

1. Pendahuluan

1.1 Latar Belakang

Twitter adalah sebuah media sosial yang memberikan kemudahan bagi penggunaannya untuk membagi dan mengakses informasi. Perkembangan jumlah pengguna *twitter* mengalami peningkatan setiap tahunnya. Berdasarkan *ranking* yang dibuat oleh situs *eBizMBA*, *twitter* berada pada urutan kedua setelah *Facebook* di dalam *Top 15 Most Popular Social Networking Sites*. Jumlah penggunaannya menembus angka 250 juta pada periode November 2013 [19]. Dan menurut *ranking* yang dilakukan situs *eMarketer*, Indonesia berada pada urutan pertama tingkat penambahan jumlah akun *twitter*. Data ini diperoleh dalam setahun dengan peningkatan sebesar 44,20% [20]. Bertambahnya jumlah pengguna berdasarkan data tersebut, membuat informasi yang tersebar pun semakin meningkat. Informasi yang tersebar di *twitter* antara lain adalah berita terbaru, ide, dan bahkan opini.

Menurut *Webster's New Collegiate Dictionary*, opini merupakan suatu pandangan, keputusan atau sebuah taksiran yang terbentuk di dalam pikiran mengenai suatu persoalan tertentu. Opini banyak berpengaruh di dalam kehidupan sosial. Pengaruhnya di antara lain adalah di dalam penjualan produk, perubahan kebijakan di dalam sistem pemerintahan, dan bahkan suara rakyat pada pemilihan umum [1]. Dengan kumpulan opini, organisasi atau individu dapat mengetahui keputusan yang tepat dalam menangani sesuatu hal. Namun, opini bukan hanya berisi hal – hal yang positif, banyak informasi dari opini yang mengandung hal negatif. Untuk itu, diperlukan sebuah metode untuk mengambil opini dan mengelompokkannya berdasarkan sentimen yang sesuai.

Sentiment Analysis adalah sebuah teknik perkembangan dari *Natural Language Processing* dengan area riset dari klasifikasi level dokumen [2] sampai pembelajaran polaritas kata dan frase [3]. Teknik ini memiliki tujuan untuk mencari opini dari seseorang tentang sesuatu hal yang spesifik. Berbagai penelitian telah dilakukan untuk menganalisis opini, khususnya pada media sosial *twitter*, diantaranya adalah “Twitter as a Corpus for *Sentiment Analysis* and *Opinion Mining*” oleh Alexander Pak dan Patrick Paroubek. Penelitian tersebut memiliki tujuan untuk membangun *sentiment classifier* yang dapat menentukan sentimen positif, negatif, dan netral pada sebuah dokumen [6].

Pada tugas akhir ini, penelitian berfokus pada *sentiment analysis* terhadap isu sosial Kota Bandung. Penelitian ini diperlukan karena *sentiment analysis* masih didominasi dengan studi kasus *review* produk. Selain itu isu sosial memiliki struktur opini yang berbeda dengan produk atau sebuah layanan [4]. Dan kebijakan baru dari walikota yang terpilih menciptakan isu sosial yang baru juga. Penelitian ini bertujuan untuk menemukan kelas *sentiment* yang tepat dari setiap opini yang disebar pada *twitter*. Di dalam sebuah *tweets* bisa terdapat satu atau lebih kata yang menggambarkan suatu opini. Namun, kata – kata tersebut terkadang memiliki hubungan dengan kata lainnya. Di dalam sebuah isu pada satu data *tweets* dapat mengandung beberapa kata opini dengan sentimen yang berbeda-beda.

Berdasarkan masalah tersebut, dibutuhkan sebuah metode yang dapat menangani permasalahan *multi-opini*. Masalah tersebut dideskripsikan dengan metode *Holistic Lexicon-Based*. Metode ini dapat mengidentifikasi sentimen dari setiap *opinion words* yang terdapat pada data *tweet* dan dapat menangani permasalahan *multi-opini* di dalam suatu data. Metode ini merupakan peningkatan dari metode *Lexicon-Based* yang tidak bisa menangani permasalahan *multi-opini*. Di dalam penanganan masalah *multi-opini*, metode ini mengumpulkan seluruh sentimen dari kata opini berdasarkan jarak antara kata opini dengan fiturnya. Sehingga akhirnya dapat dipakai untuk menentukan kelas opini dari setiap data [5].

1.2 Perumusan Masalah

Berdasarkan pada latar belakang di atas, permasalahan yang akan diuraikan dan diteliti adalah :

1. Bagaimana cara mengidentifikasi data opini dari *tweets* dan menentukan sentimen kata opini dari setiap data menggunakan *opinion lexicon*?
2. Bagaimana cara menerapkan metode *Holistic Lexicon-Based* pada *sentiment analysis* terhadap data *tweets*?
3. Bagaimana akurasi klasifikasi sentimen berdasarkan metode *Holistic Lexicon-Based* pada *sentiment analysis* terhadap data *tweets*?

1.3 Batasan Masalah

Adapun batasan masalah untuk tugas akhir ini adalah :

1. Data yang dianalisis adalah *tweets* berbahasa Indonesia yang didapat dari situs jejaring sosial <http://www.twitter.com>
2. Klasifikasi opini hanya untuk sentimen opini positif dan negatif.
3. Data akan dianalisis secara *offline*, yaitu sistem tidak terhubung ke dalam jaringan internet.
4. Suatu *tweet* dinyatakan data opini jika di dalamnya terdapat minimal satu *opinion word*.
5. Metode ekstraksi kata opini dan klasifikasi yang digunakan adalah *Holistic Lecicon-Based*.
6. Penentuan *query* pencarian adalah *tweets* yang memiliki *mention* ke :
 - @infobdg
 - @infobandung
 - @pemkotbandung
 - @ridwankamil
 - @distarcipbdg
 - @pdamtirtawening
 - @dinsos_bdg
 - @dishub_kotabdg
 - @dbmpkotabdg
 - @pdkebersihan
 - @diskamtam

- @pjudbmbpdg
- #suarabdg.

1.4 Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut :

1. Menganalisis dan mengidentifikasi kata opini dari dataset *tweets* menggunakan *opinion lexicon*.
2. Menerapkan metode *Holistic Lexicon-Based* pada *sentiment analysis* terhadap data *tweets*.
3. Mengukur akurasi klasifikasi sentimen berdasarkan metode *Holistic Lexicon-Based* pada *sentiment analysis* terhadap data *tweets*.

1.5 Metodologi Penyelesaian Masalah

Penyelesaian masalah dilakukan dalam beberapa tahap, secara garis besar sebagai berikut :

1. Studi Literatur

Mempelajari literatur – literatur yang relevan untuk mempelajari dan memahami permasalahan *sentiment analysis*, khususnya konsep dan langkah – langkah metode *Holistic Lexicon-Based*.

2. Pengumpulan Data

Pengumpulan data *tweets* berdasarkan *query* pencarian mengenai isu sosial Kota Bandung. *Tweets* yang diambil mengandung *mention* kepada user @infobdg, @infobandung, @pemkotbandung, @ridwankamil, @distarcipbdg, @pdamtirtawening, @dinsos_bdg, @dishub_kotabdg, @dbmpkotabdg, @pdkebersihan, @diskamtam, @pjudbmbpdg dan *hashtag* #suarabdg. Data yang diambil hanya *tweets* berbahasa Indonesia.

3. Perancangan Sistem

Pada tahap ini dilakukan perancangan sistem untuk analisis sentimen dan klasifikasi *tweets* ke dalam jenis sentimen dengan menggunakan metode *Holistic Lexicon-Based*.

4. Implementasi

Pada tahap ini dilakukan implementasi berdasarkan rancangan sistem yang telah dibuat sebelumnya menjadi suatu sistem klasifikasi opini berdasarkan isu sosial.

5. Analisis dan Pengujian

Pada tahap ini dilakukan analisis dan pengujian terhadap metode *Holistic Lexicon-Based* terhadap data *tweets* mengenai isu sosial Kota Bandung

6. Pengambilan Kesimpulan dan Pembuatan Laporan

Pada tahap ini dilakukan pengambilan kesimpulan berdasarkan analisis dan pengujian yang telah dilakukan serta pembuatan laporan tugas akhir untuk dokumentasi setiap proses kegiatan penelitian.