

BAB I PENDAHULUAN

1.1 Latar Belakang

Media sosial adalah sebuah layanan yang memfasilitasi dalam pertukaran informasi dan topik secara berkelanjutan dengan cakupan yang luas [1]. Pada saat ini perkembangan media sosial sebagai alat untuk berkomunikasi diantara masyarakat untuk menyampaikan opini. Penggunaan media sosial sangat populer dikalangan masyarakat, salah satu media sosial yang terkenal adalah *microblogging* Pengguna Internet berpindah dari *blog* atau *mailing list* menuju ke *microblogging* karena akses dan format penulisan pesannya yang mudah [1]. Contoh dari *microblogging* adalah Twitter, pengguna Twitter sangat populer di Indoneisa, ini ditunjukkan dengan peringkat pengguna Twitter. Twiiter di Indonesia pada tahun 2016 menurut website socialmemos mencapai 29 juta pengguna dengan 2.4% dari 10 juta *tweet worldwide* [1]. Selain itu Twitter hanya menyediakan 280 karakter saja sehingga *tweet* yang dituliskan pengguna akan mengungkapkan secara singkat dan inti dari maksud tulisannya.

Pengguna Twitter menuliskan opini dan pendapatnya melalui layanan Twitter dengan berbagai topik salah satunya tentang topik politik [1]. Hal ini Twitter menjadi sumber data yang sangat berpotensi dan efisien untuk menganalisis dan memprediksi Pemilihan Presiden (Pilpres). Namun data yang didapat dari twitter merupakan data yang tidak terstruktur sehingga membutuhkan sebuah *Text mining* yang akan memproses teks dari data yang didapatkan. *Text Mining* merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah data tak terstruktur yang nantinya

akan diperoleh pola-pola data, tren dan ekstraksi pengetahuan yang potensial dari data teks [2]. Dimana hasil pengolahan data dari *Text mining* adalah untuk menganalisis sentimen. Analisis sentimen (*Sentiment Analysis*) adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual [2].

Indonesia merupakan negara yang berpengalaman dalam demokrasi, seperti halnya pemilu yang pernah diselenggarakan sebanyak 11 kali pemilu anggota lembaga legislative, yaitu pada tahun 1955, 1971, 1977, 1982, 1987, 1992, 1997, 1999, 2004, 2009, dan 2014[11]. Pemilu pertama kali dilangsungkan pada tahun 1955 yang ditujukan untuk memilih anggota DPR dan konstituante, namun Pemilu 2004 merupakan pemilu pertama di mana para peserta dapat memilih langsung presiden dan wakil presiden sehingga disebut dengan Pemilihan Presiden (Pilpres), hingga saat ini Indonesia sudah tiga kali menyelenggarakan Pilpres yaitu pada tahun 2004, 2009, dan 2014. Pilpres selanjutnya yaitu akan diadakan pada 17 April 2019 dengan pasangan calon nomor urut 1 adalah Joko Widodo dan Ma'ruf Amin serta nomor urut 2 adalah Prabowo Subianto dan Salahudin Sandiaga Uno. Menuju Pilpres akan berlangsung debat kepresidenan dan berbagai acara lainnya, sehingga memicu diskusi antara masyarakat pengguna twitter, serta media sosial lainnya, diskusi yang dihasilkan dapat berupa hal positif atau negatif (sentimen), oleh karena itu sentimen dapat dijadikan sebagai referensi *polling* pengguna media sosial terhadap calon presiden.

Disisi lain *polling* dengan menggunakan media twitter sangat efisien terhadap biaya, dan juga menambah populasi sample, dengan kemungkinan

menjangkau geografis jajak pendapat dengan sangat luas. Sehingga dapat diprediksi pemenang dari sebuah Pilpres. Sentimen teks terbukti menjadi prediktor yang unggul pada tonggak tertentu dalam karya ilmiah [3].

Telah banyak penelitian – penelitian sebelumnya yang memiliki *related* arah penelitiannya, seperti pada penelitian yang dilakukan oleh Ratnawati dan Fajar dengan judul “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter” melakukan klasifikasi sentimen positif dan negatif terhadap Opini Film dengan mengimplementasikan algoritma Naïve Bayes Classifier dengan menghasilkan nilai akurasi pada *fold* kedua yaitu 90%, *precision* 92%, *Recall* 90% dan *f-measure* 90% dari total *fold* sebanyak 5. Kemudian penelitian yang dilakukan oleh Ali dan Ahmad dengan judul “*Sentiment Analysis on Twitter Data using KNN and SVM*” penelitian ini ditujukan untuk sentiment dengan data Twitter bersifat global saja, namun belum di implementasikan pada kasus nyata, seperti pemilihan president dll. Kemudian penelitian yang dilakukan oleh Winarno W. W., Dkk. dengan judul “*Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter*” penelitian ini membandingkan dua metode untuk masalah sentiment analisis yaitu Naïve Bayes dan SVM dengan hasil optimasi terbaik yaitu metode SVM, namun untuk lebih mengetahui metode yang lebih baik perlu sebuah komparasi algoritma lain seperti KNN.

Dari penelitian – penelitian sebelumnya yang sudah dijelaskan diatas masih perlu pengembangan, penelitian yaitu sebuah analisis sentimen dengan mengimplementasikan algoritma kedalam sebuah permasalahan nyata yaitu pada opini pengguna Twitter pada Pilpres 2019, kemudian menentukan sebuah

algoritma yang paling optimal, algoritma yang akan dibandingkan yaitu algoritma Support Vector Machine (SVM) dan algoritma *K-Nearest Neighbor* (KNN) sehingga didapat hasil algoritma terbaik dan perkiraan hasil Pilpres 2019 berdasarkan data Twitter.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan maka dapat dirumuskan permasalahannya yaitu bagaimana hasil perbandingan algoritma *Support Vector Machine* (SVM) dengan algoritma *K-Nearest Neighbor* terhadap proses analisis sentimen untuk prediksi hasil pilpres 2019 berdasarkan opini sosial media ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini yaitu untuk melakukan perbandingan algoritma *Support Vector Machine* (SVM) dengan algoritma *K-Nearest Neighbor* terhadap proses analisis sentiment untuk prediksi hasil pilpres 2019 berdasarkan opini sosial media.

1.4 Manfaat Penelitian

1. Mengetahui tingkat elektabilitas masyarakat terhadap calon presiden dan wakil presiden melalui media sosial Twitter.
2. Dapat digunakan untuk melakukan survey oleh elit politik.
3. Mengetahui efektifitas algoritma SVM dan K-NN.

1.5 Batasan Masalah

Berdasarkan permasalahan yang ada, agar masalah yang sedang di tinjau lebih terarah dan tercapai sasaran yang telah ditentukan, maka harus

diberikan batasan-batasan masalah. Berikut merupakan batasan masalah dari proposal ini, yaitu:

1. Objek yang digunakan pada penelitian ini adalah data *tweet* dari pengguna Twitter terkait Pemilihan Presiden (Pilpres) 2019.
2. Tweet yang digunakan yaitu *tweet* dengan berbahasa Indonesia saja.
3. Kata kunci untuk Search di Twitter menggunakan kata kunci *Jokowi* dan *Maruf Amin* untuk kandidat nomor urut satu, *Prabowo* dan *Sandiaga Uno* untuk kandidat nomor urut dua.
4. Rentang waktu pengambilan data yang diambil yaitu pada tanggal 27 Februari 2019, 28 Februari 2019, 7 April 2019, 9 April 2019, 11 April 2019 dan 12 April 2019.
5. Algoritma yang digunakan untuk penelitian adalah algoritma *Support Vector Machine* dan algoritma *K-Nearest Neighbor*.
6. Persentase prediksi diambil secara keseluruhan atau tidak berdasarkan letak geografis wilayah.
7. SDLC (*System Development Life Cycle*) yang digunakan menggunakan model *Waterfall*.
8. Pengujian menggunakan skenario pengujian yang terdapat pada poin 3.4.1.
9. Aplikasi yang dibuat hanya menggunakan *command prompt* sebagai media keluaran (*output*).

1.6 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini disusun dalam beberapa bab yang masing-masing bab menguraikan beberapa pokok pembahasan. Adapun sistematika penulisan laporan ini yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisikan tentang latar belakang permasalahan yang diambil penulis, perumusan masalah yang dihadapi, batasan masalah, tujuan, state of the art, kerangka pemikiran, metodologi penelitian serta bagaimana sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang konsep/teori apa saja yang berkaitan dengan topik yang diangkat, yang telah dibuat berdasarkan hasil penelitian dan hal-hal yang berguna dalam proses penulisan tugas akhir ini.

BAB III METODOLOGI PENELITIAN

Pada bagian ini merupakan bagian yang menjadi alur bagaimana penelitian dilakukan, mulai dari cara mengumpulkan data, mengolah atau menganalisis data dan menyimpulkan atau menetapkan simpulan dari sebuah hipotesis semua tercantum di dalam BAB 3.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang pengujian sistem secara umum maupun terperinci. Pengujian sistem secara umum akan membahas mengenai lingkungan uji coba untuk menggunakan sistem ini. Selanjutnya secara lebih terperinci

dijelaskan dalam pengujian sistem meliputi skenario pengujian beserta langkah - langkah dalam uji coba sistem untuk mengetahui aplikasi tersebut telah dapat menyelesaikan permasalahan yang dihadapi sesuai dengan yang diharapkan.

BAB V PENUTUP

Berisi tentang pernyataan singkat berupa kesimpulan dari pembahasan perangkat lunak yang dibuat secara keseluruhan dan saran untuk mengembangkan perangkat lunak yang lebih baik.

