

BAB I

PENDAHULUAN

1.1. Latar Belakang

Bahasa adalah salah satu bentuk komunikasi yang vital. Komunikasi adalah proses dimana pertukaran informasi terjadi anatar kelompok orang dengan bantuan bahasa. Perkembangan Teknologi Informasi yang pesat membuat manusia semakin bebas untuk menyalurkan pendapat atau opini ke dalam web atau media sosial. Ada milyaran informasi yang tersebar di berbagai macam teknologi seperti Website, dimana diantaranya menggunakan bahasa natural. Bahasa natural adalah bahasa yang secara umum digunakan oleh manusia dalam berkomunikasi satu sama lain. Bahasa yang diterima oleh komputer butuh untuk diproses dan dipahami terlebih dahulu supaya maksud dari *user* bisa dipahami dengan baik oleh komputer. Perkembangan *Artificial Intelligence* yang tumbuh secara pesat dalam bidang riset maupun penerapannya di dunia nyata. Salah satu perkembangan yang pesat terlihat pada komputasi *linguistic* contohnya adalah *Natural Language Processing*. *Natural Language Processing (NLP)* merupakan salah satu cabang ilmu *AI* yang berfokus pada pengolahan bahasa natural dengan pendekatan komputerisasi yang digunakan untuk menganalisis data, teks, ucapan, dan lain-lain. Oleh karena itu *NLP* dapat melakukan pengolahan bahasa alami seperti bahasa manusia pada umumnya (Mohini Chaudhari and Sharvari Govilkar, 2015).

Sebagian besar masyarakat Indonesia menggunakan bahasa daerah masing-masing dalam kehidupannya sehari-hari, karena Indonesia memiliki beragam bahasa daerah. Tetapi bukan berarti mereka tidak mengerti Bahasa Indonesia. Untuk berkomunikasi dengan masyarakat yang beda suku atau daerah tentunya akan menggunakan Bahasa Indonesia sebagai Bahasa Nasional. Beragam kata dan ungkapan yang diungkapkan dalam berinteraksi di kehidupan sehari-hari ataupun di media sosial mengandung sentimen. Sentimen yang dimaksud adalah pengelompokan dari kalimat yang diungkapkan baik itu positif, negatif atau

netral. Di era sekarang ini tentunya media sosial menjadi suatu media publik yang sering digunakan masyarakat untuk menuangkan kata atau ungkapan. Baik itu tentang politik, hukum, ekonomi, teknologi dan lain-lain. Oleh karena itu diperlukan sebuah model yang dapat mengelompokkan ungkapan-ungkapan tersebut apakah ungkapan tersebut positif, negatif dan netral.

Pembangunan model ini akan menggunakan *library Spacy*. *Spacy* merupakan sebuah *library* bahasa pemrograman *Python*. *Library spacy* merupakan sebuah *library open source* yang bisa dipergunakan untuk membuat sebuah sistem NLP (Natural Language Processing). Salah satu fasilitas yang dimiliki *spacy* adalah model *Text Classification* yang dapat mendeteksi kalimat berdasarkan sentimen negatif, positif dan netral. Untuk menjalankan *library* ini dibutuhkan dataset Bahasa Indonesia yang diolah menggunakan fasilitas dari *Spacy*. Model *spacy* sudah tersedia dalam beberapa bahasa diantaranya Bahasa Inggris, Bahasa Jerman, Bahasa Spanyol, Bahasa Portugis, Bahasa Prancis, Bahasa Italia, dan Bahasa Belanda. Untuk pengembangan *Text Classification* Bahasa Indonesia sejauh ini belum ada dikembangkan melalui *Library Spacy*.

Dengan adanya *Library Spacy* ini diharapkan dapat membantu dalam membuat sebuah model *Text Classification* menemukan sentimen kalimat Bahasa Indonesia. Melalui model yang dibuat diharapkan dapat menambah sistem dan sumber daya *NLP* Bahasa Indonesia yang masih terbatas. Sehingga dapat mengejar ketertinggalan perkembangan Bahasa Indonesia dalam bidang pemrosesan Bahasa alami.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah yang ditemukan adalah sebagai berikut :

- a. Bagaimana membangun sebuah model *Natural Language Processing* untuk mengidentifikasi sentimen kalimat Bahasa Indonesia dengan *library Spacy* dengan bahasa pemrograman *Python*?
- b. Bagaimana tingkat akurasi dari pelatihan yang didapat?

1.3. Batasan Masalah

Bila dilihat dari perumusan masalah diatas serta persyaratan tertentu bahwa batasan permasalahan yang ada pada Tugas Akhir ini yakni:

- a. Bahasa pemrograman yang digunakan adalah *Python*.
- b. Model yang dibuat hanya untuk mengidentifikasi sentimen kalimat Bahasa Indonesia.

1.4. Tujuan Penelitian

Tujuan yang hendak dicapai dalam penelitian ini antara lain :

- a. Mengimplementasikan *Natural Language Processing* untuk membuat model klasifikasi kalimat Bahasa Indonesia.
- b. Menganalisis kinerja pelatihan model menggunakan *NLP*.

1.5. Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini secara garis besar diuraikan sebagai berikut.

BAB I : PENDAHULUAN

BAB ini menjabarkan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

BAB ini berisi tinjauan atas penelitian-penelitian terdahulu, untuk membandingkan dan menimbang seberapa orisinil dan penting penelitian yang akan dilakukan

BAB III : LANDASAN TEORI

BAB ini menjabarkan tentang teori-teori yang relevan digunakan sebagai dasar untuk menyelesaikan permasalahan pada penelitian ini, seperti konsep *Natural Language Processing*, Algoritma, dan dataset yang digunakan.

BAB IV : LANDASAN TEORI

BAB ini menjabarkan tentang alat dan bahan serta langkahlangkah penelitian.

BAB V : HASIL DAN PEMBAHASAN

BAB ini menjabarkan pembahasan hasil model terhadap dataset yang sudah dilatih. Hasil tersebut berupa waktu yang diperlukan pelatihan dan akurasi yang didapat dari pelatihan serta validasi model dengan data diluar dataset.

BAB VI : KESIMPULAN DAN SARAN

BAB ini berisi kesimpulan dari hasil penelitian yang dilakukan serta saran.