BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini [1]. Analisis sentimen banyak diterapkan pada kasus dengan jumlah data banyak untuk mendapatkan kandungan informasi dari kumpulan data tersebut. Salah satu media yang bisa digunakan sebagai sumber data untuk dianalisis adalah media sosial seperti Twitter. Twitter adalah layanan untuk berkomunikasi dengan teman, keluarga, dan rekan kerja yang terhubung melalui pertukaran pesan secara cepat [2]. Karena banyaknya data opini, twitter sering digunakan sebagai media untuk melihat tanggapan umum terhadap sesuatu.

Bahasan yang populer diperbincangkan di kalangan pengguna Twitter di Indonesia adalah opini pengguna terhadap kinerja pemerintah. Contohnya opini terhadap kinerja Basuki Tjahaja Purnama (Ahok) yang menjabat sebagai Gubernur DKI Jakarta sejak tahun 2014. Beberapa survei tentang kepuasan warga Jakarta terhadap kinerja Beliau antara lain, Populi Center 58,2% warga Jakarta puas (Desember, 2015), CSIS 67% warga Jakarta puas (Januari, 2016) [3]. Namun terkadang ada beberapa kebijakan kontroversial yang membuat Beliau tidak disukai oleh sebagian kalangan masyarakat. Contohnya adalah kebijakan beliau tentang proyek reklamasi pantai utara, penggusuran kalijodo, penggusuran pasar ikan penjaringan, sampai penertiban bantaran kali ciliwung [4]. Segala opini yang ditujukan kepada Beliau dapat dijadikan bahan evaluasi terhadap kebijakan untuk masa yang akan datang. Namun karena banyaknya perbedaan pendapat, muncul kebutuhan untuk memilah opini yang bertentangan agar memudahkan dalam proses evaluasi dengan menggunakan Analisis Sentimen.

Ada beberapa metode yang digunakan untuk melakukan analisis sentimen, Salah satunya dengan metode *Naive Bayes*. Klasifikasi *Naive Bayes* adalah sebuah

model probabilitas sederhana yang didasari oleh aturan yang bersifat asumsi independen [5]. *Naive Bayes* dipilih karena metode ini memiliki bentuk yang sederhana dibanding dengan metode lainnya sehingga lebih mudah diterapkan. *Naive Bayes* memiliki beberapa jenis pendekatan yang bertujuan membuat akurasi keputusan semakin tinggi. Salah satu pendekatan yang bisa dipakai adalah *Bernoulli Model.* Pendekatan ini cocok digunakan pada dokumen yang hanya memiliki beberapa kata [6]. Twitter memiliki banyak data yang beragam. Untuk mempercepat klasifikasi metode *Naive Bayes* dapat dikombinasikan dengan *Feature Selection.*

Feature Selection mengurangi dimensi ruang fitur, menghapus data berlebihan, data tidak relevan dan data yang mengganggu, hal tersebut bertujuan mempercepat kinerja algoritma, meningkatkan kualitas data dan meningkatkan komprehensibilitas [7]. *Feature Selection* yang dipilih adalah *Chisquare* karena *Naive Bayes* dengan pendekatan *Bernoulli Model* menghasilkan akurasi yang lebih baik ketika dikombinasikan dengan *Chisquare* [8].

Berdasarkan pemaparan yang telah dibahas, penelitian ini dimaksudkan untuk menerapkan Analisis Sentimen Pengguna Twitter terhadap opini tentang kinerja Gubernur DKI Jakarta (Ahok) Menggunakan Metode *Naive Bayes* dengan Pendekatan *Bernoulli Chisquare*.

1.2 Rumusan Masalah

Berdasarkan latar belakang maka dapat dirumuskan masalah bagaimana menerapkan metode *Naive Bayes* dengan pendekatan *Bernoulli* dan *Chisqure* kepada opini pengguna twitter terhadap kinerja gubernur DKI Jakarta (Ahok) sebagai calon gubernur periode selanjutnya dan seberapa akurat dari metode yang diterapkan?

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah menerapkan Analisis Sentimen Pengguna Twitter terhadap Gubernur DKI Jakarta (Ahok) menggunakan Metode *Naive Bayes* dengan Pendekatan *Bernoulli Chisquare*.

Adapun Tujuan dari penelitian ini antara lain :

1. Untuk mengklasifikasikan sentimen pengguna twitter secara otomatis berkaitan dengan kinerja Basuki Tjahya Purnama sebagai Gubernur DKI Jakarta menggunakan metode *Naive Bayes* dengan pendekatan *Bernoulli Chisquare*.
2. Untuk mengetahui besarnya akurasi dari metode *Naive Bayes* dengan pendekatan *Bernoulli Chisquare* dalam melakukan analisis sentimen pengguna twitter terhadap kinerja Basuki Tjahya Purnama sebagai Gubernur DKI Jakarta.

1.4 Batasan Masalah

Adapun batasan masalah untuk mempermudah ruang lingkup dalam proses pengerjaan serta analisa. Adapun permasalahan yang dibatasi adalah sebagai berikut:

1. Ketika data uji ditemukan kata-kata selain Bahasa Indonesia maka kata tersebut akan diabaikan namun tidak dihilangkan.
2. Akun Twitter yang dianalisa adalah @basuki_btp sebagai akun o*fficial* dari figur Basuki Tjahya Purnama (Ahok).
3. Tweet yang digunakan sebagai data training adalah yang di-*mention* pada bulan April 2016 dan Agustus 2016.
4. *Stopword List* sebagai data bantu dalam proses *Preprocessing* menggunakan sebagian hasil penelitian Tala, F.Z tentang efek stemming terhadap bahasa indonesia [9].
5. Metode *Stemming* menggunakan Algoritma Nazief dan Andriani [10].

1.5 Metodologi Penelitian

Metode yang diterapkan pada penelitian ini berupa Metode Kuantitatif, karena penjabaran latar belakang masalah berawal dari umum menjadi masalah yang lebih khusus. Berikut adalah metode pengumpulan data dan pembangunan perangkat lunak:

1.5.1. Metode Pengumpulan Data

a. Studi literatur

Mengumpulkan data melalui berbagai referensi baik buku, jurnal, *paper*, dan artikel-artikel lain yang bersangkut paut dengan penelitian tentang analisis sentimen pengguna twitter, khususnya dengan metode *Naive Bayes*.

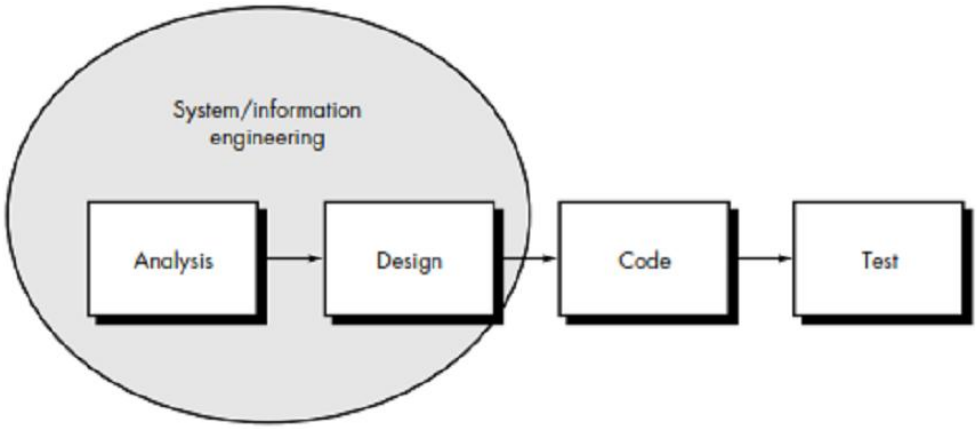
b. Pengumpulan Data

Pengumpulan data training menggunakan Bahasa Python dengan *Library Tweepy* untuk mengakses API Twitter, hasil yang diperoleh berbentuk file JSON untuk kemudian diubah menjadi file CSV. *Query* yang dimasukkan adalah "@basuki_btp" dengan periode bulan april 2016, sehingga data yang dikumpulkan hanya *tweet* yang mengandung *mention* @basuki_btp pada bulan tersebut.

Setelah mendapatkan data mentah, data dipilah secara manual untuk mendapatkan data training yang berkualias. *Tweet* yang tidak berhubungan dengan fakta atau opini tentang @basuki_btp akan dihilangkan. Selain pengumpulan data training sebagai data utama, pengumpulan data bantu pun tak kalah penting untuk menunjang proses *preprocessing.* Data yang dimaksud antara lain : Data *Emoticon, Stopword, Negation* dan *Stemming.*

1.5.2. Metode Pembangunan Perangkat Lunak

Teknik analisis dalam pembangunan perangkat lunak menggunakan *The Linear Sequential Model* dari buku Pressman [12]. Tahapan dari pembangunan perangkat lunak dapat dilihat seperti pada gambar 1.1



Gambar 1.1 Model Pembangunan Perangkat Lunak

Berikut adalah penjelasan dari tiap tahapan pada model Pembangunan Perangkat Lunak:

a. *Analysis*

Tahapan ini merupakan tahap awal pengumpulan kebutuhan dari perangkat lunak yang akan dibangun dengan melakukan pengumpulan data dengan cara observasi terhadap data yang akan digunakan berkaitan dengan klasifikasi analisis sentimen.

b. *Design*

Tahapan ini merupakan tahap penerjemahan kebutuhan perangkat lunak yang didapatkan ke dalam bentuk analisis lalu dilanjutkan ke dalam pemodelan perangkat lunak, bertujuan untuk memberikan gambaran yang harus dikerjakan, arsitektur perangkat lunak dan representasi antar muka.

c. *Code*

Tahapan ini merupakan tahap pembangunan perangkat lunak dengan menejemahkan hasil dari pemodelan perangkat lunak atau *design* ke dalam bentuk kode yang dapat dimengerti oleh komputer.

d. *Test*

Setelah perangkat lunak dibangun dilakukan pengujian terhadap logika dari perangkat lunak untuk mencari celah kesalahan pada perangkat lunak agar didapatkan perangkat lunak yang sesuai dengan kebutuhan.

1.6. Sistematika Penulisan

Untuk menentukan gambaran secara umum mengenai isi laporan skripsi, maka sistematika penulisan laporan adalah sebagai berikut:

BAB I PENDAHULUAN

Berisi pembahasan umum yang berhubungan dengan penyusunan skripsi, yang meliputi latar belakang masalah, perumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian, dan sistematika penulisan laporan.

BAB II TINJAUAN PUSTAKA

Berisi pembahasan umum rangkaian penjelasan yang berhubungan dengan teori-teori yang akan digunakan dalam analisis sentimen, penjelasan twitter, sampai teori pembahasan algoritma yang digunakan.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini dijelaskan analisis dan perancangan mencakup analisis masalah, analisis sistem, analisis sumber data, analisis algoritma analisis kebutuhan non-fungsional, analisis kebutuhan fungsional, dan perancangan tatap muka.

BAB IV PERANCANGAN DAN IMPLEMENTASI

Merupakan tahapan yang menguraikan tentang implementasi, perangkat keras, perangkat lunak, implementasi dari aplikasi dan pengujian aplikasi.

BAB V. KESIMPULAN DAN SARAN

Berisi kesimpulan dan saran yang sudah diperoleh dari hasil penelitian skripsi.