

BAB I

PENDAHULUAN

1.1 Latar Belakang

Belakangan ini, penelitian mengenai sentiment analisis pada media sosial menjadi topik bahasan oleh banyak peneliti[1] untuk melakukan penelitian. Salah satunya adalah Twitter, sebagai salah satu media sosial bertipe microblogging yang populer didunia, menjadikan twitter menjadi lahan data untuk meneliti sebuah sentimen analisis[2]. Sentimen analisis diklasifikasikan menjadi tiga jenis kategori, yaitu positif, negative dan netral[3].

Dalam membuat sebuah sentiment analisis terdapat berbagai proses *pre-processing*. Banyak penelitian yang meneliti tentang proses pre-processing dan menyatakan bahwa preprocessing memiliki peranan penting dalam kesuksesan kategorisasi teks[4][5]. Tetapi, metode pembobotan tak kalah penting dengan proses *pre-processing*[6]. Secara garis besar, metode pembobotan dibagi menjadi *Traditional Term Weighting* seperti *binary*, *TF*, dan *TF-IDF*, dan *Supervised Term Weighting* antara lain *TF-RF*, *TF-chi 2*, *TF-IG*, *TF-OR*, dan lain-lain[7].

Namun, tak semua metode pembobotan memiliki efektivitas yang sama[8]. Pada penelitian sebelumnya, ternyata metode pembobotan yang tepat berpengaruh pada hasil akhir klasifikasi teks[8]-[11]. Namun, dari beberapa penelitian tersebut menggunakan dataset selain bahasa Indonesia, mengingat bahwa sebuah kata memiliki arti atau makna yang berbeda pada bahasa yang digunakan.

Oleh karena itu, dilakukan penelitian untuk mengetahui pengaruh penggunaan metode pembobotan yang tepat dalam klasifikasi sentimen dokumen teks berbahasa Indonesia menggunakan algoritma klasifikasi *Support Vector Machine* (SVM). SVM dipilih karena fakta bahwa SVM mampu memberikan akurasi klasifikasi yang dapat diterima untuk data berdimensi tinggi ketika jumlah data latih yang tersedia terbatas[12].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka didapatkan rumusan masalah yaitu “bagaimana pengaruh penggunaan metode pembobotan TF-IDF dan TF-RF pada performa algoritma klasifikasi SVM dalam analisis sentimen dokumen teks bahasa Indonesia ?”.

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Fokus dari penelitian hanya membandingkan dua metode pembobotan, yaitu metode pembobotan *TF-Relevance Frequency*(TF-RF) dan *T F-Inverse Document Frequency* (TF-IDF).
2. Penelitian ini hanya mengukur nilai akurasi dari dua model metode pembobotan.
3. Dataset yang digunakan merupakan dataset milik peneliti sebelumnya.
4. Hanya menggunakan Algoritma *Support Vector Machine*.

1.4 Maksud dan Tujuan Penelitian

Penelitian ini bertujuan untuk membandingkan akurasi antara kedua metode pembobotan TF-RF dan TF-IDF untuk mengetahui pengaruhnya pada performa algoritma SVM.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat untuk :

1. Peneliti

Mengetahui pengaruh penggunaan metode pembobotan TF-RF dan TF-IDF dalam mengklasifikasikan sentimen teks bahasa Indonesia serta dapat dijadikan referensi pada penelitian selanjutnya yang membahas tentang metode pembobotan pada analisis sentimen.

1.6 Metode Penelitian

1. Pengumpulan Data

a. Studi Literatur

Studi literatur dilakukan untuk mempelajari penelitian-penelitian dengan tema sejenis yang pernah dilakukan sebelumnya. Penelitian-penelitian ini dijadikan referensi pada penelitian ini. Sumber yang dipakai dapat berupa buku, jurnal, paper, website, dan lain-lain. Pada penelitian ini, referensi yang dicari adalah mengenai analisis sentimen di Twitter dan metode-metode klasifikasi yang dipakai

b. Pengumpulan Data

Data input yang digunakan menggunakan dataset milik peneliti sebelumnya berupa #2019gantipresiden[13].

2. Eksperimen

a. Alat dan Bahan

Penelitian ini menggunakan bahasa Python 3.7 dengan IDE JetBrains Pycharm Community Edition 2018 ver. 2.3, dengan library Scikit-learn untuk machine learning dan library NLTK untuk praproses data, serta menggunakan Twitter API. Sedangkan bahan untuk penelitian ini adalah data tweet yang diambil dari Twitter.

b. Tahap penelitian

Pada tahap ini akan dilakukan perancangan model yang akan dibuat. Tahapan-tahapan dalam proses perancangan model ini antara lain pre-processing, pembobotan dengan metode TF-RF dan TF-IDF, perancangan model dari data latih dengan menggunakan metode Support Vector Machine.

c. Pengujian

Model yang telah dibangun kemudian diuji dan dievaluasi. Pengujian dilakukan untuk mengetahui nilai akurasi dan process time dari model yang telah dibangun. Selain menguji model, kedua model yang telah dibangun akan dibandingkan, untuk mengetahui model yang memiliki akurasi yang lebih baik. Pengujian menggunakan metode F1 Score berbasis confusion matrix dengan mencari accuracy, precision, recall dan f1 score. Kemudian model

akan divalidasi menggunakan metode K-Fold Cross Validation dengan $K=10$ untuk memeriksa apakah model melakukan klasifikasi secara overfitting.

3. Hasil dan Pembahasan

Pada tahap ini memaparkan dan membahas hasil dari eksperimen yang telah dilakukan. Pembahasan yang dilakukan mengenai akurasi dari kedua metode pembobotan.

4. Penulisan Laporan

Hasil penelitian kemudian dituliskan ke dalam sebuah laporan. Dalam penulisan laporan penelitian akan ditarik pula suatu kesimpulan berdasarkan hasil pembahasan dan pengujian, serta saran untuk penelitian berikutnya.

1.7 Sistematika Penulisan

Rincian sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini memuat paparan latar belakang penelitian, rumusan masalah, batasan masalah dalam penelitian, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab ini berisi pembahasan tentang penelitian sebelumnya yang dijadikan referensi pada penelitian ini., serta teori-teori yang digunakan dalam penelitian ini.

3. **BAB III METODOLOGI PENELITIAN**

Bab ini berisi analisis dari penelitian serta perancangan model yang akan dibuat.

4. **BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi hasil dan pembahasan dari eksperimen yang telah dilakukan.

5. **BAB V KESIMPULAN DAN SARAN**

Bab ini membahas kesimpulan dari penelitian serta saran untuk penelitian selanjutnya.

6. **DAFTAR PUSTAKA**

