

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan pengguna media sosial kian meningkat dari waktu ke waktu. Seiring berjalannya waktu jumlah pengguna media sosial bahkan semakin tinggi hingga menyentuh angka miliaran. Banyaknya pengguna media sosial ini juga berdampak terhadap banyaknya jumlah unggahan atau *posting* oleh pengguna pada akun media sosial miliknya. Unggahan ini merupakan data, sehingga semakin banyak unggahan berarti semakin besar juga data yang tersedia. Hal ini menjadi suatu peluang bagi para ilmuwan dan ahli statistika untuk melakukan analisis lebih mendalam dengan menggunakan data yang masif itu. Salah satu media sosial yang dapat digunakan untuk mengambil data adalah Twitter karena memiliki banyak pengguna yang aktif mengunggah unggahan. Twitter merupakan sebuah media sosial *microblogging* dengan pengguna dapat mengunggah sebuah *tweet* yang berisikan maksimal 280 karakter (yang dahulu hanya 140 karakter). Sebuah *tweet* biasanya menceritakan tentang apa yang dialami oleh pengunggah, namun tak jarang juga *tweet* tersebut mengungkapkan apa yang dirasakan oleh seseorang terhadap suatu barang, tokoh, ataupun peristiwa.

Ketika ada peristiwa penting, jumlah *tweet* mengenai peristiwa tersebut akan meningkat dan pada tahun 2020 ada satu momen penting yang dinantikan oleh dunia khususnya Amerika Serikat, yaitu Pemilihan Presiden Amerika Serikat 2020. Pada pemilu kali ini terdapat dua orang kandidat yang bersaing untuk mendapatkan suara terbanyak, yaitu Donald Trump dari Partai Republik dan Joe Biden dari Partai

Demokrat yang pernah telah dinyatakan secara resmi sebagai Presiden Amerika Serikat berikutnya. Terdapat banyak *tweets* dari berbagai pengguna terhadap peristiwa ini, khususnya terkait dengan para kandidat. Dengan menganalisis sentimen dari sebuah *tweet* bisa didapatkan beberapa *insight* terkait pemilu tersebut. Analisis semacam ini dinilai akurat untuk memberikan gambaran seperti apa dinamika — bahkan bisa saja *outcome* dari sebuah pemilu [1]. Sebagai perbandingan, menurut Hootsuite terdapat sebanyak 68.700.000 pengguna pengguna Twitter di Amerika Serikat pada bulan Oktober 2020 [27] dan menurut sensus Amerika Serikat, populasi warga Amerika Serikat ada sebanyak 329.850.861 jiwa [28], yang berarti sekitar 20.8% warga Amerika Serikat menggunakan Twitter.

Pada penelitian ini analisis dilakukan dengan memanfaatkan *tweets* mengenai pemilu Amerika Serikat 2020 yang diambil menggunakan Snsrape dengan menggunakan nama kedua kandidat sebagai kata kunci pencarian. *Tweets* diambil secara *query*, kemudian *tweets* akan diklasifikasikan berdasarkan polaritas sentimennya. Klasifikasi dilakukan dengan menggunakan *deep learning* dengan *Long Short-Term Memory (LSTM)*. Dipilihnya metode *deep learning* adalah karena metode ini memberikan hasil klasifikasi yang lebih akurat dibandingkan dengan *linear classifier* dengan *machine learning*, seperti *Naïve Bayes Classifier* [2][3].

1.2 Maksud dan Tujuan

Penelitian ini dilakukan dengan tujuan yaitu mengklasifikasikan sejumlah *tweets* yang berkaitan dengan Pemilu Amerika Serikat 2020 yang telah diambil

menggunakan *scraper*. *Tweets* diambil dengan menggunakan kata kunci berupa nama dari kedua kandidat. Dengan melihat perbandingan jumlah sentimen positif dan sentimen negatif dari kedua kandidat, diharapkan bisa memberikan gambaran bagaimana dinamika pemilu dan citra para kandidat di mata publik dan kandidat mana yang cenderung lebih banyak dipilih.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini ialah:

1. *Prediction dataset* yang berisikan *tweets* (teks atau *corpus*) dari beberapa pengguna,
2. *Training dataset* yang digunakan untuk men-*train neural network* yang berisikan *tweets* beserta *sentiment* labelnya.
3. *Model neural network* yang dibuat dengan *library* TensorFlow.

1.4 Metode Penelitian

Metode penelitian yang dilakukan pada penelitian ini ialah studi literatur, eksplorasi dan analisis *dataset*, pembersihan *dataset*, pengembangan model *neural network*, evaluasi model *neural network*, dan implementasi model pada *dataset*. Metode penelitian akan dibahas lebih dalam dan terperinci pada Bab III.

1.5 Sistematika Penulisan

Skripsi ini terdiri dari lima bab. Kelima bab tersebut tersusun dengan rangkaian seperti berikut:

1. BAB I: PENDAHULUAN

Bab pertama menjelaskan latar belakang, tujuan, batasan masalah, metode penelitian dan sistematika penulisan dari penelitian yang dilaksanakan.

2. BAB II: LANDASAN TEORETIS

Bab kedua berisikan tentang dasar-dasar teori yang mendasari penelitian ini. Dasar-dasar teori tersebut antara lain mengenai Pemilu Amerika Serikat 2020, analisis sentimen, *deep learning* dan *neural network*, Python, Google Colab, *library* TensorFlow, dan evaluasi model.

3. BAB III: METODE PENELITIAN

Pada bab ketiga akan dibahas mengenai proses kerja dari penelitian secara detail mulai dari pengambilan data hingga evaluasi model.

4. BAB IV: EXPLORATORY DATA ANALYSIS

Di dalam bab ini juga akan dibahas mengenai model *neural network* yang telah dibuat dan hasil evaluasinya. Selain itu akan dibahas perbandingan hasil sentimen dengan hasil *popular vote* dari suatu *newsletter* serta perbandingan dengan hasil sentimen pada penelitian sebelumnya.

5. BAB V: PENUTUP

Pada bab terakhir akan dibahas mengenai kesimpulan dari penelitian yang telah dilakukan diikuti dengan saran untuk penelitian ini atau penelitian serupa di masa yang akan datang agar penelitian bisa menjadi lebih baik atau bisa memberikan hasil yang lebih memuaskan.