

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial menjadi salah satu kebutuhan manusia untuk berinteraksi dengan orang lain, mulai dari berkomunikasi, berbagi foto hingga video. Salah satu media sosial yang populer adalah Twitter [1].

Sebagian besar orang menggunakan Twitter untuk mengetahui lebih banyak informasi tentang berita terbaru atau mengikuti berita secara umum dengan mengikuti topik yang sedang tren. Twitter adalah salah satu platform media sosial paling populer yang digunakan untuk mengekspresikan pendapat, pemikiran, dan pandangan pengguna. Pengguna Twitter dapat membaca dan memposting pesan hingga 140 karakter yang disebut 'tweet'. Jumlah pengguna aktif bulanan Twitter lebih dari 313 juta pada 2016 dan hampir 500 juta tweet yang diposting dihasilkan per hari [2].

Dengan banyaknya penggunaan *social network*, dapat dimanfaatkan dengan membuat *text mining*. *Text mining* sebenarnya merupakan bagian dari *data mining* dimana proses yang dilakukan utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks menggunakan alat analisis tertentu.[3]

Analisis sentimen atau juga biasa disebut dengan *opinion mining* adalah salah satu bagian dari *text mining*. Bidang ini melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut-atributnya. [3]

Untuk melakukan analisis sentiment perlu menggunakan pendekatan dengan algoritma *supervised learning* dimana data sudah dikumpulkan dan dikumpulkan yang terdiri dari data *input* dan *output*.

Metode *random forest* merupakan pengembangan dari *decision tree* dimana setiap *decision tree* telah dilakukan proses pelatihan dengan menggunakan sampel individu. *Random forest* yang dihasilkan memiliki banyak *tree* dan setiap *tree* ditanam dengan cara yang sama. Seiring dengan bertambahnya dataset, maka *tree* juga ikut berkembang.[4] Sedangkan *Support Vector Machine* adalah kondisi dimana mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas data. *Support Vector Machine* hanya menggunakan beberapa titik data terpilih yang berkontribusi (*support vector*) untuk membentuk model yang akan digunakan dalam proses klasifikasi.[5]

Oleh karena itu, dilakukanlah penelitian perbandingan antara dua algoritma klasifikasi *Random Forest* dan *Support Vector Machine* dari segi akurasi proses dalam analisis sentiment dengan mengambil study kasus terhadap akun Twitter *pln_123*. Perbandingan ini bertujuan untuk mengetahui algoritma mana yang memiliki akurasi terbaik.

1.2 Rumusan Masalah.

Berdasarkan latar belakang masalah yang telah dijelaskan di atas, maka rumusan masalah penelitian ini adalah di antara dua algoritma, yaitu Random Forest dan Support Vector Machine, manakah algoritma yang memiliki akurasi terbaik.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah :

1. Penelitian ini hanya membandingkan dua model algoritma klasifikasi, yaitu Support Vector Machine dan Random Forest.
2. Penelitian ini hanya menggunakan pernyataan yang mengandung sentimen terhadap akun Twitter pln_123 sebagai objek pembuatan model algoritma.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membandingkan akurasi antara kedua model metode klasifikasi yang telah dibuat terkhusus untuk klasifikasi teks sentimen.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat bermanfaat untuk.

1. Meningkatkan kemampuan untuk text mining dan analisis.
2. Dapat menjadikan hasil penelitian sebagai pertimbangan gambaran untuk objek yang diteliti.

3. Hasil penelitian dapat digunakan sebagai referensi untuk para peneliti selanjutnya yang meneliti akan analisis sentimen terutama pada media sosial Twitter.

1.6 Metodologi Penelitian.

1.6.1 Metode Pengumpulan Data.

Pengumpulan data tweet dari twitter dilakukan dengan cara scrapping menggunakan API dari Twitter lalu diolah dengan *package* dari Python. Pelabelan data dilakukan secara manual oleh beberapa orang. Teori - teori mengenai analisis sentimen, preprocessing stopword, stemming, TF-IDF, dan algoritma dihimpun sebagai referensi materi dari jurnal-jurnal, artikel dan buku guna menunjang penelitian.

1.6.2 Metode Analisis.

Melakukan analisa terhadap syarat yang dibutuhkan dalam pengembangan sistem.

1.6.3 Metode Perancangan.

Pada tahap ini dilakukan perancangan dengan menerapkan *pre-processing* data, dari menghilangkan tanda baca, setelah itu mengubah dari kalimat menjadi per kata, kata-kata tersebut diubah dari tidak baku menjadi baku dengan cara tokenisasi, menyaring kata - kata yang dianggap penting, melakukan *stemming* untuk mengubah kata ke bentuk kata dasar, memberi

bobot pada kata menggunakan TF-IDF, serta menerapkan kedua algoritma untuk melatih data dan mendapatkan model yang diinginkan.

1.6.4 Metode Implementasi.

Pada tahap ini akan dilakukan penulisan kode program.

1.6.5 Metode Testing.

Pada tahap ini akan dilakukan perhitungan akurasi dari implementasi yang sudah dilakukan. Metode yang digunakan adalah *Confusion Matrix*.

1.7 Sistematika Penulisan.

Sistematika penulisan terbagi menjadi 5 bab, yaitu :

1. BAB 1 PENDAHULUAN

Bab ini berisikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

2. BAB 2 LANDASAN TEORI

Pada bab ini menguraikan teori - teori yang digunakan sebagai dasar dalam penelitian.

3. BAB 3 METODE PENELITIAN

Bab ini menjabarkan tentang analisis yang diperlukan dalam melakukan sebuah penelitian.

4. BAB 4 HASIL DAN PEMBAHASAN

Bab ini akan diuraikan lebih rinci tentang implementasi metode Random Forest dan Support Vector Machine kedalam kode program untuk membuat serta pembahasan output yang akan ditampilkan.

5. BAB 5 KESIMPULAN DAN SARAN

Pada bab terakhir menyajikan kesimpulan atas apa yang telah dikerjakan dan juga berisi saran dari penulis untuk pengembangan aplikasi ini selanjutnya.

