

**Analisis Sentimen Berbasis Aspek pada Objek Wisata  
Pantai di Garut Menggunakan Support Vektor Machine  
dan Long-Short Term Memory**

**Proposal Tugas Akhir**

**1305210065**

**Ikhwan Wahyudin**



**Telkom  
University**

**PROGRAM STUDI SARJANA SAINS DATA  
FAKULTAS INFORMATIKA  
UNIVERSITAS TELKOM  
BANDUNG**

**2024**

# **Lembar Pengesahan**

**Analisis Sentimen Berbasis Aspek pada Objek Wisata Pantai di Garut  
Menggunakan Support Vektor Machine dan Long-Short Term Memory**

***Aspect-Based Sentiment Analysis on Beach Tourism Objects in Garut Using  
Support Vector Machine and Long-Short Term Memory***

**NIM:1305210065**

**Ikhwan wahyudin**

Proposal ini diajukan sebagai usulan pembuatan tugas akhir pada

Program Studi Sarjana Sains Data

Fakultas Informatika

Universitas Telkom

Bandung, 19/05/2024

Menyetujui

Calon Pembimbing I,

Calon Pembimbing II,

Dr. Ir. Moch. Arif Bijaksana, M.Tech.

NIP: 21650006

Dr. Kemas Muslim Lhaksana, S.T., M.ISD.

NIP: 13820075

## Daftar Isi

Lembar Pengesahan.....	2
Daftar Isi .....	3
Daftar Gambar .....	5
Daftar Tabel .....	6
Abstrak.....	7
BAB 1 Pendahuluan.....	8
1.1. Latar Belakang .....	8
1.2. Perumusan Masalah .....	9
1.3. Tujuan .....	9
1.4. Rencana Kegiatan .....	10
1.5. Jadwal Kegiatan .....	11
BAB 2 Kajian Pustaka .....	12
2.1. Penelitian Terkait .....	12
2.2. Analisis Sentimen Berbasis Aspek.....	12
2.3. Google Map .....	13
2.4. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	13
2.5. <i>Support Vektor Machine</i> .....	13
2.6. <i>Word Embedding</i> .....	14
2.7. <i>Long Short-Term Memory</i> .....	14
BAB 3 Perancangan Sistem atau Alur Pemodelan.....	16
3.1 Eksplorasi Data .....	16
3.2 Perancangan Sistem .....	17
3.3 Pengumpulan Data.....	17
3.4 Pelabelan Data.....	18
3.5 <i>Case Folding</i> .....	18
3.6 Tokenisasi.....	18
3.7 <i>Punctuation Removal</i> .....	19
3.8 <i>Stopword Removal</i> .....	19
3.9 <i>Stemming</i> .....	20
3.10 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	20
3.11 <i>Word Embedding</i> .....	21

<b>3.12</b>	<b>Data Split .....</b>	<b>21</b>
<b>3.13</b>	<b>Pembelajaran Model.....</b>	<b>22</b>
<b>3.14</b>	<b>Evaluasi.....</b>	<b>22</b>
<b>Daftar Pustaka.....</b>		<b>23</b>

## **Daftar Gambar**

<b>Gambar 2.5.1. Representasi SVM.....</b>	<b>13</b>
<b>Gambar 2.7.1. Jaringan Sel LSTM.....</b>	<b>15</b>
<b>Gambar 3.1.1. Visualisasi WordCloud .....</b>	<b>16</b>
<b>Gambar 3.1.2. Visualisasi BarPlot .....</b>	<b>16</b>
<b>Gambar 3.2. Perancangan Sistem.....</b>	<b>17</b>

## Daftar Tabel

Tabel 1.5. 1. Jadwal Kegiatan.....	11
Tabel 3.5. 1. Contoh Case Folding .....	18
Tabel 3.6. 1. Contoh Tokenisasi.....	18
Tabel 3.7. 1. Contoh Punctuation Removal.....	19
Tabel 3.8. 1. Contoh Stopword Removal.....	19
Tabel 3.9. 1. Contoh Stemming .....	20
Tabel 3.10. 1. Contoh TF-IDF Pada Scikit-Learn .....	20
Tabel 3.11. 1. Contoh Word Embedding .....	21
Tabel 3.11. 2. Contoh Word Embedding dan Padding Sequences .....	21

## Abstrak

Indonesia terkenal akan sumber daya alamnya yang melimpah, salah satunya kota Garut yang terkenal akan potensi wisatanya baik di pegunungan ataupun di daerah pesisir pantai. Dengan berkembangnya teknologi dan akses mudah terhadap media sosial pada saat ini, wisatawan semakin aktif memberikan pengalaman yang telah mereka rasakan mengenai objek wisata khususnya pantai yang ada di Garut. Banyaknya data ulasan dan penilaian terhadap berbagai aspek dari pengguna dapat menjadi suatu sumber informasi yang penting bagi pihak pengelola objek wisata agar bisa meningkatkan kepuasan dan pengalaman wisatawan. Metode Analisis sentimen berbasis aspek dianggap cocok dan memungkinkan untuk memahami reaksi dan pandangan wisatawan atau pengunjung terhadap berbagai aspek yang terkait dengan objek wisata pantai di Garut. Pada penelitian analisis sentimen berbasis aspek ini menggunakan dua algoritma pembelajaran mesin dengan metode *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM) dengan menambahkan TF-IDF untuk algoritma SVM dan *word embedding* untuk LSTM. Hasilnya pada penelitian ini akan mendapatkan performansi dari kedua model terhadap objek wisata pantai di Garut, dan akan dilakukan perbandingan model mana yang lebih baik dalam analisis sentimen berbasis aspek ini.

Kata Kunci : analisis sentimen berbasis aspek, *Support Vector Machine* , *Long Short-Term Memory*.

# **BAB 1 Pendahuluan**

## **1.1. Latar Belakang**

Indonesia terkenal akan sumber daya alamnya yang melimpah dan menakjubkan, yang bisa menjadi peluang besar untuk industri pariwisata. Salah satunya kota Garut yang terkenal akan potensi wisatanya baik di pegunungan ataupun di daerah pesisir pantai[1]. Dengan berkembangnya teknologi dan akses mudah terhadap media sosial pada saat ini, wisatawan ataupun masyarakat semakin aktif memberikan pengalaman yang telah mereka rasakan mengenai berbagai objek wisata yang telah dikunjungi. Salah satu platform yang digunakan untuk memberikan pendapat atau ulasan mengenai suatu tempat yaitu Google Map. Pada Google Map ini terdapat kolom komentar atau ulasan yang dapat digunakan pengguna untuk memberikan opini dan penilaiannya terhadap suatu tempat atau lokasi. Google Map juga dapat menampilkan informasi ulasan dari satu pengguna ke pengguna lainnya agar bisa menjadi pertimbangan dan membantu pengambilan keputusan. Menurut survei yang dilakukan oleh Sammy Paget pada situs halaman webnya menjelaskan bahwa 81% pengguna lebih sering menemukan, menggunakan dan membaca ulasan pada platform Google dibandingkan dengan yang lainnya pada tahun 2024 [2].

Didalam ulasan pada platform tersebut terdapat berbagai macam reaksi positif maupun negatif terhadap objek atau tempat yang diulas salah satunya objek wisata pantai yang terdapat di kabupaten Garut provinsi Jawa Barat. Banyaknya data ulasan dan penilaian terhadap berbagai aspek dari pengguna dapat menjadi suatu sumber informasi yang penting bagi pihak pengelola objek wisata agar bisa meningkatkan kepuasan dan pengalaman wisatawan. Maka dari itu, perlu dilakukan analisis terhadap objek wisata pantai di kabupaten Garut. Metode analisis sentimen berbasis aspek dianggap cocok dan memungkinkan untuk memahami reaksi dan pandangan wisatawan atau pengunjung terhadap berbagai aspek yang terkait dengan objek wisata pantai di Garut.

Analisis sentimen berbasis aspek merupakan proses menganalisis dan mengidentifikasi elemen sentimen pada item teks, diantaranya ada elemen sentimen tunggal atau beberapa elemen dengan setiap elemen memiliki ketergantungan satu



sama lain[3]. Pada penelitian analisis sentimen berbasis aspek ini menggunakan dua algoritma pembelajaran mesin dengan metode *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM) dengan menambahkan TF-IDF untuk algoritma SVM dan *word embedding* untuk LSTM.

## **1.2. Perumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini sebagai berikut :

### **1. Persebaran Aspek dan Polaritas**

Persebaran aspek merujuk pada distribusi dari berbagai kategori yang ditemukan dalam kalimat ulasan, sedangkan polaritas mengacu pada sentimen yang terdapat pada setiap aspek. Dengan asumsi aspek dan polaritas sentimen dapat diidentifikasi dan diekstraksi secara akurat dari kalimat ulasan menggunakan teknik pembelajaran mesin. Analisis persebaran aspek dan polaritas ini akan dibatasi hanya pada ulasan objek wisata pantai yang terdapat di wilayah Garut yang diperoleh dari google map.

### **2. Implementasi metode SVM dan LSTM**

SVM dan LSTM merupakan dua metode pembelajaran mesin. Dengan asumsi kedua metode ini dapat memprediksi sentimen pada setiap aspek dengan akurasi yang cukup tinggi. Ruang lingkup pada penelitian ini melibatkan pelatihan model SVM dan LSTM menggunakan data ulasan objek wisata pantai di Garut berdasarkan polaritas pada setiap aspeknya.

### **3. Perbandingan performa dan akurasi metode**

Performansi dan akurasi merupakan kemampuan suatu model dalam memprediksi secara benar yang dapat diukur dengan matrix evaluasi. Dengan asumsi performa dari kedua model dapat diukur dan dibandingkan secara akurat menggunakan hasil dari matrix evaluasi yang digunakan. Dengan ruang lingkup penelitian ini akan membandingkan hasil pelatihan dan prediksi kedua model .

## **1.3. Tujuan**

Penelitian pada penulisan proposal tugas akhir ini bertujuan sebagai berikut:

1. Mengetahui dan Mengidentifikasi Persebaran Aspek dan Polaritas.  
Penelitian ini bertujuan untuk mengetahui bagaimana persebaran polaritas pada tiap aspek untuk objek wisata pantai di Garut.
2. Mengimplementasi dan mengevaluasi model algoritma SVM dan LSTM.  
Penelitian ini juga bertujuan untuk menerapkan dua metode SVM dan LSTM dengan cara analisis sentimen berbasis aspek terhadap objek wisata pantai Garut dan mengevaluasi kinerja model.
3. Membandingkan Performansi dan Akurasi Model.  
Penelitian ini juga bertujuan untuk membandingkan performansi dan akurasi pada kedua model yang digunakan untuk melihat dan mengetahui model mana yang lebih bagus dalam analisis sentimen berbasis aspek pada ulasan objek wisata pantai di Garut.

#### **1.4. Rencana Kegiatan**

Rencana kegiatan yang dilakukan pada penelitian dalam penulisan proposal ini, yaitu:

1. Studi Literatur  
Kegiatan yang dilakukan dengan mencari dan mengumpulkan informasi yang sesuai dengan studi kasus yang akan diteliti. Studi literatur yang digunakan berupa jurnal terkait analisis sentimen berbasis aspek, SVM, penambangan teks, LSTM, dan studi kasus yang akan diteliti.
2. Pengumpulan Data  
Pada tahap dilakukan pengumpulan data dengan melakukan scrapping ulasan pada Google Map menggunakan website apify.com
3. Identifikasi Masalah  
Berdasarkan data yang telah dikumpulkan, maka akan diidentifikasi metode algoritma dan analisis sentimen seperti apa yang sekiranya cocok dan efektif untuk mencapai tujuan penelitian.
4. Pelabelan Data  
Setelah dilakukannya proses pengambilan data, maka langkah selanjutnya melakukan pelabelan data secara manual, dengan beberapa label sentimen pada setiap aspek.

## 5. Perancangan Sistem

Menyusun kerangka kerja penelitian dan merancang sistem yang akan digunakan. Pada tahap ini, juga dilakukan analisis terhadap kebutuhan sistem dan pembuatan alur kerja sistem yang direpresentasikan dalam bentuk *flowchart*.

## 6. Implementasi Sistem

Melakukan implementasi metode analisis sentimen berbasis aspek dengan algoritma SVM dan LSTM terhadap dataset yang telah dipersiapkan.

## 7. Analisis Hasil Implementasi

Pada tahap ini dilakukan evaluasi dan analisis untuk mengetahui kualitas hasil implementasi metode yang digunakan.

## 8. Penulisan Laporan.

Pada tahap ini dilakukan penyusunan laporan yang berisi dokumentasi dari keseluruhan hasil proses kegiatan penelitian yang dilakukan dari awal sampe akhir termasuk studi literatur, pengumpulan dan pengolahan data, perancangan sistem, implementasi, evaluasi, dan analisis hasil implementasi, serta kesimpulan yang diperoleh.

### 1.5. Jadwal Kegiatan

Berikut jadwal kegiatan penelitian direncanakan dalam 6 bulan yang dapat dilihat pada table 1.5.1.

Tabel 1.5. 1.Jadwal Kegiatan

Kegiatan	Bulan					
	1	2	3	4	5	6
Studi Literatur						
Pengumpulan Data						
Identifikasi Masalah						
Pelabelan Data						
Perancangan Sistem						
Implementasi Sistem						
Analisis Hasil Implementasi						
Penulisan Laporan						

## BAB 2 Kajian Pustaka

### 2.1. Penelitian Terkait

Penelitian analisis sentimen berbasis aspek telah dilakukan oleh Clarisa Hasya Yutika dengan metode yang digunakan *naïve bayes* pada dataset ulasan *Famale Daily* dengan melakukan berbagai eksperimen skenario diantaranya *Full Preprocessing*, *No Stopword Removal*, *no stemming*, *no stemming* dan *no stopword removal* dengan hasil pengujian tertinggi pada *no stopword removal* dengan akurasi 75,99 % dan pada aspek harga dengan hasil pengujian akurasi 79,92% [4].

Penelitian [5] melakukan analisis sentimen berbasis aspek dengan metode SVM dengan menambahkan metode *One-Against-All* dan *Sequensial Learning* pada dataset rumah makan dengan hasil pengujian berupa klasifikasi aspek dan sentimen dengan rata rata *F- Measure* pada aspek 73% dan pada klasifikasi polaritas 92%.

Penelitian [6] melakukan analisis sentimen berbasis aspek dengan menggunakan metode LSTM dengan 50 hingga 100 *epoch* serta tingka pembelajaran 0,01 hingga 0,02 dan proses *word embedding* GloVe dengan 50 dimensi pada dataset aduan mahasiswa UMSIDA dimasa pandemi dengan hasil pengujian akurasi 81%.

Pada penelitian yang dilakukan oleh Windi Astriningsih melakukan analisis sentimen dengan pendekatan multi aspek menggunakan LSTM pada dataset ulasan hotel dengan melakukan eksperimen penelitian pada dataset dengan pengujian untuk prediksi polaritas sentimen, prediksi satu aspek, kombinasi dua aspek dan kombinasi tiga aspek dengan hasil pengujian didapatkan dengan satu aspek F1-Score sebesar 85,7% dan untuk kombinasi dua dan tiga aspek memiliki nilai rendah sekitar 79% serta F1-Score 83% untuk hasil prediksi polaritas sentimen [7].

Pada penelitian dan survei yang dilakukan oleh Ambreen Nazir ,Yuan Rao,Lianwei Wu ,dan Ling Sun dengan berisi dataset, metode yang digunakan *supervised*, *unsupervised*, dan *hybrid* dan *evaluation task* yang beragam. Pada survei ini dapat menjadi ladasan penelitian analisis sentimen berbasis aspek pada berbagai bidang aplikasi, serta tujuan dan pengembangan analisis sentimen berbasis aspek agar menjadi lebih efektif dan efisien [8].

### 2.2. Analisis Sentimen Berbasis Aspek

Analisis sentimen berbasis aspek merupakan suatu proses untuk mengidentifikasi aspek – aspek dan sentimen terkait yang terdapat dalam kalimat. Tugas dari analisis sentimen berbasis aspek ini telah ada sejak 2014 dengan pada dua tahun pengajuan terakhir model yang paling bagus didapatkan oleh *support vector machine* (SVM)[9].Pada analisis sentimen tingkat aspek, setiap kalimat dapat memiliki banyak aspek. Oleh karena itu,tugas dari tingkat aspek ini memberikan polaritas untuk semua aspek yang terdapat pada kalimat[10].

### 2.3. Google Map

Google Map merupakan layanan yang menyediakan informasi terperinci mengenai wilayah geografis dan situs yang berada diseluruh dunia. API Google Map memungkinkan platform perangkat lunak yang berbeda untuk berkomunikasi satu sama lain[11]. Pengguna juga dapat memberikan ulasan mengenai suatu tempat atau situs yang terdapat pada google map. Dengan memanfaatkan API yang telah disediakan oleh google map, pengguna dapat melakukan ekstraksi data pada Google Map akan tetapi akses API ini tidak gratis.

### 2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan proses pembobotan dengan menilai signifikansi istilah tertentu dalam dokumen tertentu. TF-IDF dibangun dari komponen TF dan IDF. TF merupakan pengukuran frekuensi istilah dalam dokumen tertentu sedangkan IDF merupakan perhitungan berapa banyak dokumen dalam korpus termasuk frasa yang ditentukan [12]. Rumus TF-IDF adalah sebagai berikut :

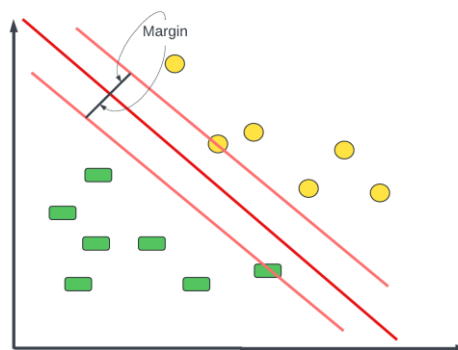
$$TF(t, d) = \frac{\text{Jumlah kemunculan } t \text{ dalam dokumen } d}{\text{Total jumlah kata dalam dokumen } d} \quad (2.4.1)$$

$$IDF(t, D) = \frac{\text{Total jumlah dokumen}}{\text{jumlah dokumen } (D) \text{ yang mengandung kata } (t)} \quad (2.4.2)$$

$$TF - IDF = TF(t, d) * IDF(t, D) \quad (2.4.3)$$

### 2.5. Support Vektor Machine

*Support Vektor Machine* merupakan metode pembelajaran mesin diawasi ( *supervised learning*) yang mampu belajar dari data dan membuat keputusan. Prinsip dasar SVM telah diperkenalkan sejak 1960-an oleh Vapnik dan Chervonenkis[13]. SVM ini bekerja dengan mencari hyperplane optimum, yaitu hyperplane yang memaksimalkan margin untuk memisahkan data yang memiliki kelas yang berbeda. Representasi SVM dinyatakan pada gambar 2.5.1.



Gambar 2.5.1. Representasi SVM

$$y_i(w * X_i + B) \geq 1, i \geq 1 ; \forall_i [14] \quad (2.5.1)$$

Pada persamaan yang diberikan diatas ,  $y_i$  merupakan data ke-I yang nilainya label kelas tergantung pada studi kasus yang dipakai, dan nilai  $X_i$  merupakan vektor

fitur input dari data ke-i. Nilai  $w$  merupakan vektor bobot yang digunakan untuk menentukan posisi hyperplane, sedangkan  $b$  merupakan bias yang memungkinkan hyperplane untuk tidak melewati titik asal dan memiliki posisi optimal dalam memisahkan kelas kelas dalam margin maksimal. Berikut representasi

## 2.6. Word Embedding

*Word embedding* merupakan sebuah teknik dalam pemrosesan bahasa alami dimana setiap kata dihubungkan kedalam vektor yang mengandung angka yang memiliki panjang dan nilai khusus untuk setiap vektor dalam kamus.[6]

## 2.7. Long Short-Term Memory

*Long Short-Term Memory* (LSTM) merupakan salah satu jenis dari algoritma RNN. RNN merupakan sejenis *neural network* tiruan dengan dirancang mengenali pola data berurutan. LSTM memiliki kelebihan dapat melakukan prediksi dengan mempelajari data sekuensial. LSTM dapat mengingat data pada jangka waktu yang cukup lama [15]. Sel LSTM mempunyai input dan output seperti halnya sel RNN biasa, tetapi sel LSTM mempunyai sistem dengan unit pengendalian (*gating*) untuk mengontrol aliran informasi. Gerbang lupa mengontrol berat status sel loop mandiri untuk memperbarui sel memori, berdasarkan input waktu saat ini  $x(t)$  dan output dari momen sebelumnya  $h(t-1)$ , maka didapatkan rumus sebagai berikut [16]:

$$F_i^{(t)} = \left( b_i^F + \sum_j U_{i,j}^f X_j^{(t)} \sum_j W_{i,j}^f h_j^{(t-1)} \right) \quad (2.7.1)$$

$b^f, u^f w^f$  menunjukkan bias masing masing, bobot input dan berulang dari gerbang lupa atau *forget gate* dengan  $\sigma$  fungsi sigmoid yang menetapkan *forget gate* ke nilai antara 0 dan 1. *Input gate* mengontrol informasi yang dimasukan kedalam sel dengan rumus sebagai berikut :

Jaringan LSTM kemudian menggunakan input sebagai berikut :

$$I_i^{(t)} = \sigma \left( b_i^I + \sum_j U_{i,j}^I X_j^{(t)} \sum_j W_{i,j}^I h_j^{(t-1)} \right) \quad (2.7.2)$$

$b^I, u^I w^I$  menunjukkan bias masing masing, bobot input dan berulang dari gerbang input. Kemudian kondisi internal sel LSTM akan diperbaharui sebagai berikut :

$$S_i^{(t)} = F_i^{(t)} S_i^{(t-1)} + I_i^{(t)} \tanh \left( b_i^I + \sum_j U_{i,j}^I X_j^{(t)} \sum_j W_{i,j}^I h_j^{(t-1)} \right) \quad (2.7.3)$$

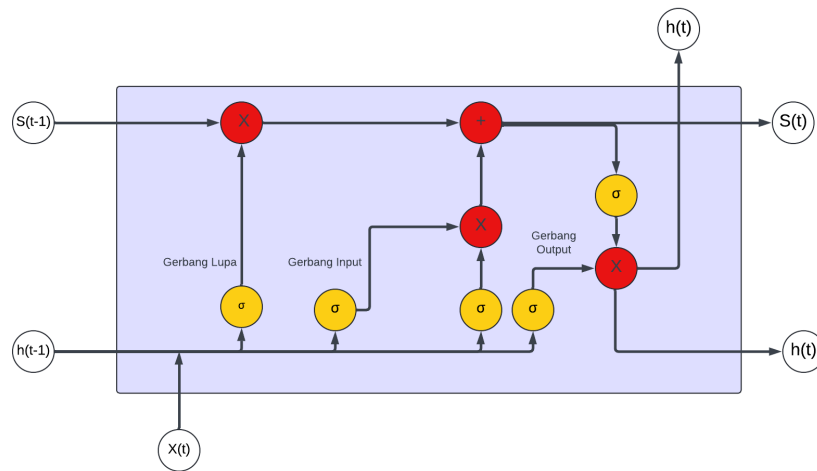
$b, U, W$  menunjukkan bias masing-masing, bobot input dan berulang ke dalam sel LSTM. *Output gate* mengontrol berat keluaran sel sebagai berikut :

$$O_i^{(t)} = \sigma \left( b_i^o + \sum_j U_{i,j}^o X_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)} \right) \quad (2.7.4)$$

$b^o, u^o, w^o$  menunjukkan bias masing masing, bobot input dan berulang berasal dari *output gate* . Pada akhirnya output dari sel LSTM sebagai berikut :

$$h_i^{(t)} = \tanh \left( S_i^{(t)} \cdot O_i^{(t)} \right) \quad (2.7.5)$$

Berikut merupakan representasi struktur sel LSTM :



Gambar 2.7.1. Jaringan sel LSTM [16]

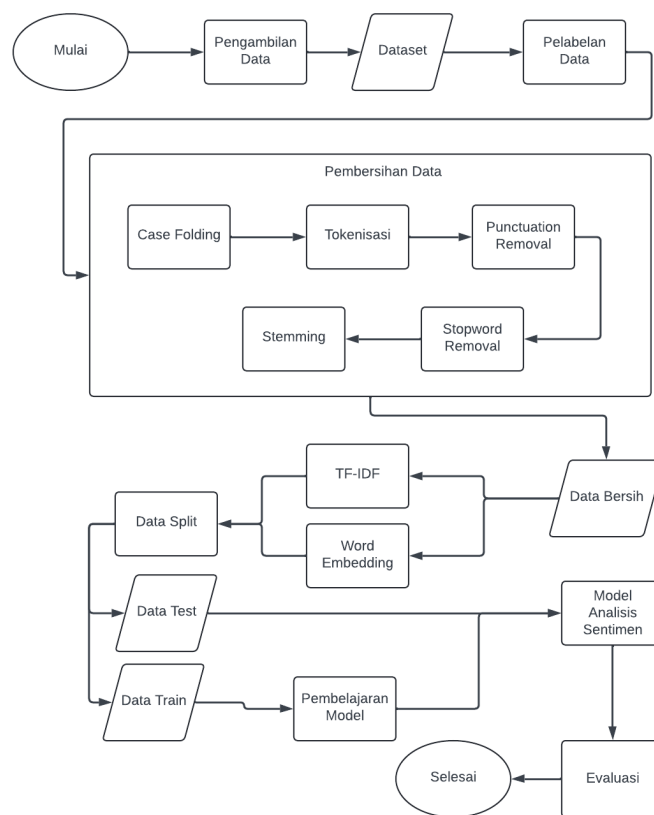
### Gambar 3.1.2. Visualisasi BarPlot



Berdasarkan gambar 3.1.2. visualisasi itu menunjukkan persebaran jumlah ulasan pada setiap pantai. Dengan ulasan terbanyak terdapat pada pantai santolo.

### 3.2 Perancangan Sistem

Pada tahap ini akan dijelaskan mengenai sistem yang akan digunakan dalam penelitian ini. Berikut flowchart dari perancangan sistem terdapat pada Gambar 3.2.1.



Gambar 3.2. 1. Perancangan Sistem

### 3.3 Pengumpulan Data

Pengumpulan data dilakukan dengan cara *crawling* dengan mengambil data ulasan pada Google Map dengan bantuan platform apify.com. Dengan menggunakan kata kunci pantai yang ada di garut selatan pada pencarian Google untuk memperoleh informasi pantai dan setelah itu dilakukan pencarian pada Google Map daerah kabupaten Garut dengan kata kunci pantai yang selanjutnya memilih pantai memiliki review cukup banyak dan sering digunakan wisatawan untuk

berwisata. Platform ini menyediakan google map API yang dapat digunakan secara gratis namun dengan adanya jumlah pembatasan pemakaian untuk pengguna gratis. Google map API ini dapat mengakses dan mengumpulkan data yang dibutuhkan dari platform google map.

### 3.4 Pelabelan Data

Pelabelan data merupakan pemberian label pada data text. Pelabelan ini dilakukan secara manual oleh peneliti sendiri. Pada setiap ulasan akan diberikan label sentimen pada setiap aspek. Untuk label sentimen terdiri dari tiga yaitu positif dengan simbol angka (1) akan melabeli setiap sentimen yang memiliki penilaian positif yang menunjukkan kepuasan atau pengalaman baik terhadap aspek tertentu dari objek wisata, negatif dengan simbol angka (-1) yang akan melabeli setiap ulasan atau sentimen memiliki penilaian negatif yang menunjukkan ketidakpuasan atau pengalaman buruk yang dialami oleh wisatawan, dan netral dengan simbol angka (0) yang akan melabeli setiap ulasan atau sentimen yang memiliki makna netral (tidak positif atau negatif) dan ulasan yang didalamnya tidak terdapat aspek yang ditentukan maka aspek tersebut akan dilabeli 0. Sedangkan untuk aspek akan terbagi menjadi 5 aspek yaitu aspek kebersihan, akses lokasi, fasilitas umum, harga, keindahan alam, dan keamanan. Pemilihan aspek dilakukan dengan melihat ulasan yang sering muncul atau dibahas oleh pengulas.

### 3.5 Case Folding

*Case Folding* merupakan salah satu pre-processing yang bertujuan untuk mengubah semua huruf dalam kalimat atau dokumen menjadi huruf kecil. Hal ini dilakukan untuk menyamakan semua huruf dalam dokumen teks yang ditulis dengan huruf besar atau kecil.

Tabel 3.5. 1. *Contoh Case Folding*

Sebelum	Sesudah
Pantainya bagus dan bersih, Tempat makan kurang, harga tiket murah.	pantainya bagus dan bersih, tempat makan kurang, harga tiket murah.
Pantai yang cukup indah, tiket berkunjung sangat murah, bersih, belum ada penginapan.	pantai yang cukup indah, tiket berkunjung sangat murah, bersih, belum ada penginapan.
Pantainya indah. Penginapan banyak dengan fasilitas lengkap.	pantainya indah. penginapan banyak dengan fasilitas lengkap.

### 3.6 Tokenisasi

Tokenisasi merupakan proses untuk memisahkan kata dalam suatu dokumen teks menjadi unit – unit kecil disebut token.

Tabel 3.6. 1. *Contoh Tokenisasi*

Sebelum	Sesudah
pantainya bagus dan bersih, tempat makan kurang, harga tiket murah.	['pantainya', 'bagus', 'dan', 'bersih', ',', 'tempat', 'makan', 'kurang', ',', 'harga', 'tiket', 'murah', '.']
pantai yang cukup indah, tiket berkunjung sangat murah, bersih, belum ada penginapan.	['pantai', 'yang', 'cukup', 'indah', ',', 'tiket', 'berkunjung', 'sangat', 'murah', ',', 'bersih', ',', 'belum', 'ada', 'penginapan', '.']
pantainya indah. penginapan banyak dengan fasilitas lengkap.	['pantainya', 'indah', '.', 'penginapan', 'banyak', 'dengan', 'fasilitas', 'lengkap', '.']

### 3.7 Punctuation Removal

*Punctuation Removal* merupakan suatu proses menghilangkan tanda baca pada dokumen teks.

Tabel 3.7. 1. Contoh *Punctuation Removal*

Sebelum	Sesudah
['pantainya', 'bagus', 'dan', 'bersih', ',', 'tempat', 'makan', 'kurang', ',', 'harga', 'tiket', 'murah', '.']	['pantainya', 'bagus', 'dan', 'bersih', 'tempat', 'makan', 'kurang', 'harga', 'tiket', 'murah']
['pantai', 'yang', 'cukup', 'indah', ',', 'tiket', 'berkunjung', 'sangat', 'murah', ',', 'bersih', ',', 'belum', 'ada', 'penginapan', '.']	['pantai', 'yang', 'cukup', 'indah', 'tiket', 'berkunjung', 'sangat', 'murah', 'bersih', 'belum', 'ada', 'penginapan']
['pantainya', 'indah', '.', 'penginapan', 'banyak', 'dengan', 'fasilitas', 'lengkap', '.']	['pantainya', 'indah', 'penginapan', 'banyak', 'dengan', 'fasilitas', 'lengkap']

### 3.8 Stopword Removal

*Stopword Removal* merupakan tahap untuk menghapus kata-kata umum yang tidak memiliki informasi penting atau kata-kata yang sering muncul pada dokumen teks dan tidak membawa makna penting untuk analisis teks seperti ‘di’, ‘yang’, ‘dan’. Pada penelitian ini library yang akan digunakan pada python yaitu library dari nltk.corpus untuk bahasa Indonesia.

Tabel 3.8. 1. Contoh *Stopword Removal*

Sebelum	Sesudah
['pantainya', 'bagus', 'dan', 'bersih', 'tempat', 'makan', 'kurang', 'harga', 'tiket', 'murah']	['pantainya', 'bagus', 'bersih', 'makan', 'harga', 'tiket', 'murah']
['pantai', 'yang', 'cukup', 'indah', 'tiket', 'berkunjung', 'sangat', 'murah', 'bersih', 'belum', 'ada', 'penginapan']	['pantai', 'indah', 'tiket', 'berkunjung', 'murah', 'bersih', 'penginapan']

['pantainya', 'indah', 'penginapan', 'banyak', 'dengan', 'fasilitas', 'lengkap']	['pantainya', 'indah', 'penginapan', 'fasilitas', 'lengkap']
--	--

### 3.9 Stemming

Pada tahap ini akan dilakukan proses pengurangan kata-kata turunan atau menghilangkan imbuhan kata pada tiap kata dalam kalimat ulasan. Misalkan pada kata “merasakan”, “dirasakan”, “merasa” akan diubah menjadi “rasa”. *Library* yang digunakan pada penelitian ini yaitu *library python* sastrawi.

Tabel 3.9. 1. Contoh *Stemming*

Sebelum	Sesudah
['pantainya', 'bagus', 'bersih', 'makan', 'harga', 'tiket', 'murah']	['pantai', 'bagus', 'bersih', 'makan', 'harga', 'tiket', 'murah']
['pantai', 'indah', 'tiket', 'berkunjung', 'murah', 'bersih', 'penginapan']	['pantai', 'indah', 'tiket', 'kunjung', 'murah', 'bersih', 'inap']
['pantainya', 'indah', 'penginapan', 'fasilitas', 'lengkap']	['pantai', 'indah', 'inap', 'fasilitas', 'lengkap']

### 3.10 Term Frequency-Inverse Document Frequency (TF-IDF)

Pada Proses ini dilakukan pembobotan kata pada tiap kata yang terdapat pada ulasan objek wisata pantai di garut. Hal ini dilakukan untuk mengukur pentingnya sebuah kata dalam kalimat ulasan terhadap kalimat ulasan lainnya yang nantinya akan dijadikan sebuah input untuk proses pembelajaran mesin pada metode SVM. Berikut contoh pembobotan TF-IDF pada python menggunakan *library* scikit-learn

Tabel 3.10. 1. Contoh TF-IDF Pada Scikit-Learn

Kata	TF-IDF		
	Kalimat 1	Kalimat 2	Kalimat 3
pantai	0.261940	0.286801	0.315444
bagus	0.443503	0	0
bersih	0.337295	0.369308	0
makan	0.443503	0	0
harga	0.443503	0	0
Tiket	0.337295	0.369308	0
Murah	0.337295	0.369308	0
Indah	0	0.369308	0.406192
kunjung	0	0.485596	0
inap	0	0.369308	0.406192
fasilitas	0	0	0.534093
lengkap	0	0	0.534093

### 3.11 Word Embedding

Pada proses ini akan dilakukan *word embedding* merupakan suatu proses dimana setiap kata dalam kosakata dikaitkan dengan vektor yang berisi angka dan memungkinkan model untuk menangkap hubungan semantik antar kata yaitu kata yang memiliki hubungan dekat akan berada dekat satu sama lain. Pada tahap ini juga akan dilakukan *padding sequence* yang merupakan proses untuk memastikan semua urutan input memiliki panjang yang sama sebelum dilakukan pelatihan model pada LSTM. Apabila urutan input tidak sama maka akan dilakukan penambahan nilai khusus agar semua input memiliki panjang yang sama.

Tabel 3.11. 1.Contoh *Word Embedding*

Kata	Vektor Embedding
pantai	[0.1, 0.3, .02]
bagus	[0.5, 0.1, 0.4]
bersih	[0.2, 0.4, 0.3]
makan	[0.3, 0.5, 0.1]
harga	[0.4, 0.3, 0.2]
tiket	[0.1, 0.4, 0.3]
murah	[0.2, 0.5, 0.4]
indah	[0.3, 0.5, 0.2]
kunjung	[0.5, 0.2, 0.4]
inap	[0.4, 0.2, 0.5]
fasilitas	[0.3, 0.4, 0.5]
lengkap	[0.5, 0.3, 0.4]

Tabel 3.11. 2.Contoh *Word Embedding* dan *Padding Sequences*

	Sebelum	<i>Word Embedding</i> dan <i>Padding Sequences</i>
Kalimat 1	pantai bagus bersih makan harga tiket murah	[0.1, 0.3, .02][0.5, 0.1, 0.4][0.2, 0.4, 0.3] [0.3, 0.5, 0.1][0.4, 0.3, 0.2][0.1, 0.4, 0.3] [0.2, 0.5, 0.4]
Kalimat 2	pantai indah tiket kunjung murah bersih inap	[0.1, 0.3, .02][0.3, 0.5, 0.2][0.1, 0.4, 0.3] [0.5, 0.2, 0.4][0.2, 0.5, 0.4] [0.2, 0.4, 0.3] [0.4, 0.2, 0.5]
Kalimat 3	pantai indah inap fasilitas lengkap	[0.1, 0.3, .02][0.3, 0.5, 0.2] [0.4, 0.2, 0.5] [0.3, 0.4, 0.5] [0.5, 0.3, 0.4][0.0, 0.0, 0.0] [0.0, 0.0, 0.0]

### 3.12 Data Split

Pada tahapan ini data yang telah dilakukan proses labeling dan pembersihan kata ,selanjutnya data akan dibagi menjadi dua bagian, 80% dari total data akan digunakan sebagai data latih untuk melatih model Support Vektor Machine dan 20% dari total data akan digunakan sebagai data tes untuk menguji performansi dari model yang digunakan.

### 3.13 Pembelajaran Model

Setelah dilakukannya pembagian pada dataset, 80% dataset yang dijadikan data latih akan digunakan untuk melatih model algoritma SVM dan LSTM untuk analisis sentimen berbasis aspek ini.

### 3.14 Evaluasi

Pada penelitian ini akan dilakukan evaluasi model untuk melihat kinerja dari model yang telah dibangun. Kinerja dari model akan dievaluasi menggunakan confusion matrix dengan berbentuk tabel pada serangkaian data yang diuji. Confusion Matrix ini terdiri dari :

1. *True Positive* (TP), apabila data aktual positif dan diprediksi positif.
2. *True Negatif* (TN), apabila data aktual negatif dan diprediksi negatif.
3. *False Positif* (FP), apabila data aktual positif dan diprediksi negatif.
4. *False Negative* (FN), apabila data aktual negatif dan diprediksi positif.

Berdasarkan nilai nilai tersebut dapat dihitung nilai akurasi, presisi, recall dan F1-scorenya. Rumus perhitungan sebagai berikut :

1. Akurasi

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.14.1)$$

2. Presisi

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (3.14.2)$$

3. Recall

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.14.3)$$

4. F1-Score

$$\text{F1 - Score} = \frac{2 \times \text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \quad (3.14.4)$$

Selain dilakukan evaluasi dengan *confusion matrix*, juga dilakukan evaluasi dengan *hamming loss*. *Hamming loss* merupakan fraksi label yang salah terhadap jumlah total label pada klasifikasi multi label, kerugian hamming dihitung sebagai jarak hamming antara  $y_{\text{prediksi}}$  dan  $y_{\text{benar}}$ . Berikut rumus hamming loss :

$$\text{Hamming Loss} = \frac{1}{N \times L} \sum_{i=1}^N \sum_{j=1}^L I(y_{ij} \neq p_{ij}) \quad (3.14.5)$$

Dengan :

$y_{ij}$  Merupakan nilai sebenarnya dari label  $j$  pada sample  $i$ .

$p_{ij}$  merupakan nilai prediksi dari label  $j$  sampai  $i$ .

$I$  merupakan fungsi indikator yang bernilai 1 jika  $(y_{ij} \neq p_{ij})$  dan 0 jika  $(y_{ij} = p_{ij})$

## Daftar Pustaka

- [1] Alamanda, D. T., Ramdhani, A., Kania, I., Susilawati, W., & Hadi, E. S. (2019). Sentiment analysis using text mining of Indonesia tourism reviews via social media. *Int. J. Humanit. Arts Soc. Sci*, 5(2), 72-82.
- [2] Paget, S. 2024. Local Consumer Review Survey 2024: Trends, Behaviors, and Platforms Explored. [Online] Available at : <https://www.brightlocal.com/research/local-consumer-review-survey/> [Accessed 18 May 2024].
- [3] Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*.
- [4] Yutika, C. H., Adiwijaya, A., & Al Faraby, S. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422-430.
- [5] Yustihan, S. R., Adikara, P. P., & Indriati, I. (2021). Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(3), 1017-1023.
- [6] Putra, B. A., Kristian, Y., Setiawan, E. I., & Santoso, J. (2022). Aspect based Sentiment Analysis Aduan Mahasiswa UMSIDA Dimasa Pandemi Menggunakan LSTM. *INSYST: Journal of Intelligent System and Computation*, 4(1), 45-54.
- [7] Astriningsih, W. (2023). *Identifikasi Multi Aspek Dan Sentimen Analisis Pada Review Hotel Menggunakan Deep Learning* (Doctoral dissertation, Universitas Islam Indonesia).
- [8] Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 13(2), 845-863.

- [9] Wang, J., Xu, B., & Zu, Y. (2021, July). Deep learning for aspect-based sentiment analysis. In *2021 international conference on machine learning and intelligent systems engineering (MLISE)* (pp. 267-271). IEEE.
- [10] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.
- [11] Zola, A. Definition Google Map. [Online] Available at : <https://www.techtarget.com/whatis/definition/Google-Maps> [Accessed 18 May 2024].
- [12] Mohammed, M. T., & Rashid, O. F. (2023). Document retrieval using term frequency inverse sentence frequency weighting scheme. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(3), 1478-1485.
- [13] Valkenborg, D., Rousseau, A. J., Geubbeldmans, M., & Burzykowski, T. (2023). Support vector machines. *American Journal of Orthodontics and Dentofacial Orthopedics*, 164(5), 754-757.
- [14] Rainforth, T. 2022. Chapter 3, Part 2 : Support Vector Machines. [Online] Available at : [https://www.robots.ox.ac.uk/~twgr/assets/teaching/lecture\\_4.pdf](https://www.robots.ox.ac.uk/~twgr/assets/teaching/lecture_4.pdf) [Accessed 18 May 2024].
- [15] NUFUS, G. K., Mustafid, M., & Gernowo, R. (2022). *ANALISIS SENTIMEN BERBASIS ASPEK UNTUK KEPUASAN PENGGUNA APLIKASI VIDEO ON DEMAND DENGAN MODEL LONG SHORT TERM MEMORY* (Doctoral dissertation, School of Postgraduate Studies).
- [16] Cai, W., Zhang, W., Hu, X., & Liu, Y. (2020). A hybrid information model based on long short-term memory network for tool condition monitoring. *Journal of Intelligent Manufacturing*, 31, 1497-1510.



