

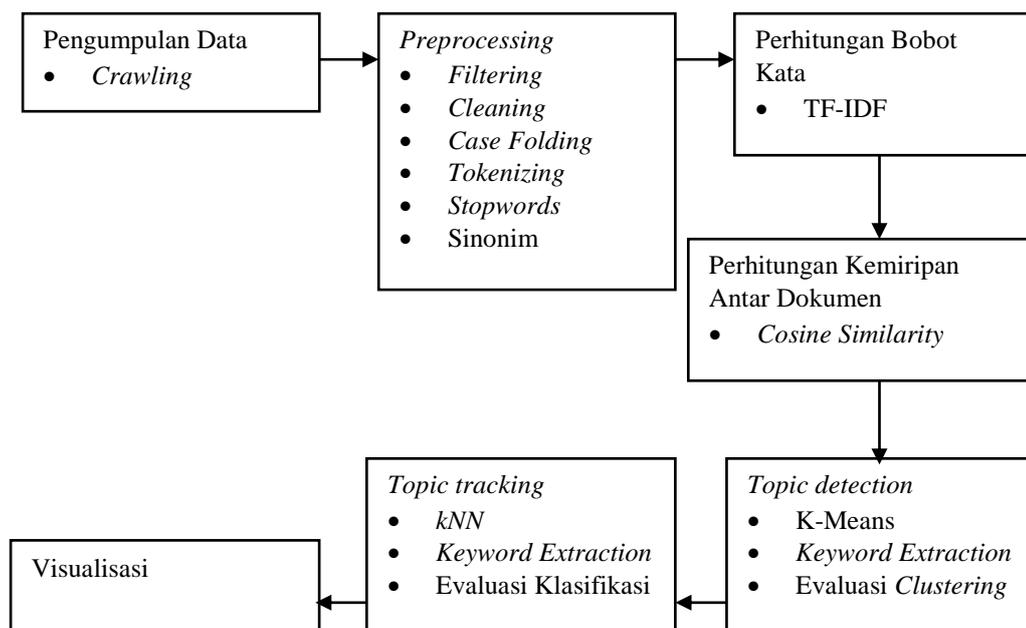
# BAB III

## METODOLOGI PENELITIAN

Untuk menunjang kegiatan penelitian dalam bab ini akan dijelaskan desain penelitian, metode penelitian yang digunakan, serta alat dan bahan penelitian.

### 2.1 Desain Penelitian

Desain penelitian adalah tahapan atau gambaran yang akan dilakukan dalam penelitian. Desain penelitian ini dibuat untuk memberikan gambaran serta kemudahan dalam melakukan penelitian. Gambar 3.1 tahapan penelitian yang digunakan :



Gambar 3. 1 Desain Penelitian

#### 1. Pengumpulan Data

Pengumpulan data dilakukan dengan proses *crawling* data *tweet* dari Twitter dengan memanfaatkan Stream API Twitter. Data yang terkumpul dari tahapan ini lebih dari 39.790 *tweet* yang mengandung kata kunci “bandung”.

## 2. *Preprocessing*

Berikut ini merupakan tahapan yang dilakukan dalam proses *preprocessing*.

### a. *Cleaning*

Pada tahapan ini dilakukan pembersihan *tweet* dari kata dan karakter yang tidak diperlukan sehingga dapat mengurangi *noise*. Contohnya adalah dengan melakukan penghapusan URL dan karakter tanda baca.

### b. *Case Folding*

*Case folding* adalah proses penyeragaman bentuk huruf, sehingga *tweet* hanya terdiri dari huruf a sampai z.

### c. *Tokenizing*

*Tokenizing* adalah proses pemecahan sebuah kalimat menjadi kata-kata.

### d. *Stopwords*

Pada tahapan ini dilakukan proses penghilangan *tweet* dari kata-kata yang termasuk dalam daftar kata *stopword* bahasa Indonesia.

### e. Sinonim

Pada tahapan ini dilakukan proses penyeragaman makna dari kata-kata yang termasuk dalam daftar kata sinonim bahasa Indonesia.

## 3. Perhitungan Pembobotan Kata

Pada tahapan ini dilakukan proses perhitungan bobot kata dengan menggunakan TF-IDF dari kata-kata *tweet* yang sebelumnya telah dikonversi menjadi bentuk *vector space model*.

## 4. Perhitungan Kemiripan Antar Dokumen

Pada tahapan ini dilakukan proses perhitungan nilai kemiripan antar dokumen dengan menggunakan *cosine similarity* dari *vector space model* yang dihasilkan oleh proses pembobotan kata.

## 5. *Topic detection*

Berikut merupakan tahapan dari proses *topic detection* :

### a. K-Means

Pada tahapan ini digunakan algoritma K-Means untuk mengelompokkan *tweet* kedalam sejumlah topik.

b. *Keyword Extraction*

Pada tahapan ini dilakukan proses pengekstraksian kata sehingga akan diperoleh kata yang dapat merepresentasikan setiap *cluster*.

c. *Evaluasi Clustering*

Pada tahapan ini dilakukan proses evaluasi *clustering*.

6. *Topic tracking*

Berikut merupakan tahapan dari proses *topic tracking* :

a. *k-NN*

Pada tahapan ini digunakan algoritma klasifikasi k-NN untuk mengelompokkan *tweet* ke dalam topik yang telah terkumpul dalam proses *topic detection* dan *cosine similarity* untuk menentukan nilai kemiripan antar *tweet*.

b. *Evaluasi Klasifikasi*

Pada tahapan ini dilakukan proses evaluasi klasifikasi.

7. *Visualisasi*

Pada tahapan ini digunakan visualisasi untuk menampilkan topik-topik yang telah terkumpul.

## 2.2 Metode Penelitian

Metode penelitian ini dibagi menjadi dua, yaitu metode pengumpulan data dan metode pengembangan perangkat lunak.

### 2.2.1 Metode Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan *tweet* yang telah dikumpulkan dengan proses *crawling* dengan menggunakan Streaming API Twitter. Data hasil *crawl* tersebut didapatkan dalam format JSON lalu dimasukkan ke dalam *database* untuk proses lebih lanjut. Sementara itu data yang digunakan untuk menilai visualisasi diperoleh dari survei. Kuesioner digunakan sebagai instrument penelitian dalam survei. Pertanyaan dari kuesioner ini berjumlah enam buah yang terdiri dari pilihan “sangat tidak setuju”, “tidak

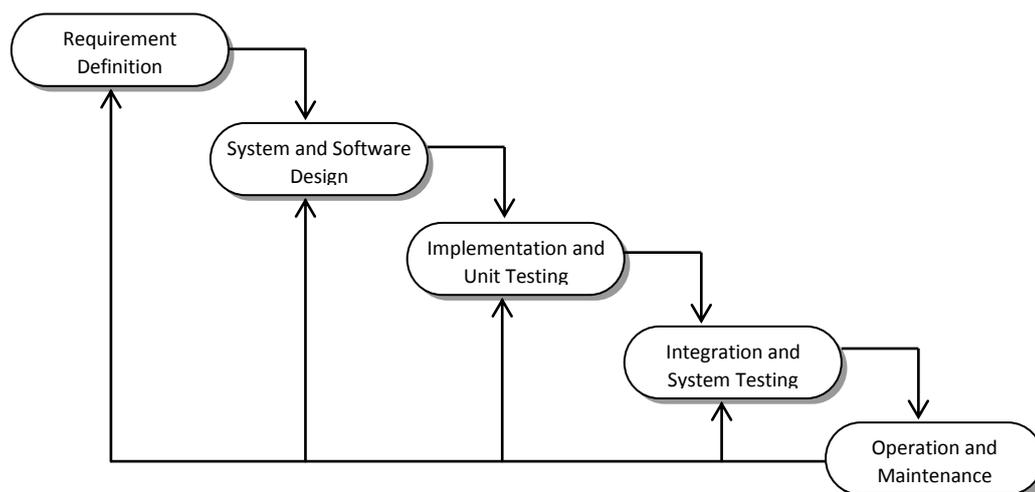
setuju”, “netral” “setuju”, dan “sangat setuju” (Likert, 1932). Hasil kuesioner disajikan dengan menggunakan metode distribusi frekuensi.

Populasi adalah keseluruhan dari objek penelitian (Notoatmodjo, 2005). Populasi pada penelitian ini yaitu jumlah pengguna internet di Indonesia dengan rentang usia 18-25 tahun. Rentang tersebut diambil berdasarkan rentang usia terbanyak pengguna internet yaitu sebesar 49% dari total 88,1 juta pengguna atau sebesar 43,16 juta (APJII, 2014).

Sampel adalah sebagian objek yang diambil dari keseluruhan objek yang diteliti dan dianggap mewakili seluruh populasi (Notoatmodjo, 2005). Jumlah *sample* untuk penelitian antara 10-20 sample (Roscoe, 1975). Pada penelitian ini, jumlah sample sebanyak 20 dengan rentang usia 17-25 tahun.

### 2.2.2 Metode Pengembangan Perangkat Lunak

Pembangunan perangkat lunak dalam penelitian ini menggunakan model *waterfall* (Sommerville, 2011). Dalam model *waterfall* Sommerville terdapat kemungkinan untuk kembali ke tahap sebelumnya apabila terjadi kesalahan atau perbaikan, dimana alur prosesnya seperti pada Gambar 3.2.



Gambar 3. 2 Model *Waterfall* (Sommerville, 2011)

1. *Requirement Definition*, Tahap awal dimana adanya analisis untuk menentukan kebutuhan, batasan, dan tujuan (*goal*) dari perangkat lunak sesuai yang diinginkan. Hal tersebut kemudian didefinisikan secara rinci dan terbentuk sebagai spesifikasi sistem. Pada tahap ini dilakukan pengumpulan data *tweet* dengan kata kunci “bandung” dan menentukan algoritma apa yang digunakan pada proses *Topic detection and tracking* (TDT).
2. *System and Software Design*, *System design* merupakan proses perancangan perangkat keras maupun perangkat lunak yang dilibatkan untuk menunjang sistem yang akan dibangun. Sementara itu *software design* merupakan proses perancangan yang melibatkan identifikasi dan menggambarkan dasar sistem serta hubungan satu sama lain. Pada tahap ini dibuat desain dari sistem yang akan dikembangkan yaitu TDT.
3. *Implementation and Unit Testing*, Pada tahap ini, *software design* yang telah dilakukan sebelumnya kemudian diimplementasikan dalam bentuk unit program. Setelah unit program dibuat, kemudian dilakukan *testing* pada unit program tersebut untuk memastikan implementasi berjalan dengan baik.
4. *Integration and System Testing*, Setelah semua unit program berhasil diimplementasikan dan lolos *testing* maka dilanjutkan dengan mengintegrasikan setiap unit untuk membentuk sistem yang diinginkan. Sistem yang sudah dibentuk kemudian di tes kembali untuk memastikan unit program dapat berjalan satu sama lain dalam sistem dan sistem yang dibuat sudah memenuhi kebutuhan.
5. *Operation and Maintenance*, Tahap ini merupakan tahap dimana sistem sudah dipasang kemudian melakukan perbaikan ketika terdapat kesalahan atau *error* yang tidak ditemukan sebelumnya saat pembangunan sistem berlangsung. Perbaikan juga dilakukan jika terdapat kebutuhan baru yang perlu ada pada sistem.

### 2.3 Alat dan Bahan Penelitian

Berdasarkan kebutuhan-kebutuhan di atas, maka ditentukan bahwa alat dan bahan yang digunakan pada penelitian ini adalah sebagai berikut:

### 2.3.1 Alat Penelitian

Dalam penelitian ini, peneliti menggunakan berbagai alat bantu penunjang baik berupa perangkat keras maupun perangkat lunak. Adapun perangkat keras yang digunakan adalah seperangkat komputer yang mempunyai spesifikasi sebagai berikut:

1. *Processor* Intel i5
2. RAM 4 GB
3. *Hard disk* 750 GB
4. Mouse dan Keyboard

Sementara itu perangkat lunak yang digunakan adalah sebagai berikut:

1. Sistem Operasi Microsoft Windows 8.1 64 bit
2. JDK (Java Development Kit)
3. JRE (Java Runtime Environment)
4. Netbeans 8.0
5. Xampp
6. HeidiSQL
7. Sublime Text
8. Google Chrome

### 2.3.2 Bahan Penelitian

Bahan penelitian yang digunakan pada penelitian kali ini adalah *tweet* yang didapat dari *crawling* data yang dilaksanakan pada tanggal 22 Mei 2015.