

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Hotel merupakan salah satu produk pariwisata yang sangat penting untuk dipertimbangkan baik dari segi fasilitas, pelayanan ataupun jarak tempuh perjalanan wisata. Saat ini sudah banyak website wisata yang menyediakan fasilitas internet untuk menuliskan opini dan pengalaman pribadinya secara online. Analisa sentimen atau opinion mining merupakan salah satu solusi mengatasi masalah untuk mengelompokkan opini atau review menjadi opini positif atau negatif secara otomatis. Namun dari ulasan tersebut selain dari banyaknya rating yang diberikan, kita terkadang masih kesulitan menentukan mana ulasan yang baik atau buruk, terlebih lagi banyaknya ulasan yang berbeda-beda membutuhkan waktu yang banyak untuk dibaca. Oleh karena itu perlunya proses klasifikasi analisis sentimen dari ulasan-ulasan tersebut untuk mengelompokkan ulasan positif dan negatif namun jika data ulasan hotel diolah dengan baik akan mempermudah pihak hotel dalam menarik kesimpulan berdasarkan ulasan pelanggan, sehingga dapat memperbaiki layanan yang dimiliki[1].

Analisis sentimen adalah sebuah proses yang memahami, mengekstraksi, dan mengolah data teks secara otomatis untuk menemukan jenis sentimen pada teks tersebut (Rozi, et al., 2012). Jenis sentimen yang didapatkan yaitu positif atau negatif. Analisis sentimen berguna untuk memudahkan pengguna pada proses

memahami sentiment sehingga dapat melakukan penentuan keputusan pada suatu objek[2].

Analisis sentimen disebut juga opinion mining, adalah bidang ilmu yang menganalisa pendapat, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik, dan atribut mereka. Analisis sentimen berfokus pada opini-opini yang mengekspresikan atau mengungkapkan sentimen positif atau negatif [3].

Metode Word2Vec dapat digunakan untuk merepresentasikan kata kata dalam bentuk matematis. Word2Vec merupakan sebuah algoritma untuk mempelajari posisi kedekatan semantic antar kata dari sebuah teks masukan, Word2Vec dapat membantu memecahkan soal analogi yang sering muncul pada tes psikotes. Word2vec memiliki dua arsitektur yaitu Skip-gram dan CBOW (Continuous Bag-of-Word)[4]. CBOW memprediksi kata dari konteks pada kata, sedang Skip-gram memprediksi konteks dari kata.

Beberapa peneliti yang telah melakukan klasifikasi analisis sentimen pada ulasan hotel salah satunya menggunakan metode word embedding. Pada penelitian menggunakan metode word embedding ini akan menghasilkan ruang vektor kata yang substruktur yang bermakna proses ini menghasilkan fitur dimensi tinggi space yang memrepresentasikan teks dalam dokumen dengan sekumpulan kata, dengan representasi vektornya dapat dibentuk menjadi semantik model distribusi dengan menangkap kata yang ada [5]. Oleh karena itu penulis mengajukan penelitian terkait *word embedding* Representasi kata menjadi vector pada ulasan hotel di Indonesia menggunakan metode Word2Vec.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat dirumuskan permasalahannya yaitu berapa persen akurasi yang dihasilkan dengan word embedding dengan menggunakan metode Word2Vec dalam representasi kata untuk menentukan sentimen suatu kalimat dalam ulasan hotel.

## 1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas, terdapat beberapa batasan masalah pada penelitian ini untuk pengelompokan ulasan hotel berbahasa Indonesia menggunakan Word Embedding dengan metode Word2Vec. adapun batasan masalah sebagai berikut :

1. Sumber data didapatkan dari halaman website Traveloka yang nantinya pengambilan data tersebut dilakukan dengan menggunakan sebuah aplikasi desktop yang bernama WebHarvy.
2. Menggunakan library Numpy, Pandas, NLTK, gensim, tensorflow keras.
3. Periode pengambilan data diabil pada bulan Juli - Agustus 2020 dengan jumlah data 5000. Dimana dari penelitian sebelumnya sudah terdapat dataset sebesar 5.000 di penelitian yang berbeda.
4. Dataset ulasan yang diambil dari kota wisata yang ada di kota Yogyakarta, Jakarta dan Bandung.
5. menggunakan algoritma Klasifikasi Long Short Term Memory (LSTM) untuk mengklasifikasikan ulasan negatif dan positif.
6. Menggunakan Confusion Matrix untuk menentukan Evaluasi.

#### 1.4 Maksud dan Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maksud dan tujuan dilakukannya penelitian ini adalah sebagai berikut :

- Maksud Penelitian:

Untuk membandingkan tingkat akurasi antara Word Embedding menggunakan metode Word2Vec dengan tanpa menggunakan Word Embedding

- Tujuan Penelitian :

Untuk mengimplementasikan Word Embedding dengan metode Word2Vec sehingga mampu mengetahui tingkat akurasi metode tersebut dalam penggunaan sentimen analisis pada ulasan hotel di Indonesia.

#### 1.5 Metode Penelitian

Dalam menjalankan penelitian ini, peneliti melakukan beberapa tahap metode penelitian. Adapun metode dijelaskan sebagai berikut :

##### 1.5.1 Pengumpulan Data

Pada tahap pengumpulan data dalam penelitian ini didapatkan dari cara *scraping data* dari situs Traveloka dengan jumlah data 10.000. Dimana dari penelitian sebelumnya sudah ada dataset sebesar 5.000 di penelitian yang berbeda. Maka dipenelitian ini dataset yang didapatkan dengan cara *scraping data* menambahkan data sebanyak 5.000, total data menjadi 10.000 dataset. Pada metode pengumpulan data ini, yaitu dengan cara observasi. Observasi adalah teknik



pengumpulan data melalui observasi langsung. Peneliti menggunakan indra untuk mengamati objek yang akan diamati di lokasi. Peneliti diposisikan sebagai pengamat atau orang luar. Dari data yang diambil, semuanya berupa data teks. Teks adalah bentuk yang paling tidak terstruktur sehingga harus diubah menjadi data yang siap diolah yang nantinya akan digunakan dalam proses pengujian penelitian. Cara pengambilan data tersebut dilakukan dengan menggunakan sebuah aplikasi desktop yang bernama WebHarvy.

### 1.5.2 Annotation

Setelah data terkumpulkan dari *web scraping*, data kemudian dianalisis dengan memberikan sebuah label pada setiap ulasan kata. Label pada penelitian ini diberikan label 1 untuk sentiment positif sedangkan label 0 diberikan untuk sentiment negatif. Tahap ini bertujuan untuk memberikan pengetahuan pada model yang akan dibentuk di tahap pelatihan data. Sehingga hasil akhir dari manual labeling yaitu 50% data sentiment positif dan 50% data sentiment negatif, data dibuat sama agar mempermudah dalam tahap proses pelatihan dan pengujian data.

Tahapan Text annotation membantu untuk memvisualisasikan teks atau kata yang penting melalui computer vision saat pelabelan teks. Kualitas dan keakuratan dilakukan untuk memastikan setiap teks penting dianotasi dengan metadata untuk memastikannya dapat menghasilkan campuran pelabelan dan anotasi yang tepat untuk berbagai kebutuhan dalam pemrosesan bahasa kemudian memberikan pembelajaran pada model yang akan dibentuk di tahap pelatihan data.

### 1.5.3 Preprocessing Data

Preprocessing data pada penelitian ini melakukan perubahan data teks yang mentah agar siap diolah dalam pengujian. Preprocessing meliputi Data celaning, case folding, filtering, dan tokenization. Cleaning merupakan proses untuk menghapus simbol-simbol yang kurang penting dalam data ulasan hotel. Pada proses preprocessing data, membutuhkan sebuah library bernama NLTK dan keras.

### 1.5.4 Ekstraksi Fitur

Ekstraksi fitur adalah proses pengambilan ciri sebuah objek yang dapat menggambarkan karakteristik dari objek tersebut. Ekstraksi fitur atau pembangunan vektor pada tahapan penelitian ini dilakukan untuk mengubah sebuah kata dari kata-kata menjadi numerik dalam proses analisis sentimen. Karena komputer hanya bisa membaca sebuah angka. Pembangunan vektor memakai ekstraksi fitur word embedding bagian dari ekstraksi fitur, pada pembangunan vektor penelitian kali ini menggunakan metode Word2Vec.

### 1.5.5 Implementasi Algoritma Klasifikasi

Pada tahapan ini dilakukan proses klasifikasi data yang telah melalui tahap ekstraksi fitur. Dalam tahap klasifikasi ini, algoritma klasifikasi yang akan digunakan dalam penelitian ini adalah Long Short Term Memory (LSTM) karena dalam beberapa penelitian LSTM menghasilkan akurasi yang tinggi ketika digunakan.

### 1.5.6 Evaluasi

Hasil dari implementasi, selanjutnya dilakukan proses pengujian. Akan dilihat perbandingan keberhasilan algoritma untuk mengklasifikasi ulasan sesuai dengan kelasnya dengan total data uji yang dilakukan. Dari hal tersebut dapat dilihat keakuratan atau performansi dari algoritma yang digunakan terhadap data yang ada. Metode evaluasi yang digunakan yaitu menggunakan metode Confusion Matrix

### 1.6 Sistematika Penulisan

Pada dasarnya penyusunan sistematika penulisan bertujuan untuk memudahkan pembaca dalam mengikuti apa yang dipaparkan dalam laporan penelitian ini. Untuk mengetahui uraian singkat yang memuat gambaran singkat secara keseluruhan isi masing-masing bab, maka dibuat sistematika sebagai berikut :

#### **BAB I : PENDAHULUAN**

Pada bab ini menjelaskan tentang dasar penelitian, yang berisi latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, metode penelitian dan sistematika penulisan dalam pembuatan laporan ini.

#### **BAB II : LANDASAN TEORI**

Bab ini membahas tentang tinjauan pustaka, yang berisi dasar-dasar teori yang berhubungan dengan topik penelitian ini.

**BAB III : METODOLOGI PENELITIAN**

Dalam bab ini berisi alur tahapan penelitian, alat berupa metode serta perangkat dan bahan berupa data yang akan digunakan dalam penelitian.

**BAB IV : IMPLEMENTASI DAN PEMBAHASAN**

Bab ini akan membahas tentang hasil penelitian yang dilakukan secara menyeluruh serta pembahasan pada penelitian yang dikaji.

**BAB V : PENUTUP**

Bab ini berisikan kesimpulan dan saran yang akan berguna dan dapat dimanfaatkan untuk penelitian selanjutnya.

