

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada saat ini banyak pelaku usaha yang mengembangkan usaha mereka menjadi online dengan melalui marketplace. Jumlah pengguna marketplace di Indonesia setiap tahun mengalami pertumbuhan dan perkembangan yang sangat cepat, hal ini tak terlepas dari banyaknya jumlah pengguna internet di Indonesia yaitu mencapai 196,7 juta jiwa [1]. Pada setiap marketplace menyediakan fitur kepada para konsumen untuk menuliskan ulasan dan pengalaman mereka terhadap suatu produk. Analisis sentimen atau *Opinion Mining* merupakan salah satu solusi untuk mengatasi permasalahan pengelompokan ulasan atau review, menjadi ulasan positif atau negatif secara otomatis. Terdapat banyaknya ulasan yang berbeda-beda, sehingga dapat mempersulit untuk menentukan ulasan yang baik atau buruk dan membutuhkan waktu yang lama untuk dibaca. Oleh karena itu perlu proses analisis sentimen untuk mengidentifikasi berbagai macam ulasan dari konsumen terhadap beberapa produk termasuk dalam ulasan positif atau negative [2].

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, sehingga analisis sentimen berguna untuk memudahkan pengguna melakukan identifikasi sentimen agar dapat melakukan penentuan keputusan pada suatu objek [3]. Analisis sentimen disebut juga *Opinion Mining* adalah suatu teknik

untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini, untuk mendapatkan informasi tentang sentimen bernilai positif atau negatif [4]. Analisis sentimen berfokus pada ulasan yang mengekspresikan atau mengungkapkan sentimen positif atau negatif.

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)* [5]. *CBOW* memprediksi kata dari konteks pada kata, sedang *Skip-gram* memprediksi konteks dari kata.

Penelitian ini melakukan klasifikasi analisis sentimen berdasarkan ulasan yang diberikan, dengan menggunakan *word embedding* metode *Word2Vec* model arsitektur *CBOW (Continuous Bag-of-Word)*. Pada analisis sentimen tersebut menggunakan metode *Word Embedding* yang akan menghasilkan ruang vektor kata yang substruktur 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 [6]. Oleh karena itu penulis melakukan penelitian terkait *Word Embedding* dengan menggunakan Metode *Word2Vec* untuk merepresentasikan kata menjadi vector pada ulasan marketplace di Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini yaitu, berapa persen akurasi yang akan dihasilkan word embedding dengan menggunakan metode Word2Vec dalam representasi kata untuk menentukan sentimen suatu kalimat dalam ulasan marketplace.

## 1.3 Batasan Masalah

Di dalam penelitian ini diberikan beberapa batasan masalah adalah sebagai berikut :

1. Bahasa yang digunakan untuk pengelompokan ulasan pada marketplace adalah Bahasa Indonesia.
2. Menggunakan algoritma Klasifikasi *Long Short Term Memory (LSTM)* untuk mengklasifikasikan ulasan negatif dan positif.

## 1.4 Tujuan Penelitian

Untuk mengimplementasikan *Word Embedding* dengan metode *Word2Vec* dan tanpa menggunakan *Word Embedding*, sehingga mampu mengetahui tingkat akurasi tersebut dalam penggunaan sentimen analisis pada ulasan marketplace di Indonesia..

## 1.5 Manfaat Penelitian

Adapun manfaat penelitian dalam penyusunan skripsi ini adalah :

1. Memberikan pengetahuan tentang metode *Word Embedding-Word2Vec*.

2. Mengetahui kemampuan tingkat akurasi dari metode *Word Embedding-Word2Vec* dalam penggunaan sentimen analisis.

## 1.6 Metode Penelitian

Tahapan awal yang dilakukan untuk mendapatkan informasi dengan cara mengumpulkan dan mempelajari literatur melalui kajian dari buku, jurnal, maupun artikel-artikel yang terdapat pada internet untuk dianalisis dan mampu dijadikan bahan pendukung skripsi penelitian yang berkaitan dengan permasalahan yang akan diteliti dan sekaligus sebagai tambahan referensi bagi penulis.

Pada tahap ini penulis menggunakan data sekunder yang sudah tersedia pada halaman website kaggle. Selanjutnya pada data tersebut dilakukan proses *preprocessing* data, pada penelitian ini melakukan perubahan data teks yang mentah agar siap diolah dalam pengujian. *Preprocessing* meliputi data *cleaning*, *case folding*, *filtering*, dan *tokenization*. *Cleaning* merupakan proses untuk menghapus simbol-simbol yang kurang penting dalam data ulasan marketplace.

Setelah data yang digunakan melalui tahap *preprocessing*, sehingga data tersebut sudah siap digunakan sesuai dengan kebutuhan penelitian. Pada tahap selanjutnya yaitu *Word embedding*, adalah sebuah pendekatan yang digunakan untuk mempresentasikan vector kata, yang merupakan pengembangan komputasi permodelan kata-kata yang sederhana seperti perhitungan menggunakan jumlah dan frekuensi kemunculan kata dalam sebuah dokumen [5].

Pada tahapan selanjutnya setelah melalui *Word Embedding*, 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.

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

### 1.7 Sistematika Penulisan

Adapun sistematika penulisan pada skripsi ini terdiri beberapa bab, adalah sebagai berikut :

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

Pada penelitian ini bab pendahuluan membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metodologi, dan sistematika penulisan laporan skripsi. Uraian di bab pendahuluan ini memberikan gambaran kepada pembaca mengenai maksud dan tujuan penulis dalam melakukan penelitian.

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

Landasan teori berisi tinjauan pustaka, dasar-dasar teori yang digunakan dalam penyusunan laporan skripsi ini. Teori yang dijabarkan akan menjadi acuan dalam melakukan analisis yang akan dibahas pada bab III.

**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**

Implementasi dan Pembahasan, pada bab ini merupakan tahapan yang penulis lakukan dalam membahas tentang hasil penelitian secara menyeluruh, mengkaji pembahasan dari hasil penelitian, dan evaluasi terhadap hasil yang didapat.

**BAB V: PENUTUP**

Penutup berisi kesimpulan dan saran yang diharapkan dapat bermanfaat untuk pengembangan penelitian selanjutnya.

