

BAB I PENDAHULUAN

1.1. Latar Belakang

Peran media sosial seperti *Facebook* dan *Twitter*, saat ini menjadi sebuah media untuk memberikan kritik terhadap kegiatan politik [1]. *Sentiment analysis* atau yang biasa disebut *opinion mining* adalah sebuah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap sebuah entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa topik dan atribut mereka [2].

Banyak metode yang digunakan untuk melakukan analisis sentimen seperti *Naive Bayes*, *Maximum Entropy* dan *Support Vector Machine*, dan lain-lain [27]. Teknik *machine learning* menggunakan satu *set* pelatihan dan tes ditetapkan untuk klasifikasi [27]. Menurut Pang [4] metode menggunakan *machine learning* dapat memberikan hasil yang lebih baik, namun klasifier *supervised learning* membutuhkan banyak data latih yang telah diberikan label.

Tanpa diberikan data latih yang telah diberi label, *supervised learning* tidak dapat bekerja [5]. Metode *Support Vector Machine* dipilih karena memiliki kemampuan generalisasi dalam mengklasifikasikan suatu *pattern* [10].

1.2. Rumusan Masalah

Berdasarkan latar belakang, permasalahan yang dihadapi adalah menerapkan sebuah pola pendekatan yang berbeda dalam melakukan analisis sentimen pada media sosial. Dalam permasalahan yang dihadapi data yang diambil adalah data mengenai pilkada DKI.

Fokus dari penelitian tidak pada isu politik tetapi pada analisa performansi metode yang digunakan untuk melakukan analisis sentimen.

1.3. Tujuan

Tujuan dari tugas akhir ini terdapat poin-poin berikut :

1. Melakukan perbandingan nilai akurasi, presisi, dan *recall* dari *Support Vector Machine* dan *Maximum Entropy*.

2. Mencari faktor-faktor yang mempengaruhi tingkat presisi, *recall* dan akurasi dari masing-masing metode.

1.4. Batasan Masalah

1. *Dataset* yang diambil adalah *tweet* yang berhubungan dengan Pilkada DKI Jakarta khususnya masing-masing kandidat.
2. Data yang digunakan merupakan *tweet* berbahasa Indonesia.
3. Data yang digunakan bersifat *offline* (tidak terintegrasi secara *realtime* dengan *Twitter*)
4. Data *tweet* yang diambil hanya komentar teks saja, tanpa atau dengan mengabaikan URL dan juga *hashtag* yang sering muncul dalam *tweet*.
5. Tidak melakukan penanganan negasi pada setiap *tweet* yang di analisis.

1.5. Hipotesis

Berdasarkan penelitian dari [4], metode *SVM* dan *Maximum Entropy* dengan menggunakan unigram mencapai 80% dalam hal akurasi. Pada penelitian [9], dengan menggunakan metode *SVM* dan *Lexicon Base* hasil dapat mencapai sampai dengan 80%. Dengan merujuk dari hasil penelitian sebelumnya, hasil akurasi, presisi dan *recall* dapat mencapai performansi di atas 70%.