

BAB I

PENDAHULUAN

1.1 Latar Belakang

Usaha perhotelan adalah bisnis yang bergerak di bidang pariwisata yang sangat penting, baik dari segi fasilitas pelayanan maupun jarak tempuh dari hotel ke tempat wisata yang dituju. Selain itu, tingkat kepuasan pelanggan juga menjadi suatu aspek yang sangat penting dalam bisnis ini. Kepuasan pelanggan dapat dilihat dari ulasan mereka tentang pengalaman menggunakan hotel. 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 atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [2]. Analisis sentimen dapat digunakan untuk melakukan klasifikasi ulasan ke dalam sentimen tertentu dengan memperhatikan faktor pada objek sentimennya. Dalam analisis sentimen, ada sebuah

metode yang lebih maju yaitu menggunakan *word embedding*, representasi kata dalam vektor, banyak peneliti yang telah menggunakan metode ini dalam penelitiannya. Penelitian terdahulu yang telah menggunakan *word embedding* menyatakan bahwa pada penggunaan *word embedding* akan mendapatkan hasil kualitas yang lebih baik dalam tugas pemrosesan teks tingkat lanjut, termasuk analisis sentimen [3]. Dalam melakukan pemrosesan data teks yang tidak terstruktur ini, para peneliti menggunakan metode yang disebut dengan *Natural Language Processing* atau biasa disingkat NLP, NLP adalah sebuah metode pembentukan model komputasi dengan perantara bahasa alami [4].

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* metode *Global Vector (GloVe)* pada analisis sentimen ulasan hotel berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat dirumuskan permasalahannya yaitu apakah ada pengaruh akurasi dari penggunaan *word*

embedding dengan metode *Global Vector (GloVe)* dalam representasi kata untuk menentukan sentimen suatu kalimat dalam ulasan hotel.

1.3 Batasan Masalah

Pada penelitian ini akan meneliti pengelompokan sentimen dari ulasan hotel berbahasa Indonesia menggunakan *word embedding* dengan metode *Global Vector (GloVe)*.

1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan dari penelitian ini adalah untuk mengetahui kecepatan akurasi dari penggunaan ekstraksi fitur khususnya dengan menggunakan *word embedding* metode *Global Vector (GloVe)* lebih baik dari tidak menggunakan *word embedding* dalam menganalisis sentimen ulasan hotel dalam bahasa Indonesia.

1.5 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan untuk memberikan manfaat berupa meningkatnya akurasi dalam menganalisis sentimen dari suatu kalimat dalam suatu teks jika menggunakan ekstraksi fitur khususnya *word embedding*.

1.6 Metode Penelitian

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

1.6.1 Pengumpulan Data

Dalam penelitian ini menggunakan dataset yang didapatkan dari

scraping data dari halaman website Traveloka dengan jumlah 10.000. 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.

1.6.1.1 Manual Labeling

Setelah data yang didapatkan dari *web scrapping*. Data kemudian dianalisis dengan memberikan label pada setiap kata ulasan. Label pada penelitian kali ini menggunakan positif dan negatif. Label pada penelitian ini digunakan label 1 untuk sentimen positif sedangkan 0 untuk sentimen negatif. Tahap ini bertujuan untuk memberikan pembelajaran 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.

1.6.1.2 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.

1.6.1.3 Ekstraksi Fitur

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 *Global Vector (GloVe)*.

1.6.1.4 Implementasi Algoritma Klasifikasi

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

1.6.1.5 Evaluasi

Hasil dari implementasi selanjutnya dilakukan proses pengujian. Akan dilihat perbandingan keberhasilan algoritma untuk mengklasifikasikan ulasan sesuai dengan kelasnya dengan total data uji yang dilakukan dengan menggunakan algoritma *confusion matrix*. Dari hal tersebut dapat dilihat keakuratan atau performansi dari algoritma yang digunakan terhadap data yang ada.

1.7 Sistematika Penulisan

Untuk mengetahui uraian singkat yang memuat gambaran singkat secara keseluruhan isi masing-masing bab, maka dibuat sistematika sebagai berikut :

BAB I PENDAHULUAN

Bab ini membahas tentang uraian latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tentang teori-teori yang digunakan serta konsep dasar yang menjadi landasan dari 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 HASIL DAN PEMBAHASAN

Bab ini membahas tentang hasil yang didapatkan dari penelitian yang sudah dilakukan oleh penulis

BAB V PENUTUP

Berisi kesimpulan dari penelitian yang dibuat serta saran untuk memperbaiki kekurangan yang ada pada penelitian.