

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bisnis hotel adalah salah satu bisnis di bidang pariwisata yang sangat penting baik dari segi fasilitas, pelayanan maupun jarak tempuh dari hotel ke tempat wisata. Selain itu, 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 dan mana yang buruk, terlebih lagi banyaknya ulasan yang berbeda-beda membutuhkan waktu yang banyak untuk dibaca. Oleh karena itu perlunya analisis sentimen dari ulasan-ulasan tersebut untuk mengelompokkan ulasan positif dan negatif sehingga pihak hotel dapat menggunakan hasilnya untuk menentukan apa yang harus mereka perbaiki dari bisnis mereka [1].

Dalam melakukan analisis sentimen, terdapat tahapan ekstraksi fitur yang berguna untuk mengubah teks menjadi sebuah data yang dapat diklasifikasi. Dalam ekstraksi fitur ada banyak cara yang bisa dilakukan seperti *Lexical Description*, TF, TF-IDF, Bigram, Unigram dan sebagainya. Selain itu *word embedding* juga termasuk ke dalam ekstraksi fitur, *word embedding* tidak hanya mencocokkan kata per kata namun juga bisa mencocokkan kalimat per kalimat. Penelitian terdahulu yang telah menggunakan *word embedding* menyatakan bahwa penggunaan *word*

embedding akan memperoleh kinerja yang lebih baik untuk tugas pemrosesan teks tingkat lanjut, termasuk analisis sentimen [2]. Penelitian yang dilakukan oleh Jordhy Fernando et al. [1] menggunakan metode *double embedding* dan *attention mechanism* untuk melakukan analisis sentimen pada ulasan hotel dan mendapatkan hasil F1-measure sebesar 0,914 dan 0,90. Julio Savigny et al. [3] membandingkan beberapa metode *word embedding* seperti vektor kata rata-rata, vektor kata rata-rata dengan TF-IDF, vektor paragraf, dan dengan menggunakan algoritma Convolutional Neural Network (CNN) untuk menganalisis emosi pengguna pada komentar di Youtube dan menghasilkan kesimpulan performa yang terbaik didapatkan menggunakan metode CNN dengan akurasi sebesar 76,2%.

Pada penelitian ini penulis akan menguji apakah penggunaan ekstraksi fitur *word embedding* dapat menghasilkan akurasi yang lebih tinggi daripada tidak menggunakan ekstraksi fitur. Oleh karena itu penulis mengajukan penelitian terkait *word embedding* menggunakan metode BERT pada analisis sentimen ulasan hotel berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan permasalahannya yaitu apakah ada pengaruh akurasi dari penggunaan *word embedding* dengan metode BERT 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 BERT.

1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan dari penelitian ini adalah untuk mengetahui apakah akurasi dari penggunaan ekstraksi fitur khususnya dengan metode *word embedding*-BERT lebih baik dari tidak menggunakan ekstraksi fitur 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 tahapan metode penelitian. Tahapan – tahapan yang dilakukan adalah sebagai berikut :

1.6.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari *scraping* data dari laman Traveloka dengan jumlah 10.000 data. Dari data yang diambil merupakan ulasan dari laman Traveloka dan

semuanya data yang diambil 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.

1.6.2 *Manual Labeling*

Setelah melakukan scraping data dari laman Traveloka, data kemudian dianalisis dengan memberikan label pada setiap kata ulasan. Label yang digunakan pada penelitian ini adalah label positif dan negatif. Label positif di representasikan dengan angka 1 sedangkan label negatif dengan angka 0. Tahap ini bertujuan untuk memberikan pembelajaran pada model yang akan dibentuk di tahap pelatihan data. Sehingga hasil akhir dari manual *labeling* yang dilakukan adalah 50% data sentimen positif dan 50% data sentimen negatif, pembagian data dibuat sama agar mempermudah dalam tahap proses pelatihan dan pengujian data.

1.6.3 *Preprocessing Data*

Preprocessing data dalam penelitian ini dilakukan untuk mengubah data teks yang mentah agar siap diolah dalam pengujian. Tahap *preprocessing data* meliputi *case folding*, dan *tokenization*.

1.6.4 Ekstraksi Fitur

Ekstraksi fitur atau pembangunan vektor pada tahap penelitian ini dilakukan untuk mengubah sebuah kata menjadi numerik dalam proses analisis sentimen. Dikarenakan komputer hanya bisa membaca sebuah angka, maka tahap ini perlu dilakukan.

Pembangunan vektor menggunakan ekstraksi fitur *word embedding* sebagai bagian dari ekstraksi fitur. Pada pembangunan vektor pada penelitian ini akan mengimplementasikan metode *Bidirectional Encoder Representations from Transformers* (BERT).

1.6.5 Evaluasi

Hasil dari implementasi selanjutnya dilakukan proses pengujian akhir. Pada tahap ini 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 akurasi atau performa dari algoritma yang digunakan terhadap data yang ada.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penelitian ini terbagi menjadi beberapa bagian, yaitu :

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 METODE PENELITIAN

Bab ini berisi tentang pembahasan metode-metode yang digunakan dalam melaksanakan penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

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

BAB V PENUTUP

Bab ini merupakan bab terakhir berisi tentang kesimpulan dari penelitian yang telah dilakukan dan juga berisi saran yang bisa diberikan oleh peneliti kepada peneliti selanjutnya.

