

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen[1].

Perkembangan media yang begitu pesat, memunculkan banyak media online dari media berita sampai media sosial. Media sosial saja sudah begitu banyak, dari Facebook, Twitter, Path, Instagram, Google+, Tumblr, LinkedIn dan sebagainya masih banyak lagi. Media sosial sekarang ini tidak hanya digunakan sebagai sarana pertemanan, mencari teman, tapi sudah banyak digunakan untuk kegiatan lain. Promo dagangan, jual beli apa saja sampai menjadi wadah orang – orang dalam menyindir atau mengkritik tokoh politik[1].

Salah satu penelitian dalam penggunaan metode SVM yaitu analisis sentimen menggunakan *Support Vector Machine (SVM)* dengan kasus “Sentiment Analysis of Product Reviews using *Support Vector Machine Learning Algorithm*”[18], menghasilkan akurasi 89,98% dalam penggunaan metode SVM dan lebih akurat dibandingkan algoritma lainnya. Dikatakan juga keakuratan yang diperoleh akan lebih ditingkatkan dengan memasukkan lebih banyak bentuk kalimat.

Namun saat melakukan proses analisis sentimen kita akan menemukan beberapa kalimat satire. Penyampaian satire dalam bentuk teks seringkali kompleks sehingga bisa saja menyebabkan pembaca salah memahami polaritas dari suatu teks. Dimana satire merupakan gaya bahasa untuk menyatakan sindiran terhadap suatu keadaan atau seseorang. Satire biasanya disampaikan dalam bentuk ironi, sarkasme dan parodi. Satire adalah gaya bahasa yang dipakai dalam kesusastraan

untuk menyatakan sindiran, hinaan terhadap suatu keadaan atau seseorang. Sedangkan arti kata hinaan adalah merendahkan, memandang rendah, memburukkan nama baik orang, menyinggung perasaan (seperti memaki-maki, menistakan)[2]. Pada penelitian yang dilakukan oleh S. Fanissa, M. A. Fauzi, S. Adinugroho [3] penelitian ini menggunakan metode *Multinomial Naïve Bayes* yang menghasilkan akurasi tertinggi sebesar 86.6%, namun masih terdapat kekurangan pada pelabelan data yang dilakukan oleh pakar dalam kelas sentimennya serta kalimat sarkastik yang dirasa kurang sempurna. Salah satu penelitian dalam pendeteksian satire telah dilakukan oleh María del Pilar Salas-Zárate, dkk dalam klasifikasi berita dengan menggunakan metode SVM menghasilkan akurasi F-Measure tertinggi sebesar 83%[4].

Dari beberapa penelitian yang telah diuraikan, dapat diketahui bahwa untuk mengklasifikasi data besar masih memiliki kekurangan dalam hal akurasi. Dari penelitian tentang SVM dapat disimpulkan bahwa Algoritma klasifikasi *Support Vector Machine* (SVM) lebih akurat dibandingkan algoritma lainnya. Berdasarkan latar belakang yang telah diuraikan, pada penelitian ini saya akan mengimplementasikan metode SVM pada analisis sentimen pendeteksian satire terhadap tokoh politik di Indonesia.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan maka dapat dirumuskan sebuah masalah yaitu bagaimana performansi dari algoritma *Support Vector Machine* dalam melakukan analisis sentimen pendeteksian satire terhadap tokoh politik di Indonesia pada komentar youtube.

1.3. Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk mengimplementasikan metode *Support Vector Machine* dalam melakukan analisis sentimen dan pendeteksian satire. Adapun tujuan dari penelitian ini yaitu untuk mengukur performansi algoritma *Support Vector Machine* pada kasus implementasi pendeteksian satire dalam analisis sentimen terhadap tokoh politik di indonesia.

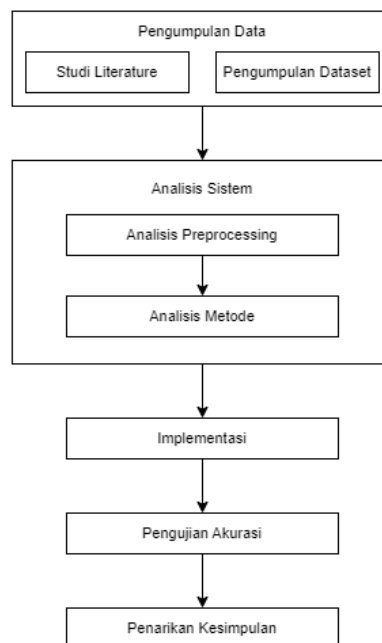
1.4. Batasan Masalah

Adapun batasan masalah pada penelitian ini yaitu sebagai berikut :

1. Kelas untuk pelabelan pada analisis sentimen terdiri dari positif, negatif dan satire.
2. Data yang diambil berupa kalimat berbahasa Indonesia.
3. Sumber Data diambil dari youtube menggunakan youtube api dengan komentar yang terdapat pada video <https://youtu.be/dl9y1A9KC-U>. Mengenai peristiwa tokoh politik saat sidang paripurna.
4. Format file untuk data latih dan data uji adalah .csv.
5. Pembobotan kata menggunakan TF-IDF.
6. Pelabelan kata dilakukan secara manual.

1.5. Metodologi Penelitian

Metodologi penelitian yang akan dilakukan pada penelitian dapat dilihat pada Gambar 1.1.



Gambar 1.1 Metodologi Penelitian

Penjelasan mengenai metode penelitian pada Gambar 1.1 , adalah sebagai berikut:

1.5.1 Metode Pengumpulan Data

Berikut ini merupakan metode pengumpulan data yang dilakukan:

A. Studi Literatur

Studi literatur merupakan proses pengumpulan data atau informasi yang berkaitan dengan penelitian yang sedang dijalankan. Pengumpulan data dilakukan dengan cara membaca buku ataupun hasil penelitian yang sudah dilakukan sebelumnya seperti jurnal, paper.

B. Pengumpulan Data Set

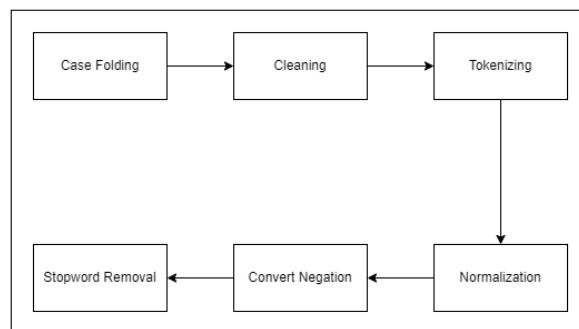
Pengumpulan data set merupakan proses pengumpulan opini netizen terhadap tokoh politik di Indonesia.

1.5.2 Metode Tahapan Analisis

Tahapan analisis pada penelitian ini adalah sebagai berikut:

A. Analisis *Preprocessing*

Analisi *Preprocessing*, merupakan tahap untuk memproses data agar data menjadi lebih terstruktur dan lebih siap untuk diolah. Langkah yang dilakukan pada tahap *preprocessing* terdiri dari *case folding*, *cleaning*, *tokenizing*, *normalization*, *convert negation*, dan *stopword removal*.



Gambar 1.2 *Preprocessing*

B. Analisis Metode

Pada tahap ini akan dilakukan klasifikasi dengan menggunakan metode SVM. Meliputi, penggunaan Pembobotan TF-IDF.

1.5.3 Implementasi

Tahapan implementasi ini adalah melakukan implementasi sistem berupa simulator, yang dimulai dari memasukkan dan membaca data komentar yang telah dilabeli secara manual, kemudian melakukan implementasi *preprocessing* yaitu:

case folding, cleaning, tokenizing, normalization, convert negation, dan stopword removal, setelah data masukkan dipraproses, dilanjut melakukan pembobotan TF-IDF memakai data yang telah dipraproses tadi, selanjutnya dilakukan pembagian data, yaitu data latih dan data uji. Lalu implementasi algoritma *support vector machine*, tahap selanjutnya yaitu pengujian.

1.5.4 Pengujian Akurasi

Dalam tahapan ini akan dilakukan pengujian akurasi dari pendeteksian satire dengan menerapkan metode *Support Vector Machine* pada analisis sentimen.

1.5.5 Penarikan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian dalam pendeteksian satire dengan menerapkan metode *Support Vector Machine* pada analisis sentimen.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab 1 menguraikan tentang latar belakang yang terdiri dari permasalahan yang berkaitan dengan penelitian, rumusan masalah, maksud dan tujuan penelitian, batasan masalah, metode penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab 2 membahas dan menjelaskan mengenai landasan teori yang mendukung penulisan skripsi.

BAB 3 ANALISIS DAN PERANCANGAN

Bab 3 berisikan tentang analisis terhadap sistem yang akan dibangun meliputi analisis masalah, analisis data masukan, dan analisis pada pemrosesan serta klasifikasi.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisikan hasil implementasi dari analisis dan pengujian terhadap sistem yang sudah dibangun.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan mengenai hasil penelitian dan saran yang diharapkan akan berguna untuk penelitian lebih lanjut pada kasus yang sama.