

BAB I

PENDAHULUAN

1.1. Latar Belakang

Twitter merupakan sebuah situs *microblogging* yang populer dibandingkan dengan situs *microblogging* lainnya. Hal ini terlihat dari jumlah pengguna Twitter yang mencapai 105 juta pada April 2010 dengan jumlah *posting* 55 juta *tweet* per hari (Jackoway, dkk., 2011). Twitter merupakan *social media* yang digunakan oleh banyak orang untuk dapat terhubung dengan orang-orang disekelilingnya dan seluruh dunia melalui komputer dan perangkat *mobile*.

Twitter sebagai salah satu situs *microblogging* mampu mengirimkan pesan pendek (140 karakter) tentang apa yang mereka lakukan, apa yang ada di sekeliling mereka, kejadian yang sedang terjadi, dan hal lainnya yang dapat dilihat oleh semua orang. Pesan tersebut biasanya disebut dengan *tweet*. Twitter dikategorikan sebagai *microblogging service*. *Microblogging* merupakan sebuah bentuk blog dimana penggunaanya dapat mengirimkan sebuah pesan teks (*status update*) yang singkat.

Indonesia merupakan negara ketiga penghasil *tweet* terbesar dengan jumlah enam juta *tweet* per hari (Wibisono, 2012). Ini bisa menjadi potensi informasi yang sangat besar untuk dimanfaatkan. Twitter pun dapat dimanfaatkan untuk mendeteksi adanya bencana karena sifatnya yang *real-time* dan *on-location update*. Misalnya, ketika terjadi bencana banjir, maka orang-orang akan mengirim sebuah pesan melalui Twitter tentang apa yang sedang terjadi terjadi di sekelilingnya.

Twitter menyediakan sumber informasi yang begitu besar yang tidak mudah didapatkan di tempat lain. Ada banyak informasi dalam Twitter yang sifatnya *up-to-date* dan tentu sangat bermanfaat untuk sebagian orang. Namun, ada juga *tweet*

yang sama sekali tidak menarik bagi sebagian orang. Dibutuhkan sebuah cara untuk menentukan informasi yang *reliable* dengan adanya kejadian di Twitter.



Dalam hal ini, informasi atau kejadian yang terdapat pada Twitter dapat diidentifikasi, sehingga nantinya informasi tersebut akan bermanfaat untuk dikonsumsi. Dengan adanya cara tersebut dapat membantu menemukan kejadian yang sudah terjadi, sedang terjadi, atau mungkin yang akan terjadi.

Twitter menyediakan banyak kumpulan *tweet* dengan banyak variasi penulisan kata didalamnya. Banyak sekali kata-kata yang jarang (*noise*) ditemui dalam bahasa Indonesia baku seperti singkatan kata, bahasa gaul, dan karakter-karakter yang tidak bermakna (Rangrej, dkk. 2011). Hal tersebut menjadi kendala dalam penentuan *event detection* sehingga harus dilakukan pemrosesan awal terlebih dahulu untuk mengurangi *noise* sehingga informasi yang didapat dari kumpulan *tweet* merupakan informasi yang bermanfaat.

Penelitian ini melakukan analisis *tweet* dengan keyword "banjir" untuk diproses lebih lanjut sehingga menghasilkan informasi yang dapat bermanfaat dan digunakan sebaik-baiknya, misalkan mengetahui kejadian banjir yang umum terjadi pada rentang waktu tertentu. Metode yang dipakai dalam penelitian ini, yaitu *density-based clustering* dengan algoritma DBSCAN. *Density-based clustering* akan melakukan *clustering* berdasarkan tingkat kepadatan dari suatu kumpulan data. Karakteristik dari *density-based clustering* ini sangat cocok digunakan untuk menemukan adanya *event* dari suatu kumpulan data tanpa terbatas berapa *event* yang ingin dideteksi. Selain itu, algoritma DBSCAN sangat cocok digunakan pada data yang memiliki banyak *noise* (Gaonkar, 2013). Banjir ini dijadikan sebagai objek penelitian karena merupakan salah satu bencana alam terbesar di Indonesia selain gempa bumi. Tahun 2013, Badan Nasional Penanggulangan Bencana (BNPB) mencatat banjir sebagai bencana yang paling sering terjadi di Indonesia dengan jumlah sebanyak 4261 bencana.

Dari penjelasan di atas, penulis tertarik untuk melakukan penelitian tentang *event detection* pada *microblogging* Twitter.

1.2. Rumusan Masalah

Merujuk dari latar belakang di atas, ada beberapa permasalahan yang timbul dalam melakukan identifikasi bencana:

- a. Bagaimana algoritma DBSCAN dapat membantu mendeteksi adanya *event* melalui Twitter?
- b. Bagaimana melakukan pengambilan informasi dari setiap *event* berdasarkan kumpulan *tweet* yang terbentuk sehingga dapat menyimpulkan dimana lokasi adanya *event*?

1.3. Batasan Masalah

Untuk memfokuskan penelitian yang akan dilakukan ada beberapa batasan masalah, yaitu sebagai berikut.

- a. Penulis hanya menggunakan Twitter sebagai sumber data.
- b. Bencana yang dideteksi dalam penelitian ini difokuskan pada bencana banjir.
- c. Penulis tidak mendeteksi kejadian berdasarkan kesamaan kata yang bersinonim dengan banjir, misalkan genangan air atau aliran air.
- d. Data *tweet* yang diambil tidak *real-time*, melainkan hasil pengumpulan dalam jangka waktu tertentu (Januari – Februari 2012 dan Januari 2013).
- e. Data lokasi banjir tidak diambil dari data *geotagging* di Twitter, melainkan diambil dari hasil ekstraksi informasi yang ada pada *tweet*.
- f. Penulis menggunakan data kumpulan sinonim yang dibuat secara manual berdasarkan data *tweet* yang ada berdasarkan asumsi yang sering muncul.

1.4. Tujuan Penelitian

Tujuan dari penelitian yang akan dilakukan yaitu.

- a. Dapat mendeteksi adanya banjir berdasarkan *tweet* dengan metode *density-based clustering* menggunakan algoritma DBSCAN.
- b. Dapat mengambil informasi penting dari setiap *event* berdasarkan kumpulan *tweet* yang terbentuk dengan menggunakan teknik *Named Entity Recognition*.

1.5. Definisi Operasional

Di dalam penelitian ini ada beberapa istilah yang umum digunakan. Diantaranya sebagai berikut.

1. *Eps* adalah *input* parameter untuk algoritma DBSCAN yang digunakan untuk menentukan radius dari suatu *cluster*.
2. *MinPts* adalah *input* parameter untuk algoritma DBSCAN yang digunakan untuk menentukan nilai *minimum point* suatu *cluster* di dalam *radius eps*.
3. *Noise* adalah data yang tidak masuk ke dalam *cluster* manapun. *Noise* dalam algoritma DBSCAN merupakan titik yang memiliki kepadatan rendah sehingga jarak *point* tersebut terhadap *point* lain jauh.
4. *Tweet* adalah pesan yang dituliskan pada Twitter.
5. *Clustering* adalah proses pengelompokkan kumpulan data ke dalam kelompok-kelompok yang memiliki kemiripan.
6. *Event Detection* adalah proses identifikasi adanya suatu *event*.
7. *Named Entity Recognition* adalah teknik pengenalan entitas dari data teks untuk mengenali informasi seperti nama, lokasi, waktu, organisasi.

1.6. Sistematika Penulisan

Sistematikan penulisan proposal skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian yang akan dilakukan, dan sistematikan penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi penjelasan tentang teori-teori dan konsep algoritma yang digunakan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini berisi penjelasan langkah-langkah yang akan dilakukan dalam penelitian.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi uraian tentang hasil penelitian dan pembahasan terhadap hasil penelitian yang dilakukan

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari keseluruhan penelitian yang telah dilakukan, serta saran dari penulis untuk kegiatan penelitian selanjutnya terkait dengan topik yang sedang dibahas.

