

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pilkada DKI Jakarta pada 2017 merupakan yang terburuk dalam sejarah. Menurut pengamat politik dari Lembaga Ilmu Pengetahuan Indonesia hal ini dikarenakan materi kampanye substantif di pilkada ini sangat minim dan lebih didominasi politisasi isu SARA (Ridhoi, 2017). Opini masyarakat Indonesia menanggapi hasil pilkada DKI Jakarta pun beragam. Ada yang positif, ada yang negatif, dan ada juga yang netral. Ragam opini tersebut juga dapat ditemui di media sosial, salah satunya adalah *Twitter*. Menurut data PT Bakrie Telecom (2013), pengguna *Twitter* di Indonesia sebanyak 19,5 juta orang. Jumlah tersebut menempati posisi kelima dunia (Kominfo, 2013). Jumlah pengguna *Twitter* yang besar di Indonesia merupakan sebuah indikator yang baik untuk melakukan analisis sentimen menggunakan data mining.

Analisis sentimen adalah salah satu cabang penelitian di domain *Text Mining* (Fachrurrozi & Yusliani, 2016). Secara umum, analisis sentimen berkaitan dengan riset komputasional dari opini, sentimen, dan emosi yang diekspresikan secara tekstual. Analisis sentimen bertujuan untuk mengekstrak atribut dari sebuah komentar (opini, sentimen, dan emosi) yang diekspresikan secara tekstual. Analisis sentimen dibagi menjadi dua

kategori, yaitu *Coarsed-grained sentiment analysis* dan *Fined-grained sentiment analysis*. *Coarsed-Grained sentiment analysis* melakukan proses analisis pada level dokumen. Pengklasifikasian berorientasi pada sebuah dokumen secara keseluruhan, yaitu positif, netral, dan negatif. Sedangkan *fined-grained sentiment analysis* melakukan proses analisis sebuah kalimat (Pang & Lee, 2006).

Data Mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan (Ridwan, Suyono, & Sarosa, 2013). Dalam data mining terdapat beberapa teknik, tiap teknik memiliki ciri khasnya sendiri. Pada kasus ini teknik yang akan penulis pakai adalah klasifikasi.

Klasifikasi merupakan suatu teknik yang digunakan untuk menyatakan suatu objek kedalam kategori yang sudah didefinisikan sebelumnya (Bertalya, 2009). Dalam klasifikasi terdapat beberapa algoritma dan tiap algoritma memiliki perbedaan dalam penggunaannya. Dari tiap algoritma, akurasi yang didapat tentunya berbeda, namun hasil yang didapat tidak berbeda jauh. *Naïve-Bayes*, *Support Vector Machine*, *Maximum Entropy* adalah algoritma yang umum digunakan pada analisis sentimen.

Support Vector Machine adalah algoritma yang diperkenalkan oleh Vapnik pada tahun 1992 dan

dikembangkan oleh Boser, Guyon, Vapnik. *Support Vector Machine* pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep *Support Vector Machine* berusaha menemukan *hyperlane* terbaik pada *input space*. Konsep *Support Vector Machine* dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input space* (Handoko, 2003). Keuntungan menggunakan *Support Vector Machine* adalah algoritma ini memiliki landasan teori yang dapat dianalisa dengan jelas dan tidak bersifat *black box*. Karena metode ini berbasis statistik, serta implementasi metode ini relatif mudah. Kekurangan dari metode ini adalah sulit dipakai jika data dalam jumlah besar.

Dengan berakhirnya Pilkada DKI Jakarta 2017 dengan Anies Baswedan dinyatakan sebagai pemenang tentunya menimbulkan banyak opini yang beragam, ada yang positif, ada yang negatif, dan ada juga yang netral. Hal ini menarik untuk dianalisa untuk mengetahui sentimen atau opini masyarakat terhadap hasil Pilkada DKI Jakarta 2017 tersebut.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, maka dapat dirumuskan permasalahan dalam penelitian ini, yaitu:

1. Bagaimana akurasi dari klasifikasi dataset dengan menggunakan *Support Vector Machine*.

2. Bagaimana opini masyarakat Indonesia terhadap hasil pilkada DKI Jakarta 2017 melalui *tweets* pada Twitter.

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Ruang lingkup yang dianalisa adalah opini terhadap hasil dari Pilkada DKI 2017 periode kedua dengan Anies Baswedan dan Basuki Tjahaja Purnama sebagai objeknya.
2. Sumber data yang digunakan untuk penelitian ini adalah data dari media sosial Twitter dengan rentang waktu bulan Mei 2017 sampai Oktober 2017.
3. Data Twitter yang diolah adalah data Twitter yang berbahasa Indonesia.
4. Data Twitter yang diolah adalah data yang telah melewati tahap *pre-processing*.
5. Data Twitter dikelompokkan ke dalam tiga kelas atribut, yaitu: positif, negatif, dan netral.
6. Algoritma yang digunakan dalam analisis sentimen ini adalah *Support Vector Machine*.
7. Algoritma pembanding yang digunakan adalah *Maximum Entropy*.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Mengetahui hasil akurasi klasifikasi sentimen dengan *Support Vector Machine*.

2. Mengetahui opini masyarakat Indonesia terhadap hasil pilkada DKI Jakarta 2017 melalui *tweets* pada Twitter.

1.5. Sistematika Penulisan Tugas Akhir

Laporan ini ditulis dengan sistematika sebagai berikut:

BAB 1: PENDAHULUAN

Bab 1 ini berisi tentang latar belakang yang menguraikan tentang konsep-konsep dasar pembuatan penelitian ini. Selanjutnya pada bab ini juga diuraikan tentang rumusan masalah, batasan penelitian, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB 2: TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab 2 ini berisi tentang tinjauan pustaka yang menguraikan tentang penelitian-penelitian terdahulu yang kemudian dibandingkan antara penelitian tersebut dengan penelitian yang dilaksanakan oleh peneliti. Selanjutnya pada Bab 2 ini diuraikan teori-teori tentang data mining, Analisa sentimen, Twitter, Algoritma *Support Vector Machine*, dan teori-teori lainnya yang berkaitan dengan judul penelitian.

BAB 3: METODOLOGI PENELITIAN

Pada bab 3 ini berisi metode, langkah-langkah proses analisis, klasifikasi, evaluasi dan interpretasi data.

BAB 4: HASIL DAN PEMBAHASAN

Pada bab 4 ini berisi tentang hasil dan pembahasan dari penelitian tugas akhir penulis.

BAB 5: KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dan saran penulis.

