

BAB I

PENDAHULUAN

1.1. Latar Belakang

Belakangan ini, penelitian tentang analisis sentimen pada media sosial banyak menarik perhatian dari para peneliti di dunia [1]. Salah satunya adalah Twitter sebagai alat *microblogging* yang memungkinkan orang bebas berekspresi. Analisis Sentimen digunakan untuk mengetahui sentimen dari sebuah komentar tentang topik tertentu yang diklasifikasikan sebagai sentimen positif, negatif, dan netral [2].

Struktur kata dari komentar di media sosial tidak beraturan dan mengandung banyak noise, hal tersebut menjadi tantangan tersendiri dalam melakukan analisis sentimen [3], [4]. Oleh karena itu peran *preprocessing* data sangat diperlukan. *Preprocessing* data merupakan proses pembersihan dan mempersiapkan data untuk dilakukan analisis [5]. Teknik *preprocessing* data yang umum dilakukan dalam klasifikasi teks yaitu *stemming* dan penghapusan *stopwords* [6], [7]. Teknik *stemming* bertujuan untuk mengubah sebuah kata menjadi kata akar dengan menghapus awalan dan imbuhan kata [8]. Sedangkan teknik penghapusan *stopwords* dilakukan untuk menghapus kata yang sering muncul dan tidak membawa makna apapun [9].

Namun, beberapa penelitian terdahulu menunjukkan hasil yang berbeda mengenai pengaruh penggunaan kedua teknik *preprocessing* tersebut pada klasifikasi teks. Beberapa penelitian menjelaskan bahwa penggunaan *stemming* maupun penghapusan *stopwords* dapat meningkatkan akurasi [10]–[13] dan beberapa lainnya mengklaim bahwa *stemming* maupun penghapusan *stopwords* tidak memberikan pengaruh signifikan terhadap akurasi [9], [14], [15]. Meskipun demikian, semua penelitian tersebut menggunakan dataset selain bahasa Indonesia yakni bahasa Arab dan bahasa Inggris, mengingat bahwa sebuah kata memiliki arti atau makna yang berbeda pada bahasa yang digunakan.

Oleh karena itu, dilakukan penelitian untuk mengetahui pengaruh penggunaan *stemming* dan penghapusan *stopwords* terhadap tingkat akurasi 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 [16].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka didapatkan rumusan masalah yaitu “bagaimana pengaruh penerapan teknik *stemming* dan penghapusan *stopwords* terhadap akurasi analisis sentimen dokumen teks bahasa Indonesia metode *Support Vector Machine*?”.

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Fokus dari penelitian hanya membandingkan dua model teknik preprocessing terhadap akurasi SVM, yaitu penggunaan *stemming* dan *stopwords* dengan yang tidak menggunakan *stemming* dan *stopwords*.
2. Penelitian ini hanya mengukur nilai akurasi dari model penggunaan *stemming* dan *stopwords* dengan yang tidak menggunakan *stemming* dan *stopwords*.
3. Teknik *preprocessing* yang dipakai berupa *case folding*, *cleaning*, penggantian kata *slang*, *stemming*, penghapusan *stopwords*, *filtering*, *tokenization*, dan pembobotan TF-IDF.
4. Dataset yang digunakan merupakan data *tweet* sentimen terhadap tagar #2019gantipresiden di Twitter.
5. Jumlah *tweet* yang dipakai pada penelitian ini berjumlah 2000 *tweet*, di mana 80% *tweet* digunakan untuk data latih dan 20% digunakan untuk data uji, yang diambil dari tanggal 1 Januari 2019 – 30 Januari 2019.
6. *Tweet* yang diambil hanya *tweet* yang berbahasa Indonesia.

1.4. Tujuan Penelitian

Tujuan penelitian dalam skripsi ini adalah meneliti pengaruh teknik *stemming* dan penghapusan *stopwords* pada tahap *preprocessing* terhadap akurasi analisis sentimen pada dokumen teks berbahasa Indonesia menggunakan algoritma SVM.

1.5. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat untuk :

1. Peneliti

Mengetahui pengaruh *stemming* dan penghapusan *stopwords* pada analisis sentimen teks bahasa Indonesia serta dapat dijadikan referensi pada penelitian selanjutnya yang membahas tentang *preprocessing* pada analisis sentimen.

1.6. Hipotesis Penelitian

Adapun hipotesis pada penelitian ini adalah :

1. H_0 : Terdapat pengaruh signifikan pada penerapan teknik *stemming* dan penghapusan *stopwords* terhadap akurasi analisis sentimen dokumen teks bahasa Indonesia metode *Support Vector Machine*.

1.7. Metode Penelitian

1. Pengumpulan Data

a. Studi Literatur

Studi literatur dilakukan untuk mempelajari penelitian-penelitian dengan tema sejenis yang pernah dilakukan sebelumnya. Penelitian-penelitian tersebut dijadikan referensi pada penelitian ini. Sumber yang dipakai dapat berupa buku, jurnal, *paper*, *proceedings*, dan lain-lain. Pada penelitian ini referensi yang dicari mengenai penelitian yang berkaitan dengan SVM, *Preprocessing*, *Stemming*, *Stopwords*, dan analisis sentimen.

b. Pengumpulan Data

Data input yang akan dipakai pada penelitian ini berupa data *tweet* sentimen terhadap tagar #2019gantipresiden dalam format *.csv.

2. Eksperimen

Pada tahap ini dilakukan perancangan dengan menerapkan *preprocessing* data, seperti melakukan *cleaning*, *tokenization*, penghapusan *stopword*, *stemming*, serta memberi bobot pada kata menggunakan TF-IDF, dan juga klasifikasi sentimen menggunakan SVM. Serta melakukan pengujian model menggunakan *confusion matrix* dan *k-fold cross validation*.

3. Hasil dan Pembahasan

Memaparkan dan membahas hasil dari eksperimen yang telah dilakukan. Pembahasan yang dilakukan mengenai akurasi model serta pengaruh penerapan teknik *stemming* dan penghapusan *stopwords*.

4. Penulisan Laporan

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

1.8. 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 topik yang sama serta definisi yang berhubungan dengan ilmu dan permasalahan yang sedang diteliti.

BAB III METODE PENELITIAN

Menguraikan alat dan bahan yang dipakai dalam penelitian, alur penelitian, dan strategi pemecahan masalah.

BAB IV HASIL DAN PEMBAHASAN

Membahas tentang implementasi dari cara kerja analisis sentimen mulai dari *pre-processing* data, TF-IDF, SVM dan hasil yang didapat dari penelitian serta pembahasannya.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dari hasil penelitian yang didapat dan juga saran untuk penelitian selanjutnya.

DAFTAR PUSTAKA