

Bab I Pendahuluan

I.1 Latar Belakang

Tahun 2018 merupakan tahun diadakannya pemilihan kepala daerah secara serentak di Indonesia, termasuk salah satunya yaitu pemilihan kepala daerah Provinsi Sumatera Utara. Pada pemilihan kepala daerah kali ini, Provinsi Sumatera Utara memiliki dua pasangan calon pemimpin yaitu Edy Rahmayadi-Musa Rajekshah dan Djarot Saiful Hidayat-Sihar Sitorus. Sebagaimana mestinya pada saat pemilihan kepala daerah, kedua pasangan calon gubernur dan wakil gubernur tersebut akan melakukan kampanye untuk mendapatkan dukungan masyarakat. Salah satu media yang digunakan untuk kampanye yaitu melalui media sosial. Penggunaan media sosial sebagai alat kampanye sendiri dilakukan bukan tanpa alasan, berdasarkan survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2017 jumlah pengguna internet di Indonesia mencapai 143,26 juta jiwa. Angka tersebut meningkat dibandingkan pada tahun sebelumnya, yakni tahun 2016 yang tercatat mencapai 132,7 juta jiwa (Setiawan, 2018). Indonesia tercatat memiliki 130 juta pengguna media sosial aktif pada 2018. Adapun, setidaknya 49% dari total populasi penduduk memiliki dan aktif menggunakan media sosial. Dengan 120 juta pengguna di antaranya menggunakan ponsel untuk terhubung dengan akun media sosialnya (Ariyanti, 2018). Pada Maret tahun 2017, jumlah akun twitter mencapai 328 juta secara global, mewakili peningkatan 14 % dari periode yang sama pada tahun sebelumnya. Indonesia muncul sebagai negara terbesar kelima dalam hal penggunaan twitter, dengan 4,1 miliar *tweet* telah berasal dari Indonesia tahun 2016. Menurut statistik online, riset pasar dan portal intelijen bisnis Statista, Indonesia memiliki hampir 24,34 juta akun Twitter aktif per Mei tahun 2016, menempatkan negara Indonesia di tempat ketiga secara global (Herman & Mononimbar, 2017).

Twitter merupakan suatu media sosial yang memungkinkan penggunaannya untuk mengirim dan membaca pesan pendek 140 karakter yang disebut *tweets*. Selain publisitasnya, twitter dapat diakses oleh pengguna yang tidak terdaftar untuk membaca dan memantau sebagian besar *tweet*, tidak seperti facebook di mana

pengguna dapat mengontrol privasi profil mereka. Twitter juga merupakan situs mikroblogging jejaring sosial yang besar. Besarnya informasi yang disediakan oleh twitter seperti pesan *tweet* di jaringan memainkan peran penting dalam analisis data, yang membuat banyak penelitian menyelidiki dan memeriksa berbagai teknik analisis untuk memahami teknologi yang digunakan baru-baru ini (Anber, Salah, & El-Aziz, 2016). Data twitter merupakan sumber yang kaya yang dapat digunakan untuk menangkap informasi tentang topik apa pun yang dapat dibayangkan. Data ini dapat digunakan dalam berbagai kasus penggunaan seperti menemukan tren yang terkait dengan kata kunci tertentu, mengukur sentimen merek, dan mengumpulkan *feedback* tentang produk dan layanan baru. Salah satu teknik yang dapat digunakan untuk mengolah data twitter yaitu *text mining*. *Text mining* adalah teknik pemrosesan bahasa alami dan metode analisis untuk data teks untuk memperoleh informasi yang relevan. *Text mining* mendapatkan banyak perhatian pada tahun-tahun terakhir ini, karena peningkatan eksponensial dalam data teks digital dari halaman web dan layanan media sosial seperti twitter (Moujahid, 2014).

Pada penelitian ini nantinya akan dilakukan analisis sentimen terhadap calon gubernur menggunakan data dari twitter, dimana langkah awalnya yaitu dengan melakukan *crawling data* twitter yang berkaitan dengan calon gubernur dengan cara menggunakan nama calon gubernur dan slogan dari para kandidat sebagai kata kunci pencarian dengan menggunakan python. Kemudian data dari twitter tersebut akan di ekstrak untuk diambil data *tweetnya* saja yang berupa file teks. File teks tersebut nantinya akan diolah dengan metode analisis sentimen *lexicon based* dimana hasil akhirnya berupa klasifikasi tweet berupa positif atau negatif. Hasil klasifikasi positif dari *tweet* tersebut diharapkan dapat dijadikan sebagai perbandingan apakah hasil dari penelitian ini sesuai dengan hasil pemilihan gubernur provinsi Sumatera Utara.

I.2 Perumusan Masalah

Berdasarkan latar belakang, maka merumuskan masalah-masalah sebagai berikut :

- a. Mengambil data dari twitter melalui proses *data crawling* sehingga data yang sudah diambil dapat digunakan sebagai bahan penelitian analisis sentimen *lexicon based*.
- b. Mendapatkan hasil analisis sentimen terhadap para kandidat Gubernur dan Wakil Gubernur pada pemilihan gubernur Provinsi Sumatera Utara tahun 2018.
- c. Hasil penelitian yang telah didapatkan dapat dibandingkan dengan hasil perhitungan resmi oleh Komisi Pemilihan Umum (KPU).
- d. Mendapatkan persentase *accuracy*, *precision*, *recall* dan *classification error* dari metode yang digunakan pada penelitian ini.

I.3 Tujuan Penelitian

Berdasar latar belakang dan rumusan masalah maka tujuan yang ingin dicapai dari penelitian ini adalah :

- a. Menganalisis tahapan proses data crawling dari twitter sehingga data yang didapatkan dapat dimanfaatkan untuk analisis sentimen menggunakan metode *lexicon based*.
- b. Meneliti faktor yang berpengaruh terhadap sentimen positif dan sentimen negatif dari kandidat gubernur dan wakil gubernur Provinsi Sumatera Utara.
- c. Menginvestigasikan pengaruh karakter data terhadap hasil analisis sentimen dan hasil perhitungan resmi Komisi Pemilihan Umum (KPU).
- d. Mengevaluasi penggunaan metode analisis sentimen *lexicon based* dengan untuk mengetahui persentase *accuracy*, *precision*, *recall* dan *classification error* menggunakan dengan menggunakan *Support Vector Machine (SVM)*.

I.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam hal :

1. Sebagai syarat kelulusan program studi Strata-1 Sistem Informasi Universitas Telkom.
2. Mengetahui sentimen pengguna twitter terhadap para kandidat Gubernur dan Wakil Gubernur Provinsi Sumatera Utara tahun 2018.
3. Pengembangan penelitian yang terkait dengan analisis sentimen dengan memanfaatkan data dari twitter.

I.5 Ruang Lingkup

Penulisan penelitian ini dibatasi pada hal-hal sebagai berikut :

1. Dataset yang didapat dari *application programming interface (API)* twitter diambil pada rentang waktu dari tanggal 24 Maret 2018 hingga 29 Juni 2018.
2. Data yang didapatkan berdasarkan dari kata kunci yang terkait dengan kandidat yaitu nama kandidat gubernur, nama kandidat wakil gubernur dan slogan yang terkait dengan kandidat calon Gubernur dan Wakil Gubernur Sumatera Utara.
3. Data twitter yang digunakan untuk analisis hanyalah data *tweet* yang berupa data teks tanpa melakukan penyaringan tweet berdasarkan lokasi pengguna twitter.
4. Tidak memperhitungkan proses dalam pengembangan kamus yang digunakan untuk analisis sentimen.

I.6 Sistematika Penulisan

Agar mempermudah pembaca, secara garis besar isi laporan ini dibagi dalam beberapa bab secara terurut sebagai berikut :

1. Bab I : Pendahuluan

Pada bab I, penelitian ini membahas tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup dan sistematika penulisan.

2. Bab II : Tinjauan Pustaka

Pada bab II akan membahas mengenai landasan teori yang dibutuhkan untuk mendukung penelitian ini.

3. Bab III : Metodologi Penelitian

Pada bab III akan menjelaskan langkah-langkah yang akan dilakukan pada penelitian ini melalui sub bab model konseptual dan sub bab sistematika penelitian.

4. Bab IV : Hasil dan Analisis

Pada bab IV akan menjelaskan mengenai detail implementasi yang dilakukan pada penelitian ini, yaitu terdiri dari tahap pengambilan data, tahap *preparing corpus*, tahap *data preprocessing*, tahap *splitting corpus*, tahap analisis sentimen dan perbandingan dengan hasil resmi Komisi Pemilihan Umum (KPU) dan akan dilakukan pengujian metode dengan menggunakan *Support Vector Machine (SVM)*.

5. Bab V : Kesimpulan dan Saran

Pada bab V berisi kesimpulan yang didapatkan dari penelitian ini serta saran dari penulis untuk pengembangan penelitian lain yang terkait dengan penelitian ini.