

BAB I

PENDAHULUAN

1.1. Latar Belakang

Presiden terpilih Joko Widodo resmi menjabat sebagai Presiden ke-7 Republik Indonesia selama periode 2014-2019. Dalam kurun waktu 4 tahun, berbagai kebijakan dan pembangunan infrastruktur telah dibuat oleh Joko Widodo secara nasional [1]. Sebagai seorang tokoh publik, tentu tidak lepas dari pro dan kontra masyarakat terhadap kinerja maupun individu Joko Widodo. Penyampaian pendapat masyarakat kerap diunggah via media sosial seperti Facebook, Twitter, Instagram, dan lain-lain [2]. Pada Twitter, banyak pengguna yang menyampaikan pendapat lewat *tweet* yang menunjukkan orientasi positif, negatif, maupun netral kepada suatu individu.

Pada Twitter, *tweet* bisa diambil datanya menggunakan API. Kemudian, dilakukan analisis sentimen pada data untuk memperoleh informasi apakah *tweet* tersebut mengandung unsur positif, netral, atau negatif. Analisis sentimen adalah bidang studi yang menganalisis pendapat seseorang, sentiment seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang ke dalam bahasa tertulis [3]. Untuk melakukan analisis sentimen, diperlukan pendekatan menggunakan algoritma *Random Forest*. Berdasarkan dari penelitian sebelumnya yang melakukan analisis sentimen menggunakan *Random Forest*, dimana algoritma *Random Forest* berhasil mengklasifikasikan data dengan tingkat akurasi rata-rata sebesar 81% [4], dan penelitian sebelumnya yang melakukan klasifikasi loyalitas pelanggan menggunakan algoritma *Random Forest*, dimana algoritma *Random*

Forest berhasil mengklasifikasikan data dengan tingkat akurasi 83,2% [5]. Konsep dari *Random Forest* adalah menentukan *voting* terbanyak dari sekumpulan hasil yang diperoleh dari pohon-pohon (*tree*) [6].

Pembelajaran dilakukan dengan menggunakan teknik pendekatan *supervised learning* dimana sudah terdapat data yang dilatih yang terdiri dari data *input* dan data *output* berupa target yang diinginkan [7]. Dengan pembelajaran *supervised learning*, akan diperoleh model yang menggambarkan ketergantungan *input* dan *output*nya. Diharapkan model tersebut dapat digunakan untuk data *input* di luar data pembelajaran.

Oleh sebab itu, dilakukanlah penelitian analisis sentimen terhadap pak Jokowi menggunakan algoritma *Random Forest* dengan menggunakan *dataset* dari Twitter untuk memperoleh tingkat persentase pandangan masyarakat terhadap pak Jokowi.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka didapatkan rumusan masalah yaitu “bagaimana algoritma *Random Forest* dapat mengklasifikasikan sentimen *tweet* terhadap Joko Widodo?”.

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Fokus dari penelitian ini adalah menerapkan algoritma *Random Forest* untuk melakukan analisis sentimen.

2. Analisis sentimen dengan algoritma *Random Forest* menggunakan *package* yang sudah tersedia dari *python*.
3. Analisis sentimen dibagi menjadi tiga kelas sentimen yaitu positif, netral, dan negatif.
4. Data penelitian diambil dari data *tweet* di Twitter yang diunduh dalam rentang waktu 1 Agustus 2018 sampai 20 Agustus 2018 (20 hari) dengan kata kunci 'Jokowi'.
5. Dataset yang digunakan berjumlah 1755 *tweet*.
6. *Tweet* yang akan dianalisis adalah yang berbahasa Indonesia.

1.4. Tujuan Penelitian

Tujuan penelitian dalam skripsi ini adalah:

1. Mengukur tingkat akurasi algoritma *Random Forest* untuk analisis sentimen terhadap Joko Widodo.
2. Mengetahui persentase tingkat sentimen masyarakat Twitter terhadap Joko Widodo.

1.5. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat untuk :

1. Peneliti
Meningkatkan kemampuan dalam data mining dan analisis.
2. masyarakat
Memperoleh informasi persentase sentimen pengguna Twitter terhadap Joko Widodo.

1.6. Metode Penelitian

Metode yang digunakan dalam pengerjaan skripsi ini adalah eksperimental.

1.6.1 Metode Analisis

Pada tahap ini melakukan analisa terhadap masalah, alat, bahan, dan langkah kerja.

1.6.2 Metode Pengumpulan Data

Pengumpulan data dilakukan dengan cara *scrapping data tweet* dari Twitter menggunakan API dan *package Tweepy* dari *python*. Untuk label data dilakukan secara manual oleh beberapa koresponden. Teori teori mengenai analisis sentimen, *preprocessing*, *stopword*, *stemming*, TF-IDF, dan algoritma *Random Forest* dihimpun sebagai referensi materi dari jurnal-jurnal, artikel dan buku guna menunjang penelitian.

1.6.3 Metode Perancangan

Pada tahap ini dilakukan perancangan dengan menerapkan *preprocessing* data, mengubah kata tidak baku menjadi baku, memfilter kata-kata yang dianggap penting menggunakan proses *stopword*, melakukan *stemming* untuk mengubah kata ke bentuk kata dasar, memberi bobot pada kata menggunakan TF-IDF, serta menerapkan algoritma *Random Forest* untuk melatih data dan mendapatkan model.

1.6.4 Implementasi

Pada tahap ini akan dilakukan penulisan kode program dan *running* program untuk melihat hasil *test*.

1.6.5 Metode Evaluasi

Pada tahap ini akan dilakukan perhitungan akurasi dari implementasi yang sudah dilakukan. Metode yang digunakan adalah *confusion matrix*.

1.7. Sistematika Penulisan

Sistematika penulisan ini menggunakan kerangka penulisan yang tersusun sebagai berikut :

BAB I PENDAHULUAN

Memuat latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Menguraikan kajian pustaka dan dasar teori yang berkaitan dengan penelitian, seperti referensi-referensi dengan tema yang sama serta definisi yang berhubungan dengan ilmu dan permasalahan yang sedang diteliti.

BAB III METODOLOGI PENELITIAN

Pada bab ini akan diuraikan tentang analisis penelitian .

BAB IV HASIL DAN PEMBAHASAN

Membahas tentang implementasi dari hasil analisis pada BAB III, selain itu juga berisi *script* program yang dibuat.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dalam penelitian serta saran-saran yang didapat dari hasil analisis, perancangan dan implementasi.

DAFTAR PUSTAKA

