

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam beberapa tahun terakhir, penggunaan media sosial telah menyebar secara luas dalam berbagai macam lapisan masyarakat tidak hanya sebagai media untuk bersosialisasi namun juga untuk menyampaikan aspirasi dan merepresntasikan tentang apa yang terjadi dan dirasakan masyarakat. Kebiasaan masyarakat mengutarakan pendapatnya melalui media sosial terutama *Twitter* dalam menanggapi kejadian atau hal-hal yang terjadi di lingkungannya dapat menjadi salah satu acuan untuk mengetahui sentimen masyarakat terhadap lingkungan atau kota tempat tinggal masyarakat tersebut. Kebutuhan klasifikasi terhadap komentar – komentar baik berupa kritik atau saran yang tersebar di *Twitter*.

Text Mining adalah salah satu solusi dalam mengatasi permasalahan di atas. *Text Mining* dapat didefinisikan sebagai proses pengambilan informasi yang bersumber dari beberapa dokumen. *Text Mining* terdapat beberapa tujuan penggunaan yang khas diantaranya adalah *text categorization*, *text clustering*, dan *sentiment analysis* (Ronen Feldman, 2007). Dalam pengklasifikasian opini yang telah dijelaskan, fungsi *text mining* yang digunakan adalah *sentiment analysis*.

Sentiment analysis atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi *sentiment* yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini

terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (Bo Pang, 2002).

Metode yang digunakan dalam penelitian ini adalah K-NN (*K-Nearest Neighbor*). *K-Nearest Neighbor* adalah salah satu metode untuk melakukan klasifikasi terhadap objek berdasarkan data latih (*data training*) yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut. Pada fase pembelajaran, algoritma ini melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data testing (yang klasifikasinya tidak diketahui). Setelah jarak dari vektor yang baru terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah k buah yang paling dekat diambil, yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut (Ronen Feldman, 2007).

Penelitian ini memiliki tujuan untuk melakukan pengklasifikasian terhadap sentimen positif dan negatif dari *tweet* berbahasa Indonesia. Sumber opini data yang diambil berdasarkan opini dari *twitter* tentang kota Bandung dan kota Yogyakarta dikarenakan dua kota tersebut memiliki banyak kesamaan serta ketersediaan data *corpus* dalam jumlah besar dan ketersediaan algoritma *stemming* dalam bahasa Indonesia yang digunakan untuk proses *preprocessing*, dan *K-Nearest Neighbor* dipilih karena pada sentiment analysis belum banyak digunakan.

Berdasarkan uraian di atas, maka disusun suatu laporan tugas akhir yang berjudul “***Klasifikasi Kategori Komentar Twitter untuk Kota Bandung dan Kota Yogyakarta***”.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana pengklasifikasian terhadap sentimen positif dan sentimen negatif pada menggunakan metode K-NN (*K-Nearest Neighbor*)?
2. Bagaimana akurasi penggunaan metode K-NN (*K-Nearest Neighbor*) dalam mengklasifikasikan opini terhadap kota Bandung dan kota Yogyakarta di *Twitter*?

1.3 Tujuan Penelitian

Adapun tujuan pada penelitian ini adalah sebagai berikut :

1. Mengimplementasikan metode *K-Nearest Neighbor* untuk mengklasifikasi sentimen terhadap data uji.
2. Menganalisis tingkat akurasi penggunaan metode *K-Nearest Neighbor* dalam mengklasifikasi opini dalam *Sentiment analysis*.

1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan adalah data *tweet* seputar kota Bandung dan kota Yogyakarta.
2. Metode *Text Mining* digunakan adalah *K-Nearest Neighbor*.
3. *Tweet* yang diambil hanya dalam bahasa Indonesia.
4. Data berupa dokumen berisikan *tweet* diambil menggunakan aplikasi KNIME Analytical Platform.

1.5 Sistematika Penulisan

Adapun sistematika penulisan laporan tugas akhir ini terdiri dari enam bab, dengan rincian sebagai berikut :

- BAB I **Pendahuluan,**** Membahas mengenai Latar Belakang Masalah Tugas Akhir, Rumusan Masalah Tugas Akhir, Tujuan Tugas Akhir, Batasan Permasalahan Tugas Akhir dan Sistematika Penulisan Tugas Akhir.
- BAB II **Landasan Teori,**** Pada Bagian ini berisi uraian tentang teori-teori yang mendukung dalam mengerjakan Tugas Akhir.
- BAB III **Analisis Sistem,**** Pada bagian ini akan dibahas metodologi penelitian yang dilakukan.
- BAB IV **Perancangan Sistem,**** Pada bagian ini akan dipaparkan kebutuhan-kebutuhan ketika melakukan penelitian. Untuk itu pembahasan ini akan terbagi atas analisis dan perancangan.
- BAB V **Implementasi dan Pengujian Sistem,**** Pada bagian ini akan dibahas tentang tahapan implementasi. Pada bagian implementasi akan melakukan pengujian pada metode-metode yang telah dibuat pada perancangan sistem dan melakukan penghitungan akurasi metode. Bab ini juga membahas tentang hasil dari implementasi dan analisis.
- BAB VI **Penutup,**** Pada bagian ini dibahas tentang kesimpulan dan saran-saran terhadap penelitian yang telah dilakukan.