

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial memungkinkan penggunanya saling berbagi, dimana saja dan kapan saja, serta dapat mengekspresikan keinginannya. Menurut (Santoso, 2012) media sosial merupakan suatu cara bagi seseorang untuk berbagi informasi satu sama lain yaitu berupa teks, gambar, dan audio. Menurut (Permatasari, 2016) media sosial adalah suatu wadah sebagai tempat berkumpulnya beberapa orang atau lebih dengan tujuan untuk berbagi informasi, mencari teman serta berinteraksi dengan teman lainnya dengan cara *online*. Saat ini media sosial yang ramai digunakan oleh kalangan masyarakat adalah Twitter.

Twitter merupakan salah satu media sosial yang dapat digunakan pengguna mengirim serta membaca pesan teks hingga 140 karakter (Agustina, Matulatan, & Bettiza, 2012). Pada Januari tahun 2013, hingga 500 juta pengguna telah mendaftar di Twitter, 200 juta adalah pengguna aktif Bahkan para pengguna mengirimkan hingga mencapai 340 juta lebih komentar (*tweet*) dalam sehari, yang membuat twitter harus menangani jumlah permintaan pencarian lebih dari 1,6 miliar dalam sehari (Mujilahwati, 2016). Oleh karena pengguna Twitter berasal dari banyak negara, menyebabkan *tweet* dituliskan dalam berbagai bahasa.

Bahasa merupakan alat utama manusia dalam berkomunikasi dengan cara lisan maupun tulisan agar dapat saling berinteraksi satu sama lain. Menurut (Setyonegoro, 2015) fungsi bahasa yaitu sebagai cara untuk dapat menyampaikan informasi-informasi. Dengan bahasa, manusia dapat menyampaikan perasaan dan pikiran kepada orang lain (Waridah, 2015). Suatu negara dapat memiliki beragam variasi bahasa yang digunakan dalam kehidupan sehari-hari. Tetapi, tidak menutup kemungkinan jika terdapat bahasa yang hampir sama atau mirip, dikarenakan masih di dalam rumpun yang sama. Salah satu bahasa yang dapat dikatakan memiliki persamaan yaitu bahasa Malaysia dan bahasa Indonesia.

Artinya, SVM merupakan sebuah metode yang dapat diterapkan untuk mengklasifikasi teks pada *tweet* dengan tingkat keakurasian relatif dibandingkan metode lain. Penelitian (Putri, 2013) yang meneliti tentang Klasifikasi Dokumen Teks menggunakan metode SVM dengan pemilihan fitur *Chi-Square*. Pemilihan metode SVM di dalam penelitian ini adalah karena menurut (Putri, 2013) menyatakan metode SVM adalah algoritma klasifikasi terbaik jika dibandingkan dengan metode lainnya. Atas dasar beberapa penelitian tersebut yang telah melakukan penelitian menggunakan metode SVM dan menghasilkan akurasi yang baik, maka penulis mencoba melakukan penelitian dengan menggunakan metode SVM.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka akan dilakukan pengklasifikasian berbeda yaitu bagaimana “Mengklasifikasikan Bahasa yang Mirip (bahasa Indonesia dan bahasa Malaysia) dengan menggunakan Metode *Support Vector Machine*”.

1.3 Batasan Masalah

Agar pembahasan masalah dapat lebih terarah dengan baik, maka dilakukan membatasi permasalahan yang akan dibahas, yaitu sebagai berikut :

1. *Dataset* yang akan digunakan yaitu sebanyak 500 *tweet* berbahasa Indonesia dan 500 *tweet* berbahasa Malaysia. Jumlah total sebanyak 1000 data *tweet*.
2. *Tweet* yang diambil dari Twitter adalah *tweet* berbahasa Indonesia dan berbahasa Malaysia yang diambil secara *random*.
3. Klasifikasi dibagi menjadi 2 yaitu untuk kelas bahasa Indonesia dan kelas bukan bahasa Indonesia (bahasa Malaysia).

1.4 Tujuan Penelitian

Adapun tujuan dari Tugas Akhir yang dilakukan yaitu, sebagai berikut :

1. Mengklasifikasi *tweet* berbahasa Indonesia dan bahasa Malaysia dengan menerapkan Metode *Support Vector Machine* (SVM).
2. Menguji akurasi klasifikasi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.5 Sistematika Penulisan

Berikut merupakan rencana susunan sistematika penulisan laporan tugas akhir yang akan dibuat:

BAB I PENDAHULUAN

Pada bab ini akan menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penulisan dan sistematika penulisan Tugas Akhir.

BAB II LANDASAN TEORI

Pada bab ini akan menjelaskan mengenai teori-teori dari penelitian yang dilakukan penulis selama tugas akhir ini berlangsung. Teori-teori yang terdapat pada bab ini seperti, Teori tentang Twitter, bahasa, *Text Mining*, Metode *Support Vector Machine*.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang tahapan-tahapan yang dilakukan penulis selama Tugas Akhir ini berlangsung. Tahapan-tahapan yang dilakukan penulis seperti, Identifikasi masalah, Studi Literatur, menganalisa penyelesaian masalah, pengumpulan data, *preprocessing*, pembobotan dan seleksi fitur, proses SVM dan analisa klasifikasi.

BAB IV ANALISA DAN PERANCANGAN

Pada bab ini berisi penjelasan tentang bagaimana cara kerja dari klasifikasi menggunakan metode SVM untuk data *tweet* pada Twitter.

BAB V IMPLEMENTASI

Pada bab ini berisi penjelasan tentang batasan implemetasi dari Penelitian dan beberapa pengujian yang dilakukan penulis.

BAB VI PENUTUP

Pada bab ini berisi penjelasan tentang kesimpulan dan saran dari pembuatan tugas akhir