

BAB I

PENDAHULUAN

A. Latar Belakang

Analisis sentiment pada text atau sering di sebut sebagai perhitungan polaritas emosional dapat dikatakan sebagai studi komputasional mengenai pendapat (opini), sikap, emosi, dan penilaian orang-orang pada subjek individu, komunitas, entitas, peristiwa, problem (masalah), dan topik. Pada perkembangannya masalah yang penting dalam analisis sentimen adalah untuk melakukan identifikasi sentiment yang diekspresikan dalam teks dimana hasil ekspresi menunjukkan suatu opini yang menguntungkan (positif) atau tidak menguntungkan (negatif) terhadap subjek yang di tujukan.

Untuk saat ini media social merupakan sumber informasi yang mudah di akses dan mudah nyebarkan segala informasi yang terkadang tidak terfilter dengan baik sehingga banyak informasi yang mengandung konten nagatif menyebar. Banyak individu dan kelompok (organisasi) menggunakan media social untuk menangkap opini publik di media social untuk pengambilan keputusan mereka. Media Social Twitter merupakan salah satu media yang digunakan orang individu dan kelompok (organisasi) untuk menyebarkan opini publick, penyebaran informasi pada twitter berupa teks, dan data informasi text pada twitter dapat di ambil menggunakan Twitter API (Application Programming Interface).

Pemilihan Umum Presiden merupakan pesta demokrasi bagi seluruh rakyat Indonesia, peristiwa lima tahunan ini sering kali menjadi sorotan pemberitaan dan

menjadi topik utama pada setiap pembahasan yang terjadi di masyarakat misal melalui media televisi, radio dan juga media sosial. Untuk media sosial merupakan sumber penyebaran informasi yang bebas dan tidak terfilter dengan baik kadang informasi buruk sering tersebar melalui media sosial. Dalam pemilu presiden tahun 2019 ini mendukung dua calon kandidat dan masing-masing kandidat memiliki pendukung dan team sukses masing-masing, sering kali media sosial dijadikan alat sebagai penyampaian pendapat atau opini yang dapat mengangkat nama baik kandidat dan juga pendapat atau opini yang dapat menjatuhkan nama baik kandidat, opini-opini ini biasanya di sebar tanpa ada filter yang jelas. Diatas telah di jelaskan bahwa twitter merupakan media sosial yang masih banyak orang gunakan dalam penyebaran informasi , dan saat ini media sosial twitter merupakan sarana yang mudah dalam mendukung penyebaran informasi pemilihan umum Presiden 2019, bahkan informasi yang ada di media lain seperti televisi dan radio sering kali di sebar kembali melalui text di media sosial twitter.

Dampak dari penyebaran informasi text pada media sosial twitter , pertama dampak positif dari penyebaran informasi text tweet adalah mudahnya kita mengakses dan menyebarkan kembali / retweet kepada orang lain jadi informasi dengan sangat mudah tersebar, dan juga informasi-informasi yang penting pun dapat segera kita dapatkan, misal informasi bencana, informasi produk, dan juga masih banyak informasi-informasi yang memiliki nilai positif. Kedua dampak negative dari penyebaran informasi text tweet antara lain adalah penyebaran berita/informasi bohong dan penipuan. Dampak informasi positif dan negative sendiri dapat berpengaruh terhadap pemikiran individu atau perorangan dan kelompok yang telah mengakses informasi tersebut. Pada pemilihan

umum Presiden 2019 ini dampak penyebaran informasi sangat banyak terutama dari media sosial, karena hampir semua informasi tersebar begitu mudahnya.

Data twitter yang berhubungan dengan pemilihan umum Presiden 2019 sangat banyak dan merupakan data text mentah yang sangat luas untuk diolah sebagai informasi penelitian, oleh sebab itu untuk mengolah data twitter hal pertama yang dilakukan adalah membuat suatu algoritma untuk secara otomatis menganalisis polaritas emosi suatu teks dengan tujuan memperoleh nilai untuk setiap bagian teks. Kemudian langkah berikutnya algoritma yang sudah di bentuk diklasifikasikan berdasarkan metode Lexicon-Based, dimana metode Lexicon based menggunakan library untuk mendukung klasifikasi sentimen yaitu SentiWordNet yang merupakan suatu hasil anotasi otomatis dari seluruh synset pada database leksikal untuk bahasa inggris, dimana suatu kata dikelompokkan ke dalam set sinonim yang disebut synset. Setiap klasifikasi diimplementasikan dengan melakukan Part-Of-Speech Tagging yaitu melakukan pelabelan kelas kata pada text kalimat tweet yang selanjutnya dilakukan uji komparatif pada text, dari proses uji komparatif itu dilakukan aggregating opinion pada entity Neural Network Proses (NNP/NN) yang bertujuan supaya tidak terjadi low recall pada metode Lexicon-Based. Kumpulan dataset yang digunakan dalam penelitian ini diperoleh dan difilter dari data Twitter yang menyoroti peristiwa Pemilihan Umum Presiden Indonesia 2019, dalam kurun waktu 100 hari sebelum tanggal pemilihan umum hingga pada hari pelaksanaan pemilihan umum.

Penelitian ini mengilustrasikan pendekatan analisis sentimen untuk mengekstrak informasi public yang tersebar melalui media social twitter dimana data text tweet diolah dan di proses dengan tujuan mendapatkan nilai dari sentimen yang terkait melalui nilai

polaritas positif dan negatif untuk subjek calon kandidat Presiden pada Pemilihan Umum Presiden 2019.

1.1. Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan dari usulan penelitian ini, rumusan masalah yang akan dikaji yaitu:

Bagaimana mengolah data text yang di dapatkan melalui crawling data, melakukan filter dan scraper data hingga di dapatkan data yang bersih , sebagai dataset analisa sentiment? Bagaimana mengolah dataset twitter yang telah di filter dengan metode dan juga menggunakan algoritma yang dapat mengekstrak suatu opini pada dataset terkait dengan polaritas positif atau negative pada Pemilihan umum Presiden Indonesia 2019?

1.2. Batasan Masalah

Berdasarkan rumusan masalah yang telah disebutkan, batasan masalah dari penelitian ini adalah sebagai berikut:

1. Pengambilan data hanya pada media Social Twitter
2. Data Twitter di ambil dari kurun waktu 100 hari sebelum masa Pemilihan Umum hingga pada hari pelaksanaan Pemilihan Umum Presiden Indonesia 2019.
3. Menggunakan metode Lexicon-Based dan Support Vector Mechine (SVM) dalam memproses data text opini masyarakat di media sosial Twitter.

1.3. Keaslian Penelitian

Keaslian dari penelitian ini adalah pengambilan sample dataset twitter yang diambil dalam kurun waktu 100 hari sebelum masa Pemilihan Umum hingga hari pelaksanaan pemilihan Umum Presiden 2019. Proses pengolahan dataset yang di lakukan secara

bertahap melalui beberapa filtering, crawling, scraper dan juga menggunakan algoritma perhitungan polaritas dalam metode Lexicon-based dan Support Vector Machine (SVM).

1.4. Manfaat Penelitian

Dengan melakukan penelitian ini, diharapkan dapat memberikan manfaat kepada:

- a. Pengembang suatu sistem analisa text Pada Media Social

Menjadi referensi dalam mengidentifikasi penyebaran suatu opini yang berkembang di masyarakat melalui media social Twitter.

- b. Komisi Pemilihan Umum sebagai Badan Pemerintah Yang Bertanggung Pemilihan Umum dan juga Badan Analisa Pemilihan Umum di luar badan milik Pemerintah.

Kerangka penelitian dan proses pengolahan data text dapat digunakan sebagai acuan KPU/Badan Analisa Pemilu dalam mengolah data opini atau isu-isu yang berkembang di media social baik Twitter ataupun media social lain yang berbasis text di dalam penentuan prediksi, penanggulangan penyebaran informasi yang tidak benar, dan juga kejujuran dalam pelaksanaan Pemilihan Umum.

1.5. Tujuan Penelitian

Tujuan utama dari penelitian ini adalah mengolah data text yang didapatkan pada media social twitter melakukan filtering, crawling, scraper dan mencari nilai polaritas menggunakan metode Lexicon-based dan Support Vector Machine (SVM), dari proses pengolahan tersebut digunakan sebagai sumber analisa sentiment mengenai suatu opini atau informasi yang positif atau negatif terhadap proses Pemilihan Umum Presiden 2019 melalui media social twitter, hingga di dapatkan juga prediksi jumlah opini positif atau negative yang tersebar pada masyarakat melalui media social twitter kepada masing-masing kandidat Presiden. Tujuan lain penelitian ini adalah memberikan rekomendasi

pengklasifikasian untuk pemrosesan data twitter dengan dataset pada suatu topik yang sedang berkembang di masyarakat berdasarkan pada subjek individu, komunitas, entitas, peristiwa, masalah.

1.6. Sistematika Penulisan

Dalam penulisan laporan ini menyesuaikan dengan sistematika tata cara penulisan yang telah ditetapkan oleh Pascasarjana Universitas Atma Jaya Yogyakarta, dalam tatacara penulisan ini memiliki beberapa urutan penyajian yang benar, antara lain :

BAB 1 PENDAHULUAN

Pada Bab ini akan dipaparkan masalah umum tentang penyusunan laporan tesis meliputi latar belakang, rumusan masalah, batasan masalah, keaslian penelitian, manfaat penelitian, tujuan penelitian, serta sistematika penulisan laporan tesis.

BAB II TINJAUAN PUSTAKA

Pada Bab ini akan dimuat tinjauan pustaka terkait dengan penelitian tesis yang dilakukan, dalam tinjauan pustaka ini akan sedikit dijabarkan informasi dari masing-masing jurnal, paper atau text penelitian yang berasal dari penelitian oleh orang lain yang berkaitan dengan penelitian tesis ini.

BAB III LANDASAN TEORI

Pada Bab ini akan dimuat landasan teori terkait dengan penelitian tesis yang dilakukan baik dari landasan teori sumber data, cara penelitian, metodologi hingga proses penelitian.

BAB IV METODOLOGI PENELITIAN

Pada Bab ini akan dimuat metodologi yang digunakan dalam penelitian tesis, mencakup pengumpulan data, langkah dalam penelitian dan juga alur metode yang digunakan dalam memproses sumber data penelitian.

BAB V HASIL PENELITIAN DAN PEMBAHASAN

Pada Bab ini akan dibahas proses pelatihan dan analisa proses pelatihan tersebut. Dilanjutkan dengan pengujian dan evaluasi hasil penelitian yang didapatkan. Selain pada bagian ini juga akan dimuat kelebihan dan kekurangan model jaringan yang digunakan.

BAB VI KESIMPULAN DAN SARAN

Pada bab ini merupakan bab akhir dari seluruh rangkain laporan tesis yang didalamnya berisi suatu kesimpulan atas penelitian yang dilakukan. Selain itu pada bagian ini juga akan dimuat saran-saran dari peneliti baik berupa kritik dan gagasan untuk penelitian dimasa yang akan datang.