

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital memberikan hal baru dalam cara berkomunikasi melalui pengguna (blog, forum, media sosial, ulasan situs website, situs website niaga, dll.). Dengan pertumbuhan teknologi ini, berbagai elemen baik secara individu maupun organisasi memiliki minat yang besar dalam teknologi data mining yang menggunakan sumber informasi. Analisis sentimen adalah salah satu dari bidang penelitian yang paling sering digunakan dalam ilmu komputer, bertujuan untuk mengidentifikasi dan mengekstraksi opini pengguna [1].

Bidang usaha perhotelan merupakan usaha bisnis yang bergerak pada bidang pariwisata yang penting baik dari segi fasilitas, pelayanan ataupun jarak tempuh perjalanan wisata. Saat ini sudah banyak website wisata yang menyediakan fasilitas untuk pengguna internet menuliskan opini dan pengalaman pribadinya secara online. Disamping itu dengan perkembangan teknologi, ulasan tidak hanya membantu calon konsumen untuk mengambil keputusan dalam menentukan suatu produk, tetapi juga memberikan umpan balik untuk pelaku usaha tersebut [2]. Analisis sentimen telah banyak digunakan di antara perusahaan untuk secara otomatis mengekstrak opini atau ulasan baik tentang produk atau layanan mereka. Namun dari ulasan tersebut selain dari banyaknya ulasan yang diberikan, kita terkadang masih kesulitan menentukan ulasan yang baik dan buruk secara mendetail, terlebih lagi banyaknya ulasan yang berbeda-beda membutuhkan waktu untuk dibaca dan beberapa perusahaan mungkin membutuhkan lebih banyak analisis mendetail dengan menggunakan analisis sentimen berbasis aspek [1]. Oleh karena itu perlunya proses analisis sentiment berbasis aspek ulasan-ulasan tersebut untuk mengelompokkan ulasan positif dan negatif secara mendetail. Jika data ulasan hotel diolah dengan baik diharapkan akan mempermudah pihak hotel dalam menarik kesimpulan berdasarkan ulasan pelanggan, sehingga dapat memperbaiki layanan yang dimiliki.

Analisis sentimen berbasis aspek atau biasa disingkat ABSA adalah analisis yang bertujuan untuk mengidentifikasi opini dan emosi yang diberikan

sesorang yang lebih dalam dari teks ulasan. Tujuan dari ABSA adalah mengidentifikasi aspek dari entitas yang diberikan dan mengidentifikasi sentiment yang diekspresikan pada tiap aspek[3]. Dalam analisis sentimen berbasis aspek, para peneliti menggunakan metode *Natural Language Processing* untuk dapat memecahkan masalah. *Natural Language Processing* adalah sebuah metode pembentukan model komputasi dengan perantara bahasa alami yang berupaya memahami bahasa alami manusia dengan segala aturan gramatika dan semantiknya, serta mengubah bahasa tersebut menjadi representasi formal yang dapat diproses oleh komputer [4].

Pada penelitian ini penulis bertujuan untuk melihat seberapa besar akurasi model BERT dalam menganalisis kalimat berbahasa Indonesia yang didapat dari ulasan hotel berbahasa Indonesia. Oleh karena itu penulis mengajukan penelitian terkait analisis sentimen berbasis aspek menggunakan metode BERT (*Bidirectional Encoder Representations from Transformers*) untuk ulasan hotel berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, diperoleh rumusan masalah yaitu berapakah hasil akurasi penggunaan metode BERT (*Bidirectional Encoder Representations from Transformers*) dalam representasi kata dalam menentukan sentimen berbasis aspek suatu kalimat dalam ulasan hotel?

1.3 Batasan Masalah

Adapun batasan masalah yang akan dirancang agar penelitian lebih terfokus dan tidak menyimpang yakni :

1. Pengelompokan sentiment berbasis aspek dari ulasan hotel Indonesia.
2. Data yang diambil dari situs Traveloka Indonesia.
3. Data yang digunakan menggunakan Bahasa Indonesia.
4. BERT model pre-trained yang digunakan pada penelitian berasal dari IndoNLU (indobenchmark.com) dengan nama `indobert-base-pl`.

5. Klasifikasi sentimen terdiri dari 2 kelas, yaitu Negative dan Positive.

1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan dari penelitian ini adalah untuk mengetahui akurasi yang didapat dari penggunaan *word embedding* metode BERT (*Bidirectional Encoder Representations from Transformers*) dalam menganalisis sentiment berbasis aspek pada ulasan hotel dalam bahasa Indonesia.

1.5 Manfaat Penelitian

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

1.6 Metode Penelitian

Berikut metode penelitian yang digunakan dalam proses pengerjaan penelitian ini dijelaskan sebagai berikut:

1.6.1 Metode Pengumpulan Data

Pada tahapan ini dilakukan pengumpulan data dengan cara melakukan scraping data pada ulasan atau komentar pada situs website Traveloka. Semua data yang diambil berupa teks. Teks terdiri dari bentuk yang tidak terstruktur sehingga harus diubah menjadi data yang diolah yang nantinya digunakan dalam proses pengujian penelitian. Craping data dilakukan menggunakan aplikasi Webharvy.

1.6.2 Manual Labeling

Setelah melakukan scraping data dari laman Traveloka, data kemudian dianalisis dengan memberikan label kategori aspek yang dibagi menjadi 5 kategori serta 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.

Selanjutnya pada setiap kata ulasan Tahap ini bertujuan untuk memberikan pembelajaran pada model yang akan dibentuk di tahap pelatihan 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

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

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

Pada tahap ini penulis melakukan perhitungan akurasi (dalam bentuk persen) atau performa dari algoritma yang digunakan terhadap data yang ada. Pada tahap ini akan dilihat tingkat keberhasilan algoritma untuk mengklasifikasi ulasan sesuai dengan kelasnya dengan total data uji yang dilakukan dengan menggunakan algoritma *confusion matrix*.

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

	<p>Bab ini berisi tentang teori-teori yang digunakan serta konsep dasar yang menjadi landasan dari penelitian ini.</p>
BAB III	<p>METODE PENELITIAN</p> <p>Bab ini berisi gambaran umum penelitian, alat dan bahan penelitian, serta langkah-langkah penelitian.</p>
BAB IV	<p>IMPLEMENTASI DAN PEMBAHASAN</p> <p>Bab ini akan membahas mengenai proses implementasi dan pengembangan yang dilakukan secara menyeluruh serta pembahasan pada penelitian yang dikaji.</p>
BAB V	<p>PENUTUP</p> <p>Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran kedepannya untuk memperbaiki kekurangan yang ada penelitian.</p>

