya adalah melakukan perhitungan *training* dengan metode *Sequential Training SVM*. Pertama, dilakukan inisialisasi terhadap parameter yang akan digunakan, yaitu nilai  $C, \lambda, \gamma, \varepsilon$ , dan iterasi maksimum. Pada Tabel 4.13 diuraikan nilai parameter-parameter yang digunakan.

Tabel 4.13 Nilai Inisialisasi Parameter yang Digunakan

c	1
$\lambda$	0,5
$\gamma$	0,1
$\varepsilon$	0,0001
Iterasi Max	3

#### 4.4.3 Menghitung Matriks Hessian

Selanjutnya dilakukan inisialisasi nilai  $\alpha=0$  dan menghitung matriks hessian. Matriks hessian didapat dari perkalian antar kernel polynomial dan y dimana merupakan vector bernilai 1 dan -1. Perhitungan matriks hessian dilakukan dengan menggunakan persamaan 2.12. Pada Tabel 4.14 ditunjukkan hasil dari perhitungan matriks hessian.

Tabel 4.14 Hasil Perhitungan Matriks Hessian

$K(x_i, x_i)$	D1	D2	D3	D4	D5	D6
D1	87,861	-1,808	5,088	1,250	-1,250	1,250
D2	-1,808	20,591	-1,808	-1,250	2,285	-1,757
D3	5,088	-1,808	36,643	1,250	-2,285	1,250
D4	1,250	-1,250	1,250	21,801	-1,250	1,250
D5	-1,250	2,285	-2,285	-1,250	101,224	-1,250
D6	1,250	-1,757	1,250	1,250	-1,250	1,250

#### 4.4.3 Menghitung iterasi, $\delta\alpha_i$ dan $\alpha_i$

Setelah mendapatkan nilai matriks hessian maka akan dilakukan perhitungan iterasi,  $\delta\alpha_i$  dan memperbarui nilai  $\alpha_i$  dengan menggunakan persamaan 2.13, 2.14, 2.15. Pada awal perhitungan iterasi dimulai dari iterasi 0 dikarenakan nilai  $\alpha$  yang didapatkan dari hasil perhitungan masih belum diketahui.. Pada Tabel 4.15 ditunjukkan hasil dari perhitungan nilai E pada iterasi ke-0.

Tabel 4.15 Hasil Perhitungan Nilai E pada Iterasi ke-0

Iterasi 0	E <sub>i</sub>	D1	D2	D3	D4	D5	D6
		0	0	0	0	0	0

Selanjutnya menghitung nilai  $\delta\alpha$  dengan menggunakan Persamaan 2.14 untuk mendapatkan nilai  $\alpha$ . Pada Tabel 4.16 ditunjukkan hasil dari perhitungan  $\delta\alpha$  pada iterasi ke-0. Contoh perhitungan  $\delta\alpha$  pada D1 :

$$\delta\alpha_i = \min \{ \max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i \}$$

$$\delta\alpha_i = \min\{\max[0,1 (1 - 0), -0], 1 - 0\}$$

$$\delta\alpha_i = \min (0,1, 1)$$

$$\delta\alpha_i = 0,1$$

Tabel 4.16 Hasil Perhitungan Nilai  $\delta\alpha$  pada Iterasi ke-0

$\delta\alpha_i$	D1	D2	D3	D4	D5	D6
	0,1	0,1	0,1	0,1	0,1	0,1

Selanjutnya setelah didapatkan nilai  $\delta\alpha$  dilakukan perhitungan untuk mendapatkan  $\alpha_i$ . Pada Tabel 4.17, ditunjukkan hasil dari perhitungan  $\alpha$  pada iterasi ke-0. Contoh perhitungan  $\alpha$  pada D1 :

$$\alpha_i = \alpha_i + \delta\alpha_i$$

$$\alpha_1 = \alpha_0 + \delta\alpha_1$$

$$\alpha_i = 0 + 0,0001$$

$$\alpha_i = 0,0001$$

Tabel 4.17 Hasil Perhitungan Nilai  $\alpha$  pada Iterasi ke-0

alfa	D1	D2	D3	D4	D5	D6
$\alpha$	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

Selanjutnya dilakukan perhitungan terus menerus sampai mencapai iterasi maksimum agar mendapatkan nilai  $\alpha_i$  yang diperlukan untuk mencari *support vector*. Pada Tabel 4.18, ditunjukkan hasil dari perhitungan  $\alpha_i$  pada iterasi ke-3.

Tabel 4.18 Hasil Perhitungan Nilai  $\alpha$  pada Iterasi ke-3

alfa	D1	D2	D3	D4	D5	D6
$\alpha$	0,084263	0,091967	0,084263	0,082174	0,113948	0,264144

#### 4.4.4 Pemilihan Support Vector

Selanjutnya dilakukan perhitungan untuk mencari *support vector* masing-masing dokumen. Berdasarkan nilai  $\alpha_i$  terakhir yang didapat, maka selanjutnya diambil nilai  $\alpha_i$  terbesar dari masing-masing kelas. Pada Tabel 4.19 ditunjukkan hasil dari pemilihan  $\alpha_i$  terbesar dari masing-masing kelas.

Tabel 4.19 Hasil Dari Pemilihan  $\alpha_i$  Terbesar dari Masing-Masing Kelas

alfa	D1	D2	D3	D4	D5	D6
Kelas	1	-1	1	1	-1	1
$\alpha$	0	0	0	0	0	0
$\alpha$	0,1000	0,1000	0,1000	0,1000	0,1000	0,1000
$\alpha$	0,000000	0,037460	0,000000	0,000000	0,000000	0,180071
$\alpha$	0,084263	0,091967	0,084263	0,082174	0,113948	0,264144
X+ =	0,2641444					
X- =	0,1139476					

#### 4.4.5 Perhitungan Nilai Bias

Selanjutnya dilakukan perhitungan nilai bias b dengan menggunakan persamaan 2.16. Pada Tabel 4.20 ditunjukkan nilai dari bias.

Tabel 4 20 Hasil Perhitungan Nilai Bias

xi	K(xi,x+)	K(xi,x-)	$\alpha_i y_i K(xi,x+)$	$\alpha_i y_i K(xi,x-)$
D1	1,0000	1,5577	0,0842626	0,1312531
D2	1,5071	20,3410	-0,1386038	-1,8706916
D3	1,0000	1,5577	0,0842626	0,1312531
D4	1,0000	1,0000	0,0821736	0,0821736
D5	1,0000	2,0354	-0,1139476	-0,2319330
D6	1,0000	1,5071	0,2641444	0,3980950
Total	$\sum_{(i=0)}^N (\alpha_i * y_i (K(xi,x+)))$		0,2622918	
	$\sum_{(i=0)}^N (\alpha_i * y_i (K(xi,x-)))$			-1,3598497
bias	-0,548778948			

#### 4.4.6 Perhitungan Testing

Setelah dilakukan perhitungan training, maka selanjutnya barulah dilakukan perhitungan *Testing*. *Testing* yang dijelaskan di sini sebanyak 3 dokumen yang diambil dari dokumen training nomer 1, 2 dan 3. Perhitungan testing dilakukan dengan menggunakan persamaan 2.17, dimana Nilai  $f(x)$  didapat dari hasil penjumlahan  $\alpha_i$  dikali hasil perhitungan kernel dikali kelas ditambah b. Jika fungsi bernilai positif maka dokumen akan diklasifikasikan pada kelas sentimen positif, sebaliknya jika fungsi bernilai negatif maka akan diklasifikasikan pada kelas sentimen negatif. Pada Tabel 4.21, ditunjukkan hasil klasifikasi *support vector machine* terhadap data 3 data *testing*.

Tabel 4.21 Hasil Data Testing

<b>SV</b>	<b>T1</b>	<b>T2</b>	<b>T3</b>
<b>D1</b>	0,0027310	0,0027310	0,0049897
<b>D2</b>	-0,0060076	-0,0085636	-0,0066067
<b>D3</b>	0,0058647	0,0083599	0,0064495
<b>D4</b>	0,0029314	0,0029314	0,0029314
<b>D5</b>	-0,0027172	-0,0027172	-0,0027172
<b>D6</b>	-0,0029941	-0,0029941	-0,0029941
<b>Total</b>	-0,0001918	-0,0002526	0,0020526
<b>Total + bias</b>	-0,0226966	-0,0002526	0,0020526
<b>Kelas</b>	<b>-1</b>	<b>-1</b>	<b>1</b>

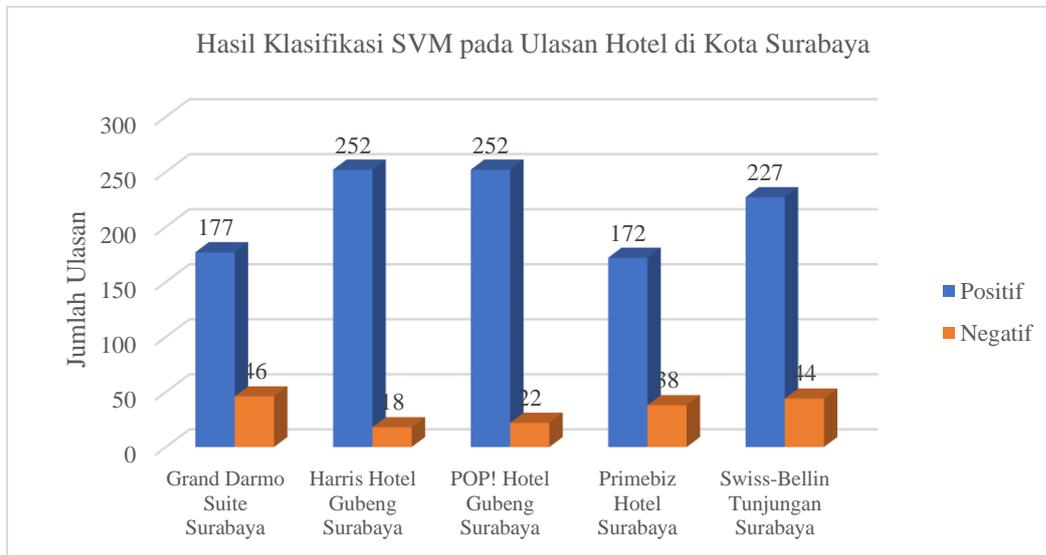
Berdasarkan hasil perhitungan data testing yang terdapat pada Tabel 4.20, data testing 1 mendapatkan nilai  $f(x)$  sebesar -0,0226966 sehingga diklasifikasikan ke dalam kelas negatif (-1), data testing 2 mendapatkan nilai  $f(x)$  sebesar -0,0002526 sehingga juga diklasifikasikan ke dalam kelas negatif (-1), Sedangkan data testing 3 mendapatkan nilai  $f(x)$  sebesar 0,0020526 sehingga diklasifikasikan ke dalam kelas positif (1).

#### 4.5 Hasil Klasifikasi Support Vector Machine

Setelah dilakukan klasifikasi dengan menggunakan *support vector machine*, didapatkan hasil sentimen ulasan pelanggan hotel pada situs TripAdvisor di kota Surabaya dan Malang.

#### 4.5.1 Hasil Klasifikasi pada Ulasan di Kota Surabaya

Berikut adalah hasil klasifikasi sentimen dengan menggunakan metode *support vector machine* terhadap data ulasan wisatawan pada 5 hotel di kota Surabaya.

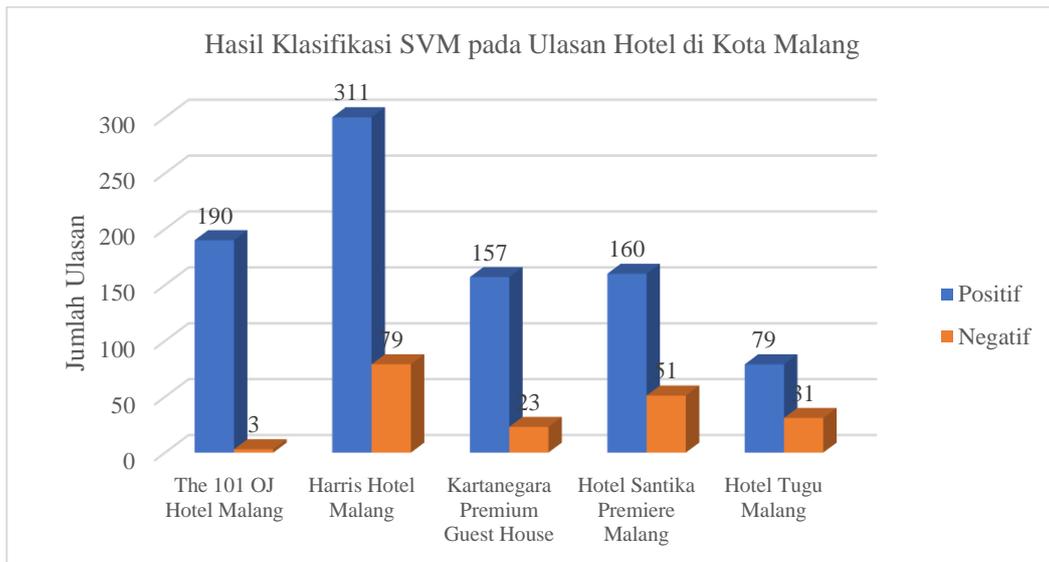


Gambar 4.1 Grafik Hasil Klasifikasi SVM pada Ulasan Hotel di Kota Surabaya

Berdasarkan Gambar 4.1, dapat disimpulkan hotel di kota Surabaya yang memiliki paling banyak ulasan dengan sentimen positif adalah Harris Hotel Gubeng dan Pop! Hotel Gubeng dengan jumlah ulasan sama yaitu 252 ulasan. Sedangkan hotel yang memiliki paling sedikit ulasan dengan sentimen positif adalah Primebiz Hotel Surabaya dengan total 172 ulasan. Selanjutnya, hotel yang memiliki paling banyak sentimen negatif adalah Grand Darmo Suite dengan total 46 ulasan. Sedangkan hotel yang memiliki paling sedikit sentimen negatif adalah Harris Hotel Gubeng dengan total 18 ulasan.

#### 4.5.2 Hasil Klasifikasi pada Ulasan di Kota Malang

Berikut adalah hasil klasifikasi sentimen dengan menggunakan metode *support vector machine* terhadap data ulasan wisatawan pada 5 hotel di kota Malang.



Gambar 4.2 Grafik Hasil Klasifikasi SVM pada Ulasan Hotel di Kota Malang

Berdasarkan Gambar 4.2, dapat disimpulkan hotel di kota Malang yang memiliki paling banyak ulasan dengan sentimen positif adalah Harris Hotel Malang dengan total 311 ulasan. Sedangkan hotel yang memiliki paling sedikit sentimen positif adalah Hotel Tugu Malang dengan total 79 ulasan. Selanjutnya, hotel yang memiliki paling banyak sentimen negatif adalah Harris Hotel Malang dengan total 79 ulasan. Sedangkan hotel yang memiliki paling sedikit sentimen negatif adalah The 101 OJ Hotel dengan total 3 ulasan.

#### 4.6 Hasil Klasifikasi Support Vector Machine Berdasarkan Aspek

Langkah selanjutnya adalah melakukan analisis lanjutan dengan melihat hasil sentimen berdasarkan kategori aspek dari ulasan, sehingga dapat diketahui kategori aspek yang memiliki paling banyak sentimen positif dan sentimen negatif. Terdapat 3 kategori aspek yang digunakan yaitu lokasi, kebersihan dan layanan.

##### 4.6.1 Analisis Hasil Klasifikasi pada Ulasan Berdasarkan Aspek di Kota Surabaya

Hasil klasifikasi sentimen yang dikategorikan ke dalam tiga aspek pada ulasan di 5 hotel kota Surabaya dapat dilihat pada Tabel 4.22.

Tabel 4.22 Hasil Klasifikasi Sentimen Berdasarkan Aspek Di Kota Surabaya

No.	Nama Hotel	Kategori Aspek	Sentimen	Jumlah Ulasan	Nilai Persentase Dari Total Ulasan
1.	Grand Darmo Suite Surabaya	Aspek Lokasi	Positif	49 Ulasan	22%
			Negatif	13 Ulasan	5,8%
		Aspek Kebersihan	Positif	39 Ulasan	17,5%
			Negatif	5 Ulasan	2,2%
		Aspek Layanan	Positif	89 Ulasan	39,9%
			Negatif	28 Ulasan	12,6%
2.	Harris Hotel Gubeng Surabaya	Aspek Lokasi	Positif	49 Ulasan	18,1%
			Negatif	4 Ulasan	1,5%
		Aspek Kebersihan	Positif	70 Ulasan	25,9%
			Negatif	3 Ulasan	1,1%
		Aspek Layanan	Positif	133 Ulasan	49,3%
			Negatif	11 Ulasan	4,1%
3.	POP! Hotel Gubeng Surabaya	Aspek Lokasi	Positif	38 Ulasan	13,9%
			Negatif	4 Ulasan	1,5%
		Aspek Kebersihan	Positif	79 Ulasan	28,8%
			Negatif	4 Ulasan	1,5%
		Aspek Layanan	Positif	135 Ulasan	49,3%
			Negatif	14 Ulasan	5,1%
4.	Primebiz Hotel Surabaya	Aspek Lokasi	Positif	35 Ulasan	16,7%
			Negatif	7 Ulasan	3,3%
		Aspek Kebersihan	Positif	49 Ulasan	23,3%
			Negatif	12 Ulasan	5,7%
		Aspek Layanan	Positif	88 Ulasan	41,9%
			Negatif	19 Ulasan	9%
5.	Swiss-Bellin Tunjungan Surabaya	Aspek Lokasi	Positif	68 Ulasan	25,1%
			Negatif	9 Ulasan	3,3%
		Aspek Kebersihan	Positif	39 Ulasan	14,4%

No.	Nama Hotel	Kategori Aspek	Sentimen	Jumlah Ulasan	Nilai Persentase Dari Total Ulasan
			Negatif	4 Ulasan	1,5%
		Aspek Layanan	Positif	120 Ulasan	44,3%
			Negatif	31 Ulasan	11,4%

1. Pada Grand Darmo Suite Surabaya, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 89 ulasan atau 39,9% dari total ulasan.
2. Pada Harris Hotel Gubeng Surabaya, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 133 ulasan atau 49,3% dari total ulasan.
3. Pada ulasan POP! Hotel Gubeng Surabaya, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 135 ulasan atau 49,3% dari total ulasan.
4. Pada ulasan Primebiz Hotel Surabaya, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 88 ulasan atau 41,9% dari total ulasan.
5. Pada ulasan Swiss-Bellin Tunjungan Surabaya, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 120 ulasan atau 44,3% dari total ulasan.

Berdasarkan hasil pada Tabel 4.22, dapat disimpulkan bahwa aspek terbanyak yang diulas oleh pelanggan pada 5 hotel di kota Surabaya adalah aspek layanan. Aspek layanan adalah aspek yang paling banyak memuaskan pengunjung hotel di kota Surabaya.

#### **4.6.2 Analisis Hasil Klasifikasi pada Ulasan Berdasarkan Aspek di Kota Malang**

Hasil klasifikasi sentimen yang dikategorikan ke dalam tiga aspek pada ulasan di 5 hotel kota Malang dapat dilihat pada Tabel 4.23.

Tabel 4.23 Hasil Klasifikasi Sentimen Berdasarkan Aspek Di Kota Malang

No.	Nama Hotel	Kategori Aspek	Sentimen	Jumlah Ulasan	Nilai Persentase Dari Total Ulasan
1.	The 101 OJ Hotel Malang	Aspek Lokasi	Positif	21 Ulasan	10,9%
			Negatif	0 Ulasan	0%
		Aspek Kebersihan	Positif	55 Ulasan	28,5%
			Negatif	0 Ulasan	0%
		Aspek Layanan	Positif	114 Ulasan	59,1%
			Negatif	3 Ulasan	1,6%
2.	Harris Hotel Malang	Aspek Lokasi	Positif	50 Ulasan	12,8%
			Negatif	27 Ulasan	6,9%
		Aspek Kebersihan	Positif	77 Ulasan	19,7%
			Negatif	15 Ulasan	3,8%
		Aspek Layanan	Positif	184 Ulasan	47,2%
			Negatif	37 Ulasan	9,5%
3.	Kartanegara Premium Guest House Malang	Aspek Lokasi	Positif	45 Ulasan	25%
			Negatif	2 Ulasan	1,1%
		Aspek Kebersihan	Positif	49 Ulasan	27,2%
			Negatif	3 Ulasan	1,7%
		Aspek Layanan	Positif	63 Ulasan	35%
			Negatif	18 Ulasan	10%
4.	Hotel Santika Premiere Malang	Aspek Lokasi	Positif	30 Ulasan	14,2%
			Negatif	10 Ulasan	4,7%
		Aspek Kebersihan	Positif	28 Ulasan	13,3%
			Negatif	3 Ulasan	1,4%
		Aspek Layanan	Positif	102 Ulasan	48,3%
			Negatif	38 Ulasan	18%
5.	Hotel Tugu Malang	Aspek Lokasi	Positif	22 Ulasan	20%
			Negatif	2 Ulasan	1,8%
		Aspek Kebersihan	Positif	13 Ulasan	11,8%

No.	Nama Hotel	Kategori Aspek	Sentimen	Jumlah Ulasan	Nilai Persentase Dari Total Ulasan
			Negatif	0 Ulasan	0%
		Aspek Layanan	Positif	44 Ulasan	40%
			Negatif	29 Ulasan	26,4%

1. Pada The 101 OJ Hotel Malang, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 114 ulasan atau 59,1% dari total ulasan.
2. Pada Harris Hotel Malang, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 184 ulasan atau 47,2% dari total ulasan.
3. Pada Kartanegara Premium Guest House Malang, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 63 ulasan atau 35% dari total ulasan.
4. Pada Hotel Santika Premiere Malang, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 102 ulasan atau 48,3% dari total ulasan.
5. Pada Hotel Tugu Malang, aspek yang mendapatkan sentimen positif tertinggi adalah aspek layanan dengan 44 ulasan atau 40% dari total ulasan.

Berdasarkan hasil pada Tabel 4.23, dapat disimpulkan bahwa aspek terbanyak yang diulas oleh pelanggan pada 5 hotel di kota Malang adalah aspek layanan. Aspek layanan adalah aspek yang paling banyak memuaskan pengunjung hotel di kota Malang.

#### **4.7 Hasil Perbandingan Rating Klasifikasi Sentimen dengan Rating pada Situs TripAdvisor**

Selanjutnya hasil klasifikasi sentimen ulasan berdasarkan aspek dibandingkan dengan nilai rating tiap aspek pada situs TripAdvisor. Analisis ini bertujuan untuk membandingkan nilai rating yang diberikan TripAdvisor dengan nilai rating klasifikasi sentimen ulasan berdasarkan aspek. Nilai rating pada situs TripAdvisor berkisar antara skala 1 sampai dengan 5 sehingga data hasil klasifikasi sentimen akan di *convert* terlebih dahulu agar memiliki range nilai yang sama. Detail dari range nilai pada TripAdvisor dapat dilihat pada tabel 4.24 dan penilaian rating pada TripAdvisor dapat dilihat pada Gambar 4.3.

Tabel 4.24 Range Nilai Pada Situs TripAdvisor

Range Nilai	Penilaian
0 - 1	Sangat Buruk
1,1 - 2	Buruk
2,1 - 3	Rata-rata
3,1 - 4	Bagus
4,1 - 5	Luar Biasa



Gambar 4.3 Penilaian Rating Pada TripAdvisor

#### 4.7.1 Analisis Hasil Perbandingan Rating Klasifikasi Sentimen dengan Rating pada Situs TripAdvisor di Kota Surabaya

Hasil perbandingan rating klasifikasi sentimen dengan rating situs TripAdvisor pada 5 hotel di kota Surabaya dapat dilihat pada Tabel 4.25 dan perhitungan lengkap dari rating dapat dilihat pada lampiran 1.

Tabel 4.25 Hasil Perbandingan Rating Klasifikasi Sentimen dan rating TripAdvisor di 5 hotel kota Surabaya

No.	Nama Hotel	Aspek	Rating Klasifikasi Sentimen	Rating TripAdvisor
1.	Grand Darmo Suite Surabaya	Lokasi	3,95	4
		Kebersihan	4,45	4,5
		Layanan	3,8	4
	Nilai Rating Secara Umum		3,95	4
2.	Harris Hotel Gubeng Surabaya	Lokasi	4,6	4,5
		Kebersihan	4,8	4,5
		Layanan	4,6	4,5

No.	Nama Hotel	Aspek	Rating Klasifikasi Sentimen	Rating TripAdvisor
	Nilai Rating Secara Umum		4,65	4,5
3.	POP! Hotel Gubeng Surabaya	Lokasi	4,8	4,5
		Kebersihan	4,4	4,5
		Layanan	4,6	4,5
	Nilai Rating Secara Umum		4,6	4,5
4.	Primebiz Hotel Surabaya	Lokasi	4,8	4,5
		Kebersihan	4,4	4,5
		Layanan	4,1	5
	Nilai Rating Secara Umum		4,1	5
5.	Swiss-Bellin Tunjungan Surabaya	Lokasi	4,4	4,5
		Kebersihan	4,5	4,5
		Layanan	4	4,5
	Nilai Rating Secara Umum		4,2	4,5

Berdasarkan hasil perbandingan klasifikasi sentimen dan rating TripAdvisor pada Tabel 4.25, dapat disimpulkan bahwa :

1. Pada Grand Darmo Suite Surabaya, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 3,95 (bagus) dan *rating* aspek lokasi pada TripAdvisor sebesar 4 (bagus). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu bagus. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,45 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 3,8 (bagus) dan *rating* aspek layanan pada TripAdvisor sebesar 4 (bagus), Penilaian pada TripAdvisor untuk aspek layanan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu bagus.
2. Pada Harris Hotel Gubeng Surabaya, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,6 (luar biasa) dan *rating* aspek lokasi pada

TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,8 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 4,6 (luar biasa) dan *rating* aspek layanan pada TripAdvisor sebesar 4,5 (luar biasa), Penilaian pada TripAdvisor untuk aspek layanan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa.

3. Pada POP! Hotel Gubeng Surabaya, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,8 (luar biasa) dan *rating* aspek lokasi pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,4 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 4,6 (luar biasa) dan *rating* aspek layanan pada TripAdvisor sebesar 4,5 (luar biasa), Penilaian pada TripAdvisor untuk aspek layanan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa.
4. Pada Primebiz Hotel Surabaya, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,8 (luar biasa) dan *rating* aspek lokasi pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,4 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 4,1 (luar biasa) dan *rating* aspek layanan pada TripAdvisor sebesar 5 (luar

biasa), Penilaian pada TripAdvisor untuk aspek layanan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa.

5. Pada Swiss-Bellin Tunjungan Surabaya, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,4 (luar biasa) dan *rating* aspek lokasi pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,5 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 4 (bagus) dan *rating* aspek layanan pada TripAdvisor sebesar 4,5 (luar biasa), Penilaian pada TripAdvisor untuk aspek layanan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian luar biasa dan klasifikasi sentimen memberikan penilaian bagus.

Berdasarkan hasil perhitungan rating dapat diketahui urutan peringkat hotel pada 5 hotel di kota Surabaya. Hotel yang mendapatkan peringkat pertama adalah Harris Hotel Gubeng Surabaya dengan nilai rating 4,65 (Luar Biasa), peringkat kedua adalah POP! Hotel Gubeng Surabaya dengan nilai rating 4,6 (Luar Biasa), peringkat ketiga Swiss-Bellin Tunjungan Surabaya dengan nilai rating 4,2 (Luar Biasa), peringkat keempat Primebiz Hotel Surabaya dengan nilai rating 4,1 (Luar Biasa), dan yang mendapatkan peringkat kelima atau terakhir adalah Grand Darmo Suite Surabaya dengan nilai rating 3,95 (bagus). Peringkat hotel berdasarkan hasil rating klasifikasi sentimen di kota Surabaya dapat dilihat pada Tabel 4.26.

Tabel 4.26 Peringkat Berdasarkan Hasil Rating Klasifikasi Sentimen di Kota Surabaya

<b>Peringkat</b>	<b>Rating Klasifikasi Sentimen</b>	<b>Nama Hotel</b>
1	4,65 (Luar Biasa)	Harris Hotel Gubeng Surabaya
2	4,6 (Luar Biasa)	POP! Hotel Gubeng Surabaya

Peringkat	Rating Klasifikasi Sentimen	Nama Hotel
3	4,2 (Luar Biasa)	Swiss-Bellin Tunjungan Surabaya
4	4,1 (Luar Biasa)	Primebiz Hotel Surabaya
5	3,95 (bagus)	Grand Darmo Suite Surabaya

#### 4.7.1 Analisis Hasil Perbandingan Rating Klasifikasi Sentimen dengan Rating pada Situs TripAdvisor di Kota Malang

Hasil perbandingan rating klasifikasi sentimen dengan rating situs TripAdvisor pada 5 hotel di kota Malang dapat dilihat pada Tabel 4.27 dan perhitungan lengkap dari rating dapat dilihat pada lampiran 2.

Tabel 4.27 Hasil Perbandingan Rating Klasifikasi Sentimen dan rating TripAdvisor pada 5 hotel di kota Malang

No.	Nama Hotel	Aspek	Rating Klasifikasi Sentimen	Rating TripAdvisor
1.	The 101 OJ Hotel Malang	Lokasi	4,4	4
		Kebersihan	5	4
		Layanan	4,85	4
	Nilai Rating Secara Umum		4,9	4,5
2.	Harris Hotel Malang	Lokasi	3,25	4
		Kebersihan	4,2	4
		Layanan	4,15	4
	Nilai Rating Secara Umum		4	4,5
3.	Kartanegara Premium Guest House Malang	Lokasi	4,8	4,5
		Kebersihan	4,7	4,5
		Layanan	3,9	4,5
	Nilai Rating Secara Umum		4,35	4,5
4.		Lokasi	3,75	4

No.	Nama Hotel	Aspek	Rating Klasifikasi Sentimen	Rating TripAdvisor
	Hotel Santika Premiere Malang	Kebersihan	4,5	4
		Layanan	3,65	4
	Nilai Rating Secara Umum		3,8	4
5.	Hotel Tugu Malang	Lokasi	4,6	4,5
		Kebersihan	5	4,5
		Layanan	3	4,5
	Nilai Rating Secara Umum		3,6	4,5

Berdasarkan hasil perbandingan klasifikasi sentimen dan rating TripAdvisor pada Tabel 4.27, dapat disimpulkan bahwa :

1. Pada The 101 OJ Hotel Malang, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,4 (luar biasa) dan *rating* aspek lokasi pada TripAdvisor sebesar 4 (bagus). Penilaian pada TripAdvisor untuk aspek lokasi tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian bagus dan klasifikasi sentimen memberikan penilaian luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 5 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4 (bagus). Penilaian pada TripAdvisor untuk aspek kebersihan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian bagus dan klasifikasi sentimen memberikan penilaian luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 4,85 (luar biasa) dan *rating* aspek layanan pada TripAdvisor sebesar 4 (bagus), Penilaian pada TripAdvisor untuk aspek layanan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian bagus dan klasifikasi sentimen memberikan penilaian luar biasa.
2. Pada Harris Hotel Malang, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 3,25 (bagus) dan *rating* aspek lokasi pada TripAdvisor sebesar 4 (bagus). Penilaian pada TripAdvisor untuk aspek lokasi telah

sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu bagus. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,2 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4 (bagus). Penilaian pada TripAdvisor untuk aspek kebersihan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian bagus dan klasifikasi sentimen memberikan penilaian luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 4,15 (luar biasa) dan *rating* aspek layanan pada TripAdvisor sebesar 4 (bagus), Penilaian pada TripAdvisor untuk aspek layanan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian bagus dan klasifikasi sentimen memberikan penilaian luar biasa.

3. Pada Kartanegara Premium Guest House Malang, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,8 (luar biasa) dan *rating* aspek lokasi pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,7 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 3,9 (bagus) dan *rating* aspek layanan pada TripAdvisor sebesar 4,5 (luar biasa), Penilaian pada TripAdvisor untuk aspek layanan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian luar biasa dan klasifikasi sentimen memberikan penilaian bagus.
4. Pada Hotel Santika Premiere Malang, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 3,75 (bagus) dan *rating* aspek lokasi pada TripAdvisor sebesar 4 (bagus). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu bagus. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 4,5 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4

(bagus). Penilaian pada TripAdvisor untuk aspek kebersihan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian bagus dan klasifikasi sentimen memberikan penilaian luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 3,65 (bagus) dan *rating* aspek layanan pada TripAdvisor sebesar 4 (bagus), Penilaian pada TripAdvisor untuk aspek layanan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu bagus.

5. Pada Hotel Tugu Malang, Berdasarkan klasifikasi sentimen aspek lokasi mendapatkan nilai *rating* sebesar 4,6 (luar biasa) dan *rating* aspek lokasi pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek lokasi telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek kebersihan mendapatkan nilai *rating* sebesar 5 (luar biasa) dan *rating* aspek kebersihan pada TripAdvisor sebesar 4,5 (luar biasa). Penilaian pada TripAdvisor untuk aspek kebersihan telah sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang sama yaitu luar biasa. Berdasarkan klasifikasi sentimen aspek layanan mendapatkan nilai *rating* sebesar 3 (bagus) dan *rating* aspek layanan pada TripAdvisor sebesar 4,5 (luar biasa), Penilaian pada TripAdvisor untuk aspek layanan tidak sesuai dengan hasil klasifikasi sentimen karena memiliki penilaian yang berbeda, dimana TripAdvisor memberikan penilaian luar biasa dan klasifikasi sentimen memberikan penilaian bagus.

Berdasarkan hasil perhitungan *rating* dapat diketahui urutan peringkat hotel pada 5 hotel di kota Malang. Hotel yang mendapatkan peringkat pertama adalah The 101 OJ Hotel Malang dengan nilai *rating* 4,9 (Luar Biasa), peringkat kedua adalah Kartanegara Premium Guest House Malang dengan nilai *rating* 4,35 (Luar Biasa), peringkat ketiga Harris Hotel Malang dengan nilai *rating* 4 (Bagus), peringkat keempat Hotel Santika Premiere Malang dengan nilai *rating* 3,8 (Bagus), dan yang mendapatkan peringkat kelima atau terakhir adalah Hotel Tugu Malang dengan nilai *rating* 3,6 (Bagus). Peringkat hotel berdasarkan hasil *rating* klasifikasi sentimen di kota Malang dapat dilihat pada Tabel 4.28.

Tabel 4.28 Peringkat Berdasarkan Hasil Rating Klasifikasi Sentimen di Kota Malang

Peringkat	Rating Klasifikasi Sentimen	Nama Hotel
1	4,9 (Luar Biasa)	The 101 OJ Hotel Malang
2	4,35 (Luar Biasa)	Kartanegara Premium Guest House Malang
3	4 (Bagus)	Harris Hotel Malang
4	3,8 (Bagus)	Hotel Santika Premiere Malang
5	3,6 (bagus)	Hotel Tugu Malang

#### 4.8 Hasil Evaluasi

Tingkat keberhasilan dari classifier dapat diukur dengan melakukan evaluasi. Pada umumnya *confusion matrix* digunakan untuk mengukur evaluasi dan dilakukan dengan mengukur nilai akurasi, precision, recall dan F1-Score. Tes akurasi dilakukan untuk mengukur keberhasilan classifier dalam menentukan label data. *Precision* adalah pengukuran untuk dapat mengukur ketepatan sistem dan *Recall* adalah pengukuran untuk mendapatkan keberhasilan sistem. Sedangkan *F1 Score* atau *F-Measure* adalah perhitungan yang merupakan harmonic mean dari *Precision* dan *Recall*.

##### 4.8.1 Hasil Evaluasi Data Ulasan Hotel di Kota Surabaya

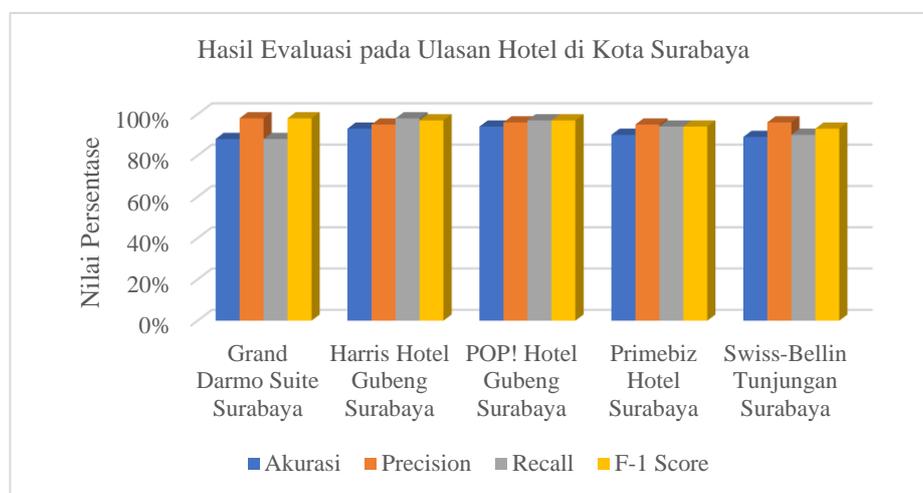
Tabel 4.29 menunjukkan hasil evaluasi data ulasan pada masing-masing hotel di kota Surabaya. Pada Gambar 4.4 dapat disimpulkan bahwa, Nilai akurasi tertinggi terdapat pada data POP! Hotel Gubeng Surabaya dengan tingkat akurasi sebesar 94% dan nilai akurasi terendah terdapat pada data Grand Darmo Suite Surabaya dengan tingkat akurasi sebesar 88%. Nilai *precision* tertinggi terdapat pada data Grand Darmo Suite Surabaya dengan nilai *precision* sebesar 98% dan terendah pada data Harris Hotel Gubeng Surabaya dan Primebiz Hotel Surabaya dengan nilai *precision* sebesar 95%. Nilai *recall* tertinggi terdapat pada data Harris Hotel Gubeng Surabaya dengan nilai *recall* sebesar 98% dan terendah pada data Grand Darmo Suite Surabaya dengan nilai *recall* sebesar 95%. Nilai *F1-Score* tertinggi terdapat pada data Grand Darmo Suite

Surabaya dengan nilai *F1-Score* sebesar 98% dan terendah pada data Swiss-Bellin Tunjungan Surabaya dengan nilai *F1-Score* sebesar 93%. Hasil perhitungan evaluasi pada 5 hotel di kota Surabaya secara lengkap dapat dilihat pada lampiran 3.

Tabel 4.29 Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Surabaya

No.	Nama Hotel	Akurasi	Precision	Recall	F1-Score
1.	Grand Darmo Suite Surabaya	88%	98%	88%	98%
2.	Harris Hotel Gubeng Surabaya	93%	95%	98%	97%
3.	POP! Hotel Gubeng Surabaya	94%	96%	97%	97%
4.	Primebiz Hotel Surabaya	90%	95%	94%	94%
5.	Swiss-Bellin Tunjungan Surabaya	89%	96%	90%	93%

Pada Gambar 4.4 ditampilkan grafik hasil evaluasi data ulasan pada 5 hotel di kota Surabaya.



Gambar 4.4 Grafik Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Surabaya

#### 4.8.2 Hasil Evaluasi Data Ulasan Hotel di Kota Malang

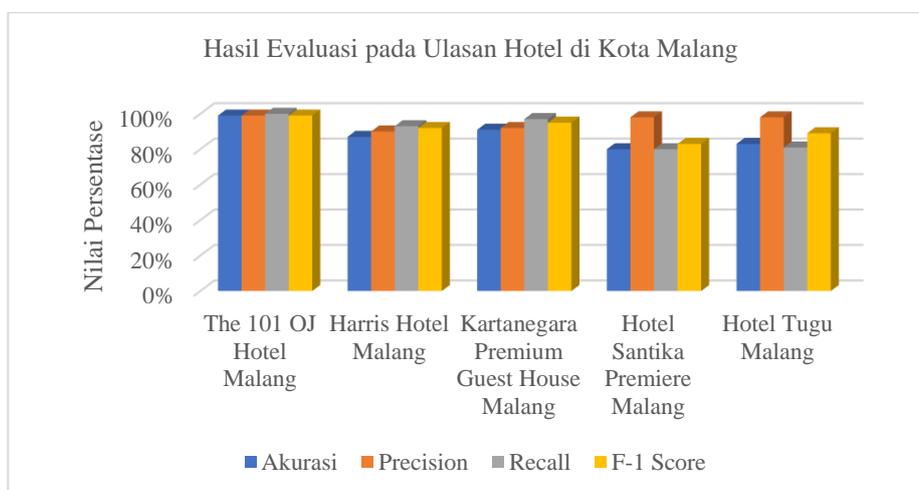
Tabel 4.30 menunjukkan hasil evaluasi data ulasan pada masing-masing hotel di kota Malang. Pada Gambar 4.5 dapat disimpulkan bahwa, Nilai akurasi tertinggi terdapat pada data The 101 OJ Hotel Malang dengan tingkat akurasi sebesar 99% dan terendah pada data Hotel Santika Premiere Malang dengan tingkat akurasi sebesar 80%.

Nilai *precision* tertinggi terdapat pada data The 101 OJ Hotel Malang dengan nilai *precision* sebesar 99% dan terendah pada data Harris Hotel Malang dengan nilai *precision* sebesar 90%. Nilai *recall* tertinggi terdapat pada data The 101 OJ Hotel Malang dengan nilai *recall* sebesar 100% dan terendah pada data Hotel Santika Premiere Malang dengan nilai *recall* sebesar 80%. Nilai *F1-Score* tertinggi terdapat pada data The 101 OJ Hotel Malang dengan nilai *F1-Score* sebesar 99% dan terendah pada data Hotel Santika Premiere Malang dengan nilai *F1-Score* sebesar 83%. Hasil perhitungan evaluasi pada 5 hotel di kota Malang secara lengkap dapat dilihat pada lampiran 4.

Tabel 4.30 Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Malang

No.	Nama Hotel	Akurasi	Precision	Recall	F1-Score
1.	The 101 OJ Hotel Malang	99%	99%	100%	99%
2.	Harris Hotel Malang	87%	90%	93%	92%
3.	Kartanegara Premium Guest House Malang	92%	92%	97%	95%
4.	Hotel Santika Premiere Malang	80%	98%	80%	83%
5.	Hotel Tugu Malang	83%	98%	81%	89%

Pada Gambar 4.5 ditampilkan grafik hasil evaluasi data ulasan pada 5 hotel di kota Malang.



Gambar 4.5 Grafik Hasil Evaluasi Data Ulasan pada 5 Hotel di Kota Malang

BAB V  
KESIMPULAN DAN  
SARAN

## **BAB 5**

### **KESIMPULAN DAN SARAN**

Bab ini menampilkan kesimpulan dari penelitian serta akan dituliskan saran bagi penelitian selanjutnya.

#### **5.1 Kesimpulan**

- 6 Metode support vector machine yang dikombinasikan dengan TF-IDF dapat menyelesaikan permasalahan dalam klasifikasi sentimen. Hal ini dibuktikan dengan kemampuan TF-IDF dalam memberikan nilai bobot pada suatu kata dan kemampuan Support vector machine dalam memberikan label sentimen pada masing-masing ulasan, yaitu ulasan yang positif dan ulasan yang negatif.
- 7 Pada data 5 hotel di kota Surabaya, hotel yang memiliki paling banyak sentimen positif adalah Harris Hotel Gubeng dan Pop! Hotel Gubeng dengan jumlah ulasan sama yaitu 252 ulasan. Pada sentimen negatif, hotel yang memiliki paling banyak sentimen negatif adalah Grand Darmo Suite dengan total 46 ulasan. Pada data 5 hotel di kota Malang, hotel yang memiliki paling banyak sentimen positif adalah Harris Hotel Malang dengan total 311 ulasan. Pada sentimen negatif, hotel yang memiliki paling banyak sentimen negatif adalah Harris Hotel Malang dengan total 79 ulasan.
- 8 Pada hasil klasifikasi sentimen pada ulasan berdasarkan aspek, dari 3 kategori aspek yang digunakan, aspek yang paling banyak mendapatkan ulasan positif (ulasan dengan opini yang baik) adalah aspek layanan. Dimana dari 10 hotel yang dianalisis di kota kota Surabaya dan Malang, semua hasil menunjukkan jika aspek layanan adalah aspek yang paling banyak diulas oleh pelanggan hotel.
- 9 Pada hasil perhitungan rating berdasarkan klasifikasi sentimen, Hotel di kota Surabaya yang berada pada peringkat pertama adalah Harris Hotel Gubeng Surabaya dengan nilai rating 4,65 (Luar Biasa). Sedangkan hotel di kota Malang yang berada pada peringkat pertama adalah The 101 OJ Hotel Malang dengan nilai rating 4,9 (Luar Biasa).
- 10 Pada hasil tes akurasi, nilai akurasi tertinggi pada data hotel di kota Surabaya terdapat pada data POP! Hotel Gubeng Surabaya dengan tingkat akurasi sebesar 94%, sedangkan nilai akurasi tertinggi pada data hotel di kota Malang terdapat pada

data The 101 OJ Hotel Malang dengan tingkat akurasi sebesar 99%. Dengan tingkat akurasi demikian juga dapat disimpulkan jika classifier yang digunakan sudah berkerja dengan baik dalam mengklasifikasikan ulasan.

- 11 Sentimen Analisis pada ulasan hotel dapat dilakukan untuk mengetahui bagaimana opini dari pelanggan hotel, sehingga bisa memberikan kontribusi bagi pengelola hotel dalam meningkatkan pelayanan untuk kedepannya dan juga agar dapat membangun opini yang baik di kalangan wisatawan bagi tempat yang mereka kelola sehingga dapat meningkatkan jumlah kunjungan.

### **11.1Saran**

1. Pada proses pemilihan aspek sebaiknya dapat menggunakan bantuan dari sebuah metode atau dapat menggunakan pendekatan ekstraksi aspek dengan metode Part-of-Speech Tagger yang dapat memisahkan ulasan secara otomatis berdasarkan kamus data pada setiap aspeknya. Hal ini dikarenakan proses pelabelan aspek yang dilakukan secara manual dapat menghabiskan banyak waktu jika data yang harus dilabeli sangat banyak
2. Pada tahap pre-processing dapat ditambah suatu tahapan lagi untuk memperbaiki penggunaan kata yang tidak baku dan untuk dapat mentranlasi kata bahasa asing yang juga terkandung di dalam ulasan sehingga lebih dapat mengoptimalkan tahap klasifikasi dikarenakan pada penelitian ini penelitian ini ulasan yang dianalisis adalah ulasan bahasa Indonesia.

# DAFTAR PUSTAKA

## DAFTAR PUSTAKA

- Akbari, M. I. H. A. D., Astri Novianty S.T., M. & Casi Setianingsih S.T., M., 2012. Analisis Sentimen Menggunakan Metode Learning Vector Quantization. Telkom University.
- Alma, Buchari. 2007. *Manajemen Pemasaran dan Pemasaran Jasa*. Bandung: CV Alfabeta.
- Banados J., A., Espinosa K., J., 2014. *Optimizing Support Vector Machine in Classifying Sentiments on Product Brands from Twitter. IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*.
- Basari, 2013. *Opinion mining of movie review using hybrid method of support vector machine and partical swarm optimazion. Procedia Engineering, Volume 53, pp. 453-462*.
- Chory, R., N., Muhammad Nasrun, Casi Setianingsih. 2018. *Sentiment Analysis on User Satisfaction Level of Mobile Data Services Using Support Vector Machine (SVM) Algorithm. The 2018 IEEE International Conference on Internet of Things and Intelligence System (IoTais)*.
- Gupta, V., Lehal G. S. 2009. *A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence*, vol. 1, no. 1, pp. 60–76, 2009.
- Han, J., Kamber, M. & Pei, J., 2012. *Data mining: concepts and techniques*, Amsterdam: Elsevier/Morgan Kaufmann.
- Indriati, Ridok A., 2016. *Sentiment Analysis For Review Mobile Application Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). Journal of Environmental Engineering & Sustainable Technology Vol. 03 No.01, July 2016, Pages 23-32*.
- Khotimah D., A., K., Riyanarto Sarno. 2018. *Sentimen Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis. 2018. 6th International Conference on Information and Communication Technology (ICoICT)*.
- Kotler, Philip & Gary Armstrong. 2008. *Prinsip-prinsip Pemasaran Jilid I edisi kedua belas. Terjemahan Bob Sabran*. Jakarta: Erlangga.

- Kotler, Philip & Kevin L. Keller. 2009. *Manajemen pemasaran jilid 1, edisi Ketiga belas, Terjemahan Bob Sabran*. Jakarta: Erlangga.
- Liao, S. H., Chu P. H., and Hsiao P. Y., 2012. *Data mining techniques and applications—a decade review from 2000 to 2011 Expert Systems with Applications*, vol. 39, no. 12, pp. 11 303–11 311, 2012.
- Lidya, S. K., Sitompul, O. S. & Efendi, S., 2015. *Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN)*. Seminar Nasional Teknologi Informasi dan Komunikasi.
- Liu, Bing. 2010. *Sentiment Analysis and Subjectivity, Handbook of Natural Language Processing 2: 627-666*.
- Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher.
- Luqyana, W. A., Cholissodin, I., Perdana, R. S., *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*. 2018. JPTIHK Vol 2 No 11 (2018).
- Mahtab S., A., Nazmul Islam, Md Mahfuzur Rahaman. 2018. *Sentiment Analysis on Bangladesh Cricket with Support Vector Machine. International Conference on Bangla Speech and Language Processing(ICBSLP), 21-22 September, 2018*.
- Moore, H.F., 1988. *Hubungan masyarakat: prinsip, kasus dan masalah (satu & dua)*, Bandung: Remadja Karya CV.
- Mustafa, Atika, Ali Akbar, dan Ahmer Sultan. 2009. *Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization. International Journal of Multimedia and Ubiquitous Engineering Vol.4, No.2, April*.
- Naz S., Aditi Sharan, Nidhi Malik. *Sentiment Classification on Twitter Data Using Support Vector Machine*. 2018. IEEE/WIC/ACM International Conference on Web Intelligence (WI) 2018.
- Santosa B., 2010. *Tutorial Support Vector Machine*. Teknik Industri ITS.

Sawakoshi Y., Makoto Okada, Kiyota Hashimoto. 2015. *An Investigation of Effectiveness of “Opinion” and “Fact” sentences for Sentiment Analysis of Customer Reviews*. 2015 *International Conference on Computer Application Technologies*.

Shandra, E. N., Setiawan B. D., Sari Y. A., 2019. *Klasifikasi Pola Sidik Bibir untuk Menentukan Jenis Kelamin Manusia dengan Metode Gray Level Co-Occurrence Matrix dan Support Vector Machine*. JPTIIK Vol. 3 No.3 (2019).

Soenarjo, Djoenasih. 1997. *Opini Publik*. Yogyakarta : Liberty

Talib, R., Hanif, M. K., Ayesha, S., Fatima F., 2016. *Text Mining: Techniques, Applications and Issues*. (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 7 No. 11, 2016.

Vijayakumar, S. & Wu, S., 1999. *Sequential Support Vector Classifiers and Regression*. *International Conference on SoftComputing*, pp. 610-619.

Vijayarani, S., Ilamathi, J., & Nithya., 2015. *Preprocessing Techniques for Text Mining - An Overview*. *International Journal of Computer Science & Communication Networks*, Vol 5(1), 7-16.

W. He. 2013. *Examining students online interaction in a live video streaming environment using data mining and text mining*,” *Computers in Human Behavior*, vol. 29, no. 1, pp. 90-102, 2013.

Zheng W., Ye Q., 2009. *Sentiment Classification of Chinese Traveler Reviews by Support Vector Machine Algorithm*. 2009 *Third International Symposium on Intelligent Information Technology Application*.

# LAMPIRAN

## LAMPIRAN 1

### Hasil Perhitungan Rating pada 5 Hotel di Kota Surabaya dan Rating TripAdvisor

#### 1) Grand Darmo Suite Surabaya

Rating TripAdvisor :

**4,5** Luar biasa  
767 ulasan

No. 3 dari 189 hotel di Surabaya

●●●●○ Lokasi  
●●●●● Kebersihan  
●●●●○ Layanan  
●●●●○ Nilai

Nilai Rating Secara Umum :

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 79 = 3,95 = 4 \end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{49}{62} \times 100 = 79\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{13}{62} \times 100 = 21\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 79 = 3,95 \end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{39}{44} \times 100 = 89\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{5}{44} \times 100 = 11\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 89 = 4,45 \end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{89}{117} \times 100 = 76\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{28}{117} \times 100 = 24\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 76 = 3,8 \end{aligned}$$

## 2) Harris Hotel Gubeng Surabaya

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 93 = 4,65\end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{49}{53} \times 100 = 92\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{4}{53} \times 100 = 8\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 92 = 4,6\end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{70}{73} \times 100 = 96\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{3}{73} \times 100 = 4\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 96 = 4,8\end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{133}{144} \times 100 = 92\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{11}{144} \times 100 = 8\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 92 = 4,6\end{aligned}$$

### 3) POP! Hotel Gubeng Surabaya

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 92 = 4,6 \end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{51}{53} \times 100 = 96\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{2}{53} \times 100 = 4\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 96 = 4,8 \end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{64}{73} \times 100 = 88\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{9}{73} \times 100 = 12\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 88 = 4,4 \end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{133}{144} \times 100 = 92\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{11}{144} \times 100 = 8\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 92 = 4,6 \end{aligned}$$

#### 4) Primebiz Hotel Surabaya

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 82 = 4,1\end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{51}{53} \times 100 = 96\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{2}{53} \times 100 = 4\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 96 = 4,8\end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{64}{73} \times 100 = 88\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{9}{73} \times 100 = 12\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 88 = 4,4\end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{88}{107} \times 100 = 82\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{19}{107} \times 100 = 18\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 82 = 4,1\end{aligned}$$

## 5) Swiss-Bellin Tunjungan Surabaya

Rating TripAdvisor :

**4,5** Luar biasa  
 736 ulasan

No. 5 dari 189 hotel di Surabaya

 Lokasi  
 Kebersihan  
 Layanan  
 Nilai

Nilai Rating Secara Umum :

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 84 = 4,2\end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{68}{77} \times 100 = 88\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{9}{77} \times 100 = 12\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 88 = 4,4\end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{39}{43} \times 100 = 91\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{4}{43} \times 100 = 9\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 91 = 4,5\end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{120}{151} \times 100 = 80\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{31}{151} \times 100 = 20\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 80 = 4\end{aligned}$$

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN 2

### Hasil Perhitungan Rating pada 5 Hotel di Kota Malang dan Rating TripAdvisor

1) The 101 OJ Hotel Malang

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 98 = 4,9 \end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{21}{21} \times 100 = 100\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{0}{21} \times 100 = 0\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 88 = 4,4 \end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{55}{55} \times 100 = 100\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{0}{55} \times 100 = 0\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 100 = 5 \end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{114}{117} \times 100 = 97\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{3}{117} \times 100 = 3\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 97 = 4,85 \end{aligned}$$

## 2) Harris Hotel Malang

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 80 = 4 \end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{50}{77} \times 100 = 65\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{27}{77} \times 100 = 35\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 65 = 3,25 \end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{77}{92} \times 100 = 84\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{15}{92} \times 100 = 16\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 84 = 4,2 \end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{184}{221} \times 100 = 83\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{37}{221} \times 100 = 17\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 83 = 4,15 \end{aligned}$$

### 3) Kartanegara Premium Guest House Malang

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 87 = 4,35 \end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{45}{47} \times 100 = 96\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{2}{47} \times 100 = 4\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 96 = 4,8 \end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{49}{52} \times 100 = 94\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{3}{52} \times 100 = 6\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 94 = 4,7 \end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{63}{81} \times 100 = 78\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{18}{81} \times 100 = 22\%$$

$$\begin{aligned} \text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 78 = 3,9 \end{aligned}$$

#### 4) Hotel Santika Premiere Malang

Rating TripAdvisor :



Nilai Rating Secara Umum :

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 76 = 3,8\end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{30}{40} \times 100 = 75\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{10}{40} \times 100 = 25\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 75 = 3,75\end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{28}{31} \times 100 = 90\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{3}{31} \times 100 = 10\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 90 = 4,5\end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{102}{140} \times 100 = 73\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{38}{140} \times 100 = 27\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 73 = 3,65\end{aligned}$$

## 5) Hotel Tugu Malang

Rating TripAdvisor :

**4,5** Luar biasa  
674 ulasan

No. 1 dari 77 hotel di Malang

●●●●● Lokasi  
●●●●● Kebersihan  
●●●●● Layanan  
●●●●● Nilai

Nilai Rating Secara Umum :

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 72 = 3,6\end{aligned}$$

Aspek lokasi :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{22}{24} \times 100 = 92\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{2}{24} \times 100 = 8\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 92 = 4,6\end{aligned}$$

Aspek Kebersihan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{13}{13} \times 100 = 100\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{0}{13} \times 100 = 0\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 100 = 5\end{aligned}$$

Aspek layanan :

$$\text{Positif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{44}{73} \times 100 = 60\%$$

$$\text{Negatif} = \frac{\text{Banyak data}}{\text{total data}} \times 100 = \frac{29}{73} \times 100 = 40\%$$

$$\begin{aligned}\text{Rating} &= \frac{\text{Nilai Max Rating}}{100} = \frac{5}{100} \times \text{Presentase ulasan positif} \\ &= 0,05 \times 60 = 3\end{aligned}$$

*(Halaman ini sengaja dikosongkan)*

### LAMPIRAN 3

#### Hasil Perhitungan Evaluasi pada 5 Hotel di Kota Surabaya

##### 1) Grand Darmo Suite Surabaya

Nilai Confusion Matrix :

TP	173
TN	24
FP	4
FN	22

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FP+FN}$$
$$\text{Accuracy} = \frac{173+24}{24+173+4+22} = 0,88 = 88\%$$

$$\text{Precision} = \frac{TP}{TP+FP}$$
$$\text{Precision} = \frac{173}{173+4} = 0,98 = 98\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$
$$\text{Recall} = \frac{173}{173+22} = 0,89 = 89\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$
$$\text{F1 Score} = 2 \times \frac{0,98 \times 0,89}{0,98 + 0,89} = 0,93 = 93\%$$

##### 2) Harris Hotel Gubeng Surabaya

Nilai Confusion Matrix :

TP	240
TN	13
FP	12
FN	5

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FP+FN}$$
$$\text{Accuracy} = \frac{240+13}{13+240+12+5} = 0,94 = 94\%$$

$$\text{Precision} = \frac{TP}{TP+FP}$$
$$\text{Precision} = \frac{240}{240+12} = 0,95 = 95\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$
$$\text{Recall} = \frac{240}{240+5} = 0,98 = 98\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$

$$\text{F1 Score} = 2 \times \frac{0,95 \times 0,98}{0,95 + 0,98} = 0,97 = 97\%$$

### 3) POP! Hotel Gubeng Surabaya

Nilai Confusion Matrix :

TP	242
TN	16
FP	10
FN	6

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FP+FN}$$

$$\text{Accuracy} = \frac{242+16}{16+242+10+6} = 0,94 = 94\%$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Precision} = \frac{242}{242+10} = 0,96 = 96\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{242}{242+6} = 0,98 = 98\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$

$$\text{F1 Score} = 2 \times \frac{0,96 \times 0,98}{0,96 + 0,98} = 0,97 = 97\%$$

### 4) Primebiz Hotel Surabaya

Nilai Confusion Matrix :

TP	163
TN	27
FP	9
FN	11

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FP+FN}$$

$$\text{Accuracy} = \frac{163+27}{27+163+9+11} = 0,90 = 90\%$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Precision} = \frac{163}{163+9} = 0,95 = 95\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{163}{163+11} = 0,94 = 94\%$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times recall}{Precision + recall}$$

$$F1 \text{ Score} = 2 \times \frac{0,95 \times 0,94}{0,95 + 0,94} = 0,94 = 94\%$$

5) Swiss-Bellin Tunjungan Surabaya

Nilai Confusion Matrix :

TP	219
TN	21
FP	8
FN	23

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN}$$

$$Accuracy = \frac{219+21}{21+219+8+23} = 0,89 = 89\%$$

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{219}{219+8} = 0,96 = 96\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{219}{219+23} = 0,90 = 90\%$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times recall}{Precision + recall}$$

$$F1 \text{ Score} = 2 \times \frac{0,96 \times 0,90}{0,96 + 0,90} = 0,93 = 93\%$$

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN 4

### Hasil Perhitungan Evaluasi pada 5 Hotel di Kota Malang

1) The 101 OJ Hotel Malang

Nilai Confusion Matrix :

TP	189
TN	3
FP	1
FN	0

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN}$$

$$Accuracy = \frac{189+3}{3+189+1+0} = 0,99 = 99\%$$

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{189}{189+1} = 0,99 = 99\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{189}{189+0} = 1 = 100\%$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times recall}{Precision + recall}$$

$$F1 \text{ Score} = 2 \times \frac{0,99 \times 1}{0,99 + 1} = 1 = 100\%$$

2) Harris Hotel Malang

Nilai Confusion Matrix :

TP	280
TN	59
FP	31
FN	20

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN}$$

$$Accuracy = \frac{280+59}{59+280+31+20} = 0,87 = 87\%$$

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{280}{280+31} = 0,90 = 90\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{280}{280+20} = 0,93 = 93\%$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times recall}{Precision + recall}$$

$$F1 \text{ Score} = 2 \times \frac{0,90 \times 0,93}{0,90 + 0,93} = 0,92 = 92\%$$

### 3) Kartanegara Premium Guest House Malang

Nilai Confusion Matrix :

TP	146
TN	18
FP	12
FN	4

$$Accuracy = \frac{TP + TN}{TN + TP + FP + FN}$$

$$Accuracy = \frac{146 + 18}{18 + 146 + 12 + 4} = 0,91 = 91\%$$

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{146}{146 + 12} = 0,92 = 92\%$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{146}{146 + 4} = 0,97 = 97\%$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times recall}{Precision + recall}$$

$$F1 \text{ Score} = 2 \times \frac{0,92 \times 0,97}{0,92 + 0,97} = 0,95 = 95\%$$

### 4) Hotel Santika Premiere Malang

Nilai Confusion Matrix :

TP	159
TN	10
FP	2
FN	40

$$Accuracy = \frac{TP + TN}{TN + TP + FP + FN}$$

$$Accuracy = \frac{159 + 10}{10 + 159 + 2 + 40} = 0,80 = 80\%$$

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{159}{159 + 2} = 0,99 = 99\%$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{159}{159 + 40} = 0,80 = 80\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$

$$\text{F1 Score} = 2 \times \frac{0,99 \times 0,80}{0,99 + 0,80} = 0,88 = 88\%$$

5) Hotel Tugu Malang

Nilai Confusion Matrix :

TP	78
TN	13
FP	1
FN	18

$$\text{Accuracy} = \frac{TP + TN}{TN + TP + FP + FN}$$

$$\text{Accuracy} = \frac{78 + 13}{13 + 78 + 1 + 18} = 0,83 = 83\%$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{78}{78 + 1} = 0,99 = 99\%$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Recall} = \frac{78}{78 + 18} = 0,81 = 81\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$

$$\text{F1 Score} = 2 \times \frac{0,99 \times 0,81}{0,99 + 0,81} = 0,89 = 89\%$$

BIODATA  
PENULIS

## BIODATA PENULIS

Nama : Siti Azza Amira  
Tempat Tgl. Lahir : Banjarmasin, 21 Juni 1995  
Jenis Kelamin : Perempuan  
Email : [azzaamira95@gmail.com](mailto:azzaamira95@gmail.com)  
Handphone : 08970443661

### Riwayat Pendidikan:

- SD : SD Kartika VI-6 Banjarmasin
- SMP : SMPN 9 Banjarmasin
- SMA : SMKN 5 Malang
- S1 : Teknik Informatika Universitas Brawijaya

