

BAB I

PENDAHULUAN

1.1. Latar Belakang

Twitter merupakan salah satu layanan jejaring sosial berbasis lokasi yang kini menjadi sangat populer di kalangan masyarakat pengguna internet. Twitter memungkinkan para penggunanya untuk menulis apapun yang mereka inginkan, seperti pengalaman hidup, pendapat, pandangan terhadap topik populer, hingga apapun yang sedang mereka lakukan. Twitter menamai tulisan para penggunanya itu dengan istilah *tweet*.

Berdasarkan data yang dikeluarkan pada April 2014, diketahui bahwa pada kuartal kedua tahun 2014 Twitter memiliki pengguna aktif sebanyak 271 juta jiwa per bulannya (Statista, 2014), dan memiliki jumlah total pengguna aktif sebanyak 645 juta jiwa, dengan jumlah rata-rata *tweet* yang dihasilkan sebanyak 58 juta *tweet* per hari (Statisticbrain, 2014). Sebagai sumber informasi, Twitter terbukti menyediakan data yang berjumlah sangat besar dan mudah didapat. Selain itu informasi yang disediakan oleh Twitter terdiri dari beranekaragam topik, ini membuat Twitter sebagai jejaring sosial berbasis lokasi menjadi salah satu pusat informasi *ter-up-to-date* yang sangat berkualitas, sehingga banyak pengetahuan baru dapat diambil darinya.

Informasi dan pengetahuan baru yang diperoleh dari Twitter dapat digunakan untuk berbagai macam keperluan, salah satunya pada penelitian ini. *Tweet* dari Twitter digunakan untuk mendeteksi aktivitas yang sedang dilakukan oleh para pengguna Twitter di lokasi-lokasi yang ditentukan. Hal ini nantinya dapat digunakan untuk memetakan perilaku pengguna Twitter pada kehidupan nyatanya dengan mengetahui aktivitas yang mereka lakukan dalam kurun waktu tertentu.

Untuk mendapatkan informasi-informasi yang diinginkan tersebut, perlu dilakukan pemrosesan awal terhadap *tweet* yang akan dijadikan sumber data. Ini dikarenakan pola penulisan *tweet* yang beraneka ragam, dimana banyak ditemukan hal-hal yang sangat jarang digunakan seperti kata tidak baku, pengulangan kata yang berlebihan, penggunaan kata-kata gaul, dan pengguna karakter-karakter yang tidak memiliki arti. Hal-hal tersebut harus dihilangkan dari *tweet* agar dapat dihasilkan informasi yang tepat dan bermanfaat.

Penelitian ini melakukan analisis terhadap *tweet* untuk mendeteksi informasi-informasi mengenai aktivitas dengan menggunakan metode Klasifikasi dan metode *clustering*. Metode klasifikasi digunakan untuk memisahkan *tweet* ke dalam tiga kelas yang ditentukan, yaitu aktivitas_lokasi untuk *tweet* yang berisi informasi mengenai aktivitas dan lokasi, non_aktivitas_non_lokasi untuk *tweet* yang tidak berisi informasi mengenai aktivitas dan lokasi, dan kelas lokasi_non_aktivitas untuk *tweet* yang tidak berisi informasi mengenai aktivitas tetapi mengindikasikan pemilik *tweet* sedang berada di lokasi yang ditentukan. Metode *clustering* digunakan untuk mengelompokkan *tweet* dalam kelas aktivitas_lokasi berdasarkan informasi aktivitas yang ada di dalamnya.

Klasifikasi dilakukan dengan menggunakan algoritma Naive Bayes. Naive Bayes merupakan salah satu metode *machine learning* yang menggunakan konsep dasar Teorema Bayes, yaitu melakukan klasifikasi dengan perhitungan nilai probabilitas (Trisedya dan Jais, 2009: 2). Naive Bayes dalam berbagai penelitian empiris terbukti memiliki performa pengklasifikasian yang cukup tinggi, selain itu algoritma ini sangat mudah di implementasikan ke dalam berbagai studi kasus sehingga cocok digunakan untuk mengklasifikasikan *tweet*.

Metode *Clustering* yang akan dilakukan menggunakan algoritma K-Means yang mempartisi data ke dalam *cluster* (kelompok) sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan yang memiliki karakteristik yang berbeda dikelompokkan pada kelompok yang

lain (Narwati, 2010: 2). Algoritma ini memiliki performa yang cukup baik, sangat mudah diadaptasi dan diimplementasikan ke berbagai kasus, selain itu waktu yang diperlukan untuk proses *clustering*-nya pun cukup singkat, sehingga cocok digunakan untuk mendeteksi aktivitas dari sekumpulan *tweet* yang diberikan.

Pemilihan aktivitas sebagai objek pada penelitian ini didasarkan pada perubahan pola hidup manusia yang kini mengarah pada perilaku mobilitas yang tinggi, yang mengakibatkan mereka sering berpindah-pindah lokasi dalam melakukan aktivitas kesehariannya. Karenanya penelitian ini dilakukan untuk dapat menggambarkan perilaku mobilitas manusia tersebut, agar nantinya dapat diketahui apa saja aktivitas yang dilakukan, dan dimana lokasi tempat aktivitas tersebut dilakukan, agar nantinya dapat digunakan sebagai informasi dan pengetahuan baru yang bermanfaat dan tepat guna.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas maka permasalahan yang akan diidentifikasi dalam penelitian ini adalah:

1. Bagaimana algoritma Naive Bayes *Classifier* dapat membantu memisahkan *tweet* yang berisi informasi mengenai aktivitas dan yang tidak.
2. Bagaimana algoritma *clustering* K-Means dapat membantu mengelompokkan informasi mengenai aktivitas dari data kumpulan *tweet*.
3. Bagaimana mengetahui aktivitas yang sedang berlangsung di tempat-tempat tertentu dengan memanfaatkan kumpulan *tweet* para pengguna layanan jejaring sosial berbasis lokasi Twitter.

1.3. Batasan Masalah

Berdasarkan identifikasi masalah serta dengan mempertimbangkan banyak aspek seperti waktu, kemampuan peneliti dan kepentingan penelitian, maka permasalahan dibatasi pada hal-hal sebagai berikut:

1. Layanan jejaring sosial berbasis lokasi yang penulis gunakan sebagai sumber data hanya satu, yaitu Twitter.
2. Data yang diambil dari Twitter merupakan kumpulan *tweet* dari para pengguna yang berlokasi di Indonesia.
3. Data yang diambil dari Twitter merupakan kumpulan *tweet* yang mengandung kata kunci berupa nama-nama lokasi yang telah ditentukan dan dibatasi sebelumnya.

1.4. Tujuan Penelitian

Adapun tujuan yang hendak dicapai dari penelitian ini adalah:

1. Untuk mengetahui cara dan kinerja algoritma *Naive Bayes Classifier* dalam memisahkan *tweet* yang berisi informasi mengenai aktivitas dan yang tidak.
2. Untuk mengetahui cara dan kinerja algoritma *clustering K-Means* dalam mengelompokkan informasi mengenai aktivitas dari data kumpulan *tweet*.
3. Untuk mengetahui aktivitas yang sedang berlangsung di tempat-tempat tertentu dengan memanfaatkan kumpulan *tweet* para pengguna layanan jejaring sosial berbasis lokasi Twitter.

1.5. Manfaat Penelitian

Hasil dari penelitian ini diharapkan mampu menjadi media informasi yang tepat guna bagi siapapun yang membutuhkan, terutama bagi mereka yang memiliki mobilitas cukup tinggi yang membutuhkan informasi dan rekomendasi

mengenai tempat-tempat yang sedang ramai dikunjungi untuk didatangi, juga diharapkan dapat menjadi media informasi yang membantu penggunanya untuk menentukan apakah suatu tempat layak dikunjungi atau malah sebaiknya dihindari, selain itu juga diharapkan bisa memberikan informasi historis mengenai tempat-tempat tersebut bagi para pengguna yang membutuhkan.

1.6. Sistematika Penulisan

Adapun sistematika penulisan penelitian ini dibagi kedalam lima bab, dan masing-masing bab terdiri dari beberapa sub bab, yaitu:

BAB I Pendahuluan

Bab ini berisi Latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat Penelitian, dan sistematika penulisan.

BAB II Kajian Pustaka

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

BAB III Metode 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.