

BAB I

PENDAHULUAN

1.1 Latar Belakang

Manusia merupakan makhluk sosial, bersosialisasi merupakan hal yang sangat dibutuhkan oleh manusia. Bersosialisasi memiliki tujuan untuk dapat berbagi informasi dan saling tolong menolong terhadap sesama manusia. Salah satu syarat untuk melakukan sosialisasi adalah harus ada manusia lainnya yang berperan sebagai sarana untuk bersosialisasi atau menyampaikan informasi. Bersosialisasi dapat meningkatkan serta menyempurnakan kesejahteraan hidupnya demi kelangsungan hidup dan memenuhi kebutuhan untuk menyesuaikan diri dengan lingkungannya.

Teknologi semakin berkembang saat ini, akhir-akhir ini masyarakat dipermudah untuk bersosialisasi dengan menggunakan teknologi. Teknologi komunikasi yang sering digunakan saat ini adalah internet. Media sosial yang tersedia saat ini di internet seperti *Facebook*, *Twitter*, *Instagram* dan lainnya. Media Sosial tersebut biasa digunakan masyarakat untuk menyampaikan sebuah opini, berbagi informasi dan kegiatan lainnya. Pengguna media sosial dapat mencari data atau informasi di media sosial secara manual di media sosial, yaitu dengan cara mencari melalui fitur pencarian yang tersedia pada media sosial tersebut.

Selain pencarian data secara manual di media sosial, pengguna dapat mencari atau menggali data dalam bentuk teks di media sosial dengan menggunakan metode *Data Mining* atau *Text Mining*. *Text Mining* merupakan metode klasifikasi bentuk variasi dari *data mining* berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar[1]. *Data mining* dan *Text Mining* merupakan metode yang memiliki proses yang sama, namun data yang dimasukkan pada *Text Mining* berbeda dengan data yang dimasukkan pada *Data Mining*. Data yang dimasukkan pada *Text Mining* merupakan data-data yang tidak terstruktur, seperti dokumen word, PDF, XML dan sejenisnya. Salah satu penerapan *Text Mining* adalah pada penelitian analisis sentimen.

Analisis sentimen adalah bidang studi yang menganalisis pendapat seseorang, sentimen seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang dalam bentuk teks tertulis[2]. Data yang dianalisis biasanya bersumber dari komentar masyarakat di sosial media. Penelitian ini mengambil data yang bersumber dari komentar-komentar masyarakat di Twitter. Twitter merupakan salah satu media sosial yang biasa digunakan oleh penggunanya untuk berbagi informasi atau beropini. Informasi yang terdapat di Twitter biasanya disebut dengan tweet, dengan batas penulisan maksimal 280 karakter. Berdasarkan artikel yang ditulis oleh Robit Mikrojul Huda mengenai pengguna Twitter, mengatakan bahwa Indonesia menduduki urutan ketiga dalam data negara pengguna Twitter dengan jumlah penggunanya mencapai 24,3 juta[3]. Tingginya jumlah pengguna Twitter berbanding lurus dengan banyaknya informasi yang terdapat di Twitter. peneliti tidak memilih media sosial Instagram sebagai studi kasus dikarenakan Instagram merupakan media sosial yang berbasis gambar dan video dalam menyampaikan sebuah informasi.

Analisis sentimen memiliki beberapa permasalahan yang menyebabkan hasilnya kurang maksimal. Masalah-masalah tersebut biasanya berupa penulisan yang tidak banyak dimengerti masyarakat atau kesalahan dalam penulisan seperti penggunaan bahasa *modern*, bahasa tidak baku dan tulisan yang disingkat. Penggunaan penulisan seperti itu pada komentar-komentar di Twitter menyebabkan hasil akurasi pada analisis sentimen menjadi kurang baik. Hal tersebut disebabkan karena pada saat melakukan proses *Preprocessing*, program hanya bisa membaca suatu kata dengan penulisan yang benar sesuai dengan kamus yang sudah disediakan. Apabila kata yang dimasukkan tidak sesuai dengan kamus yang ada, maka program tidak akan membacanya.

Rendahnya nilai akurasi dapat dioptimalisasi dengan perbaikan pada tahap *Preprocessing* dan penambahan proses perluasan makna atau *Semantic Expansion*. *Semantic Expansion* adalah proses mereformulasikan kembali *query* awal dengan melakukan penambahan beberapa *term* atau kata yang memiliki makna sama[4]. Perluasan *query* biasa digunakan pada Sistem Temu Kembali Informasi, namun dapat digunakan pula untuk melakukan analisis sentimen dengan tujuan yang sama.

Perluasan *query* atau *Query Expansion* berbeda dengan *Semantic Expansion*. Kata yang diperluas pada proses *Query Expansion* merupakan kata yang relevan dengan kata awalnya. Sedangkan kata yang diperluas pada proses *Semantic Expansion* merupakan kata yang memiliki makna sama dengan kata awalnya.

Pada penelitian ini, *Semantic Expansion* digunakan setelah tahap *preprocessing* lalu selanjutnya menerapkan metode *Naive Bayes Classifier* untuk pengklasifikasian. *Naive bayes classifier* adalah pendekatan yang mengacu pada teorema Bayes yang merupakan prinsip peluang statistika dengan mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru[5]. Metode tersebut mengklasifikasikan sebuah data menggunakan probabilitas dan statistik. Keuntungan dari penggunaan metode tersebut adalah karena metode tersebut dapat menangani kuantitatif dan data diskrit, algoritmanya mudah untuk dipahami dan lebih cepat dalam melakukan perhitungan.

Bersumber pada penelitian terdahulu dengan judul *Sentiment Analysis* pada Review Aplikasi *Mobile* menggunakan Metode *Naive Bayes* dan *Query Expansion* yang ditulis oleh Firmansyah, dkk pada tahun 2016, melakukan penelitian tentang analisa sentimen terhadap komentar-komentar masyarakat mengenai sebuah aplikasi *mobile* di *Play Store*[6]. Metode yang digunakan pada penelitian tersebut adalah *Naive Bayes* dengan *Query Expansion*. *Naive Bayes* digunakan untuk melakukan klasifikasi sentimen terhadap komentar-komentar masyarakat, sedangkan *Query Expansion* digunakan untuk memperbaiki penggunaan kata sifat yang berbeda namun memiliki makna yang sama dengan menambahkan sinonim pada kata sifat yang diambil dari API Kateglo. Berdasarkan hasil pengujiannya, penambahan metode *Query Expansion* pada penelitian tersebut menghasilkan nilai akurasi yang tinggi yaitu 98% dan mendapatkan nilai akurasi 95% jika tanpa menggunakan metode *Query Expansion*. Pada penelitian tersebut terdapat permasalahan yang tidak terpecahkan yaitu pada tahap proses *Preprocessing* mengenai normalisasi kata, negasi kata dan *Stemming*. Tahap *Preprocessing* yang dilakukan pada penelitian tersebut masih belum maksimal, sehingga sangat mempengaruhi hasil dari proses klasifikasi.

Berdasarkan deskripsi diatas maka penelitian dengan judul “**Optimasi Nilai Akurasi pada Analisis Sentimen Menggunakan Metode *Naïve Bayes Classifier* dan *Semantic Expansion***” dilakukan. peneliti melakukan penelitian tersebut dengan studi kasus yang berbeda untuk mengetahui pengaruh penambahan metode *Semantic Expansion* dalam Analisis Sentimen. Diharapkan dengan adanya penambahan metode tersebut dan perbaikan pada tahap *Preprocessing* dapat meningkatkan akurasi dalam hasil penelitian.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat ditentukan sebuah rumusan masalah yang dikaji adalah “Bagaimana hasil dari analisis sentimen menggunakan metode *Naïve Bayes Classifier* dengan penambahan metode *Semantic Expansion* dan tanpa penambahan metode *Semantic Expansion*?”.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah menganalisis nilai akurasi yang dihasilkan dari kombinasi antara metode *Naïve Bayes Classifier* dan *Semantic Expansion* pada analisis sentimen.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah :

1. Data yang dicari hanya bersumber dari komentar-komentar di twitter.
2. Data bersumber dari tweet di Twitter mengenai Ridwan Kamil saat menjabat sebagai Wali Kota Bandung pada tahun 2013 hingga tahun 2018.
3. Penelitian ini hanya menghasilkan model klasifikasi.
4. Peneliti hanya menggunakan dua kelas sentimen dalam melakukan klasifikasi data yaitu kelas sentimen positif dan kelas sentimen negatif.
5. Tweet yang memiliki kalimat sama dengan tweet lainnya(Retweet) dihapus.
6. Program yang dibuat menggunakan bahasa Python.

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini antara lain:

1. Penggunaan metode Semantic Expansion dapat meningkatkan hasil akurasi.
2. Kombinasi antara Metode Naïve Bayes Classifier dan Semantic Expansion dapat dijadikan sebagai metode dalam mengklasifikasi teks dengan hasil akurasi yang baik.
3. Penelitian ini dapat dijadikan studi literatur untuk penelitian selanjutnya.