

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi pada bidang informasi telah membuat berbagai aplikasi media sosial bermunculan seperti halnya Facebook, Twitter, Instagram dan lain-lain. Menurut Andreas Kaplan dan Michael Heinlein, mendefinisikan “media sosial adalah suatu pengelompokan *software* atau perangkat lunak berbasis Internet yang berada di atas dasar ideologi dan teknologi Web 2.0, serta yang dapat memungkinkan *user* atau pengguna untuk penciptaan dan pertukaran “user-generated content” (Kaplan & Haenlein, 2010). Terdapat beberapa jenis media sosial yang berkembang sampai saat ini dan penggunaanya(*user*) masih terbilang juga cukup banyak yang aktif, salah satunya adalah situs jejaring sosial Twitter.

“Twitter adalah salah satu dari media sosial, aplikasi yang berbasis microblogging. Microblogging merupakan jenis media sosial yang memfasilitasi pengguna untuk menulis dan memublikasikan aktivitas atau pendapat.” (Viani, 2017). Melanjutkan penelitian dari Afiandiary data yang diperoleh pada Januari 2016 berdasarkan kisaran umur 20-25 tahun, peringkat pertama masih ditempati oleh Facebook sebesar 86,1%, kedua adalah Instagram sebanyak 75,8% dan ketiga adalah sebanyak Twitter sebanyak 41,5%. Pada tahun 2016 30,1 juta pengguna. Pada tugas akhir ini, peneliti memilih subject Twitter karena media tersebut sangatlah simple dalam penggunaannya serta dapat mengirimkan pesan teks mencapai 140 karakter.

Dengan adanya media sosial Twitter. Setiap orang dapat saja saling berbagi informasi terhadap orang lain tanpa harus bertemu satu dengan yang lainnya dan juga memiliki kebebasan untuk mengemukakan pendapat. Dengan media sosial pengguna juga dapat mempengaruhi hal buruk pengguna lain dengan membuat dan menyebarkan informasi yang bersifat tuduhan, fitnah, berita hoax, maupun SARA. “Dalam media sosial dikenal istilah Ucapan kebencian atau dikenal dengan *Hate Speech*, yang makin populer saat ini, hal ini disebabkan gesekan atau perbedaan yang mewakili kelompok-kelompok tertentu baik Suku, Agama, Ras, Etnis, Golongan.” (Rohman, 2016). Atas dasar berbagai permasalahan di media sosial tersebut akhirnya pemerintah Indonesia membuat aturan terkait berbagai kejahatan yang terjadi di sosial media. (Pemerintah Indonesia, 2008).

Berdasarkan permasalahan di atas peneliti akan membangun sistem untuk menganalisis sentiment tweet dan pengklasifikasian atau mengelompokkan tweet tersebut mengandung *hate speech* dengan menggunakan metode text mining dan metode Naïve Bayes sebagai pengklasifikasi. “Text mining adalah salah satu perkembangan dari analisis teks prosesnya dikerjakan secara otomatis dengan komputasi oleh computer berguna sebagai penggalian informasi yang berkualitas mencari inti sari dari suatu rangkaian teks atau kata yang terangkum dalam sebuah dokumen” (Han, Kamber, & Pei, 2011). Metode ini akan peneliti gunakan sebagai pengolah data text dalam tweet pengguna Twitter dari hasil *crawling* data secara *real time* pada Twitter. Selain itu pengguna juga menggunakan metode Naïve Bayes. Metode Naïve Bayes Classifier dapat mengklasifikasi dari hasil *Crawling* data di Twitter dengan *Hate Space* atau perkataan kebencian. Dengan sistem tersebut diharapkan dapat bermanfaat untuk mengetahui performa dari algoritme *Naive Bayes Classifier* sebagai pengklasifikasi pengguna media sosial Twitter terkait *Hate Speech* dalam tweet pengguna Twitter.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah

- a. Bagaimana menganalisis sentimen pada *tweet* dan pengklasifikasian *tweet* menggunakan metode *Naïve Bayes Classifier*?
- b. Bagaimana akurasi model *Naïve Bayes Classifier* dengan menambahkan teknik *cross validation* untuk teknik memvalidasi data ?
- c. Bagaimana perbandingan dari akurasi model *Naïve Bayes Classifier* dengan model *Logistic Regression* sebagai model klasifikasi ?

## 1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah :

- a. Peneliti dapat mengolah data Twitter untuk memperoleh inti dari tweet pengguna menggunakan metode *text mining*.
- b. Melakukan pengklasifikasian tweet berbahasa indonesia dengan menggunakan metode *Naïve Bayes Classifier*.
- c. Mengetahui akurasi model klasifikasi menggunakan metode *Naïve Bayes Classifier* dengan tambahan fitur *Unigram* sebagai pengujian .

- d. Mencari metode terbaik untuk membangun sistem pengklasifikasian pada tweet dengan membandingkan metode *Naïve Bayes Classifier* dengan *Logistic Regression*.

#### 1.4 Batasan Masalah

Berdasarkan dari rumusan masalah di atas, terdapat beberapa batasan masalah untuk sistem tersebut sehingga ruang pengerjaan tidak terlalu melebar. Berikut adalah batasan masalah yang akan dilakukan:

- a. Sistem yang dibuat menggunakan metode *text mining* sebagai pengolahan data text hasil dari *crawling data* di Twitter .
- b. Jumlah data yang di *crawling* hanya 2500 tweet.
- c. Hasil akhir dari penelitian adalah analisis sentimen dari metode *Naïve Bayes Classifier* sebagai pengklasifikasi data Twitter.
- d. Data yang di- *Crawling* pada Twitter berupa Bahasa Indonesia.

#### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah peneliti mengharapkan dapat membantu pihak-pihak yang ingin menganalisa tweet oleh pengguna Twitter serta mengetahui performa dari algoritme *Naive Bayes Classifier* sebagai pengklasifikasi berdasarkan status tweet pengguna Twitter.

#### 1.6 Metodologi Penelitian

Metode penelitian dalam penelitian membangun Implementasi Text Mining Untuk Mendeteksi *Hate Speech* pada Twitter ini menggunakan metode penelitian *waterfall*. Metode ini juga sering disebut dengan metode air terjun. Metode *waterfall* adalah permodelan suatu sistem informasi dengan cara sistematis dan urut, metode ini dimulai dari tahap pengumpulan data, analisis kebutuhan, perancangan sistem, implementasi, dan pengujian sistem. Permodelan sistem tersebut sangat cocok dimanfaatkan agar sistem tetap terjaga karena pengembangan metode ini terstruktur.

Adapun beberapa langkah penyelesaian yang dapat dilakukan dalam pembangunan sistem tersebut adalah :

- a. Pengumpulan Data

Pada pengumpulan data, peneliti akan mengumpulkan data dengan mencari referensi jurnal dan diktat yang berhubungan dengan topik penelitian ini. Pengumpulan data juga

dilakukan dengan mengumpulkan data *tweet* dengan mengambil data langsung (*Crawling data*) Twitter menggunakan API (*Application Interface*) pada Twitter.

b. Analisis Kebutuhan

Pada analisis kebutuhan, peneliti akan menentukan apa saja yang akan dibutuhkan dalam pembangunan sistem seperti kebutuhan *input*, *proses* dan *output*. Dalam hal ini peneliti menganalisa sentimen terhadap tweet dari pengguna Twitter terhadap suatu topik berdasarkan *hashtag* sebagai sumber data. Setelah data terkumpul peneliti akan mempersiapkan *tools* pendukung untuk menganalisis sentimen seperti *Anaconda* sebagai pendistribusi *Python*, *Python package library* sebagai bahasa pemrograman yang peneliti gunakan, dan *library* yang dapat sebagai pendukung melakukan analisis sentimen menggunakan metode *Naive Bayes Classifier*.

c. Analisis Perancangan

Pada analisis perancangan peneliti akan membuat gambaran rancangan alur pengerjaan analisis data yang nantinya berguna mempermudah orang lain untuk memahami proses analisis data sentimen. Cara yang dilakukan peneliti yaitu dengan membuat diagram alur analisis sentimen.

d. Implementasi

Pada implementasi peneliti akan melakukan analisis kebutuhan dan analisis perancangan setelah dari proses tersebut peneliti akan mengimplementasikan kebutuhan sebagai analisis sentimen ini sesuai dengan alur yang dibuat pada saat melakukan proses analisis perancangan.

e. Pengujian Sistem dan evaluasi sistem

Pada pengujian sistem, sistem akan diuji terlebih dahulu sebelum digunakan. Hal tersebut berguna meminimalisir adanya kesalahan dalam sistem. Sistem akan diuji dengan mengambil data tweet di Twitter secara *realtime* pengambilan tersebut berdasarkan fitur pencarian *hashtag* pada Twitter, setelah data diambil sistem akan memberikan laporan hasil pengambilan berupa nama akun, tanggal, tweet, dan lain-lain. Setelah mendapatkan laporan hasil pengambilan data, data tersebut akan dilakukan pengujian apakah hasil dari data tersebut sesuai dengan analisis sentimen tujuan atau tidak. Pengujian ini juga dapat mengetahui kekurangan dan kelemahan analisis sentimen yang sudah berjalan. Jika dalam

proses pengujian tidak berjalan maksimal atau tidak berhasil dan tidak sesuai akan kembali lagi ke dalam proses implementasi dan dilakukan pengujian kembali.

### **1.7 Sistematis Penulisan**

Sistematika penulisan sebagai gambaran singkat struktur penulisan laporan, serta isi setiap struktur bagian. Struktur bagian tersebut dijelaskan sebagai berikut :

**BAB I PENDAHULUAN** dalam bab ini berisi terkait latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian dan langkah penyelesaian.

**BAB II LANDASAN TEORI** dalam bab ini berisi terkait teori dasar yang berkaitan dengan analisis sentimen dengan menggunakan metode Naive Bayes Classifier, dan tinjauan pustaka yang nantinya sebagai dasar pengacuan pengerjaan tugas akhir.

**BAB III METODOLOGI PENELITIAN** dalam bab ini berisi terkait langkah-langkah penyelesaian masalah dari tahap pengumpulan data, analisis kebutuhan, dan pengimplementasian.

**BAB IV HASIL DAN PEMBAHASAN** dalam bab ini berisi terkait hasil dari penyelesaian masalah dari sistem di atas dan juga pembahasan sistem.

**BAB V SIMPULAN DAN SARAN** dalam bab ini berisi terkait rangkuman dari hasil analisis sistem kinerja pada bagian sebelumnya serta saran-saran yang perlu diperhatikan guna pengembangan sistem.