

BAB I

PENDAHULUAN

1.1 Latar Belakang

Machine Learning sudah ada sejak puluhan tahun lalu, pada tahun 1959, seorang peneliti dari IBM bernama Arthur L. Samuel pertama kali mencetuskan istilah *machine learning* dalam penelitiannya mengenai *machine learning* pada permainan *checkers* [1]. Seiring meningkatnya kemajuan teknologi dan digitalisasi di dunia ini, semakin maju pula teknologi *machine learning* yang ada. Pengambilan data menjadi semakin mudah sehingga memungkinkan untuk melatih mesin menggunakan *dataset* yang banyak. Salah satu tempat di mana terdapat banyak data adalah Twitter, cukup dengan menggunakan *snsrape*, jutaan data dari Twitter dapat diambil dan dianalisis dengan *machine learning* [2].

Salah satu kegunaan analisis sentimen adalah untuk melakukan survei, survei bisa dilakukan untuk melihat opini orang-orang terhadap baik atau buruknya topik tertentu. Contohnya opini masyarakat terhadap suatu merek, atau bahkan opini masyarakat terhadap calon presiden tertentu. Dengan mengetahui opini masyarakat terhadap calon presiden, maka hasilnya dapat dipakai sebagai semacam survei hitung cepat untuk memprediksi pemenang suatu pemilu.

1.2 Maksud dan tujuan

Laporan Tugas Akhir ini bertujuan untuk membangun model *machine learning* yang cocok untuk melakukan analisis sentimen. Model tersebut kemudian dipakai untuk analisis sentimen data Twitter terhadap suatu merek untuk

membandingkannya dengan merek lain, dan analisis sentimen terhadap calon presiden untuk melakukan prediksi hasil pemilu di Amerika Serikat dan Indonesia.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Model *machine learning* atau *deep learning* yang diuji adalah *multinomial naive Bayes*, CNN 1 dimensi, dan RNN LSTM.
2. *Dataset* yang digunakan untuk *training* dan menguji keefektifan model adalah data tulisan Twitter berbahasa Inggris yang sudah diberi label kategori positif dan negatif sebanyak 1578627 sampel dan di-download dari thinknook [3].
3. Data Twitter yang dipakai untuk prediksi pemilu presiden adalah *tweet* yang di-post dari 2 hari sebelum pemilu berlangsung sampai hari pemilu selesai dan per hari diambil 12000 sampel.
4. Data Twitter yang dipakai untuk analisis sentimen terhadap suatu merek adalah *tweet* yang di-post dari 1 Januari sampai 31 Desember pada tahun tertentu, dan per hari diambil 100 sampel.

1.4 Metodologi Penelitian

Metodologi penelitian yang penulis lakukan adalah studi literatur, pencarian *dataset*, pembersihan data, mengatur konfigurasi model yang digunakan, evaluasi model *machine learning* dan *deep learning*, melakukan prosedur-prosedur untuk optimasi model, dan mengimplementasikan model yang sudah jadi untuk menganalisis sentimen data Twitter terhadap beberapa merek dan beberapa calon presiden.

1.5 Sistematika Penulisan

Laporan Tugas Akhir ini terdiri dari lima bab dengan sistematika penulisan sebagai berikut:

1. BAB I: PENDAHULUAN

Bab ini berisi latar belakang, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penelitian.

2. BAB II: LANDASAN TEORETIS

Bab ini berisi tentang dasar-dasar teori penelitian ini yaitu *Natural Language Processing*, analisis sentimen, *machine learning*, *multinomial naive Bayes*, *deep learning*, CNN, RNN, dan Google Colab.

3. BAB III: METODE PENELITIAN

Bab ini berisi metode yang digunakan penulis dalam melakukan penelitian.

4. BAB IV: PERANCANGAN

Bab ini berisi proses yang dilakukan penulis dalam melakukan penelitian.

5. BAB V: HASIL DAN ANALISIS

Bab ini berisi konfigurasi dan akurasi dari setiap model yang digunakan, serta hasil analisis sentimen data Twitter yang dilakukan oleh model.

6. BAB VI: PENUTUP

Bab ini berisi kesimpulan mengenai hasil penelitian Tugas Akhir dan saran untuk penelitian serupa ke depannya.