

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan zaman serta perkembangan teknologi yang mengikuti dengan pesat membuat arus informasi menjadi semakin mudah untuk diproduksi, diterima dan dibagikan kembali. Media masa mengalami peningkatan tajam dari media konvensional seperti radio, televisi maupun media cetak hingga kini memasuki era media baru atau yang biasa disebut media digital. Dalam media digital internet memainkan peran yang sangat penting. Betapa tidak, keberadaan internet membuka sekat dan menciptakan ruang yang lebih luas untuk bertukar informasi secara cepat dan fleksibel, kapanpun dan dimanapun. Keberadaan internet sebagai penopang utama penyebaran informasi pada saat ini membuat peranan internet menjadi sangat penting. Informasi yang actual dan factual kini menjadi prioritas bagi masyarakat. Masyarakat dapat dengan mudah mengakses berita-berita yang sedang trend melalui penggunaan media sosial. Salah satu contohnya adalah tragedi tewasnya seorang supporter sepak bola yang beberapa waktu lalu terjadi saat ini sedang menjadi *trending topic* di hampir seluruh media sosial tidak terkecuali media *twitter*.

Ketersediaan informasi yang banyak dan cepat dari *twitter* membuka peluang baru untuk melaksanakan sebuah penelitian yang bertujuan untuk menganalisis nilai sentiment pada *tweet* dari seorang pengguna *twitter*. Analisis sentiment merupakan sebuah pemikiran bahwa nilai dari suatu data merupakan hal penting dan dibutuhkan. Sentimen berhubungan erat dengan penilaian suatu data, wacana, konteks atau informasi. Nilai positif menandakan nilai yang baik

sedangkan nilai negatif menandakan nilai dari suatu konteks dalam teks tidak baik atau mengandung sarkas dan kritik. Sedangkan nilai netral menandakan bahwa konteks tersebut tidak masuk dalam kategori baik maupun sebaliknya.

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentiment yang terkandung dalam suatu kalimat opini [1].

Struktur kata dan gaya penulisan sebuah *tweet* menjadi tantangan tersendiri ketika melakukan analisa sentimen. Pertama, gaya bahasa yang digunakan sering menggunakan singkatan, menggunakan angka untuk mengganti huruf, penggunaan tanda baca yang tidak tepat dan juga *emoticon* dan karakter seperti "#", kurangnya struktur formal membuat analisa sentimen menjadi sulit. Kedua, nilai sentimen sebuah konteks teks dapat dipengaruhi oleh makna dari sebuah kata, pada sebuah *tweet* sendiri kalimat yang digunakan terkadang dapat menimbulkan ambiguitas terhadap pembacanya. Seperti kata "tahu", "tahu" dapat menjadi kata kerja dan juga kata benda.

Proses klasifikasi yaitu mendukung (positif) atau tidak mendukung (negatif) terhadap partai politik yang terdapat dalam *tweet*, penulis menggunakan metode *Support Vector Machine* (SVM). Metode SVM memiliki nilai *accuracy* terbaik yakni 81.10% [2]. Oleh karena itu penulis menggunakan svm untuk menentukan nilai sentimen yang diperoleh dari data twitter.

1.2 Rumusan Masalah

Perumusan masalah dalam penelitian ini adalah :

1. Bagaimana memanfaatkan algoritma *Support Vector Machine* sebagai alat untuk mengklasifikasi sifat terhadap partai politik Indonesia berdasar pada *tweet*.
2. Mencari nilai akurasi dari algoritma *Support Vector Machine* pada klasifikasi sifat politik terhadap partai politik Indonesia.

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah dipaparkan diatas, terdapat beberapa pembatasan masalah terhadap pokok pembahasan, yaitu:

1. Data yang digunakan berbahasa Indonesia, yang berupa teks dan tidak menyertai gambar yang diunduh dari Twitter selama periode keempat debat calon presiden dan wakil presiden pada tanggal 3 – 4 April 2019.
2. Proses pencarian data menggunakan API Twitter dengan kata kunci '*Jokowi*', '*Prabowo*', '*Gerindra*', '*PDIP*'.
3. Data yang di klasifikasi sejumlah 900 data dan diklasifikasi secara manual.
4. Klasifikasi terbagi atas dua kelas sentiment, positif(mendukung) dan negatif(tidak mendukung/kritik).
5. Analisis sentiment untuk data pengujian data testing dalam penelitian ini menggunakan data *tweet* dari Twitter yang diunduh secara *realtime*.
6. Analisis sentiment dengan metode *Support Vector Machine* menggunakan *tools* yang sudah tersedia dari library *python*
7. Penelitian ini menggunakan *TF – IDF* sebagai metode pembobotan.

8. Bahasa pemrograman yang digunakan adalah *python* dengan versi 3.6.5.
9. Pengujian metode dilakukan dengan menggunakan *K-Fold Cross Validation* dan *Confussion Matrix*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memanfaatkan metode SVM pada analisis sentiment terhadap data twitter.
2. Mengevaluasi kinerja dari metode Support Vector Machine dalam mengklasifikasi sentimen terhadap partai politik Indonesia.

1.5 Manfaat Penulisan

a. Bagi Partai Politik

Hasil sentimen dari data twitter dapat digunakan sebagai acuan untuk meningkatkan mutu atau sebagai alat pengukur kinerja dari partai politik dan kedua pasangan calon presiden dan wakil presiden.

b. Bagi Ilmu Pengetahuan

1. Memperoleh informasi mengenai tata cara mengolah data tekstual untuk mendapatkan informasi berupa sentimen dengan menggunakan *Support Vector Machine (SVM)*.
2. Menganalisis implementasi data mining dalam pemrograman *python*.

1.6 Metode Penelitian

Metode yang digunakan untuk mengerjakan tugas akhir ini adalah metode eksperimental yang dijabarkan dalam tahapan-tahapan berikut:

1.6.1 Metode Pengumpulan Data

1.6.1.1. Studi Pusaka

Melakukan pengumpulan bahan-bahan referensi yang dibutuhkan seperti jurnal, paper, thesis, handbook, dsb yang mendukung penelitian ini mengenai klasifikasi yang dihasilkan oleh *Support Vector Machine*

1.6.1.2. Klasifikasi Data

Melakukan klasifikasi data yang telah diunduh secara manual dan dibantu oleh teman dari peneliti.

1.6.1.3. Eksperimen

Memuat kode program yang berguna untuk mengunduh data dari Twitter dengan menggunakan API Twitter dengan parameter-parameter tertentu.

1.6.2 Metode analisis

Dalam metode ini terdapat beberapa tahapan seperti dibawah ini:

1. Pengambilan data

Data diambil dari Twitter menggunakan API Twitter yang telah didapatkan dengan mengajukan izin pada developer Twitter.

2. Data Preprocessing

Data yang telah ada kemudian diambil beberapa data untuk dikelompokkan menjadi tiga kelas yaitu, positif, negatif, netral. Data ini kemudian akan menjadi data training. Setelah itu melakukan *preprocessing* terhadap data yang telah ada, *preprocessing* sendiri memiliki beberapa tahapan yaitu:

- a. *Tweet Cleaning*
 - b. *Case folding*
 - c. Normalisasi kata
 - d. *Stopword removal*
 - e. *Tokenization*
 - f. *Stemming*
3. Fitur Pembobotan
- a. *Term Frequency (TF)* merupakan frekuensi kemunculan kata pada kalimat.
 - b. *Document Frequency (IDF)* kemunculan berapa banyak kata dalam satu kalimat.
4. Klasifikasi Support Vector Machine
- Tahapan berikutnya yaitu klasifikasi menggunakan *Support Vector Machine*, penentuan suatu kelas berdasarkan pola yang terbentuk dari hasil proses pelatihan data.
5. Menghitung akurasi hasil pengujian data
- Untuk pengujian tingkat akurasi dari algoritma *Support Vector Machine* peneliti menggunakan 10-fold sebagai parameter untuk melakukan *Cross Validation* yang diuji menggunakan *tools* pada *Python*.

1.6.3 Perancangan Sistem

1. Perancangan alur aplikasi menggunakan flowchart
2. Perancangan antarmuka pengguna

1.6.4 Implementasi

1. Implementasi sistem pada shell python
2. Manual kode program
3. Hasil implementasi

1.6.5 Pengujian

Untuk pengujian kinerja metode SVM penulis menggunakan metode *k-fold cross validation* dan *Confussion Matrix*. Metode ini digunakan untuk mengetahui rata-rata keberhasilan suatu sistem.

1.7 Sistematika Penulisan

Berikut merupakan penjabaran secara sistematis dari tiap Bab beserta ringkasan penjelasannya:

BAB I PENDAHULUAN

Pada bab ini peneliti menguraikan secara singkat mengenai latar belakang penulisan, rumusuan masalah yang diangkat, batasan masalah pada penelitian ini serta tujuan dan metode penelitian yang digunakan

BAB II LANDASAN TEORI

Pada bab ini peneliti menguraikan lebih lanjut mengenai landasan teori atau tinjauan pustaka tentang metode *Support Vector Machine*, *Sentiment Analysis*, *Preprocessing*, *TF-IDF*, *K-Fold Cross Validation* dan *Confussion Matrix*. Tinjauan pustaka dari penelitian yang lain secara singkat dibahas dan dibandingkan dengan penelitian penulis.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini penulis membuat analisa serta perancangan pada kebutuhan sistem. Serta realisasi dari metode penelitian yang penulis gunakan.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Dalam bab ini penulis mengimplementasikan *opinion mining* dan sentiment analisis dengan metode SVM. Bab ini juga membahas tentang proses pelatihan dan percobaan yang dilakukan beserta mengenai hasil yang didapatkan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan, saran-saran untuk penelitian selanjutnya, serta rangkuman dari hasil penelitian selama proses penelitian perjalan.

