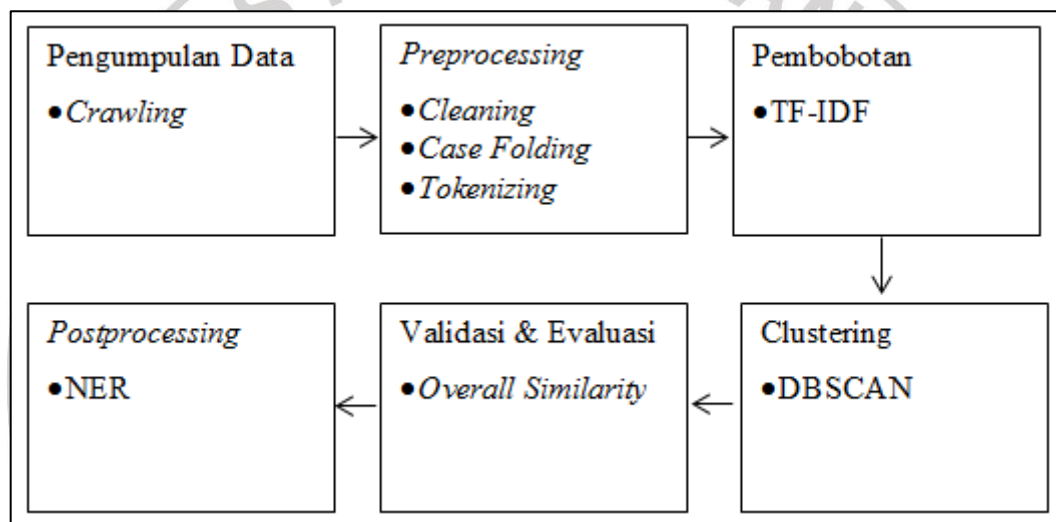


## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Desain Penelitian

Desain penelitian merupakan tahapan yang akan dilakukan oleh penulis untuk memberikan gambaran serta kemudahan dalam melakukan penelitian. Berikut tahapan penelitian yang digunakan:



Gambar 3. 1 Desain Penelitian

##### 1) Pengumpulan data

Data yang digunakan diambil dari situs jejaring sosial Twitter. *Tweet* yang saat ini penulis dapatkan ada lebih dari 250.000 *tweet* terkait dengan banjir. Data tersebut didapatkan dengan cara melakukan *crawling* dengan memanfaatkan Twitter Search API. Data diambil dari bulan Januari 2012 sampai Februari 2012 dan bulan Januari 2013.

##### 2) *Preprocessing*

Yang akan dilakukan pada tahapan ini adalah sebagai berikut:

- a) *Cleaning*, yaitu proses membersihkan data *tweet* dari kata yang tidak diperlukan untuk mengurangi *noise*. Beberapa contoh kata yang dihilangkan adalah *hashtag* (#), *username* (@*username*), dan url. Selain penghapusan kata, dilakukan penghapusan tanda baca seperti koma (,) dan titik (.
  - b) *Case Folding*, yaitu penyeragaman bentuk huruf, sehingga data hanya menjadi huruf latin dari a sampai z.
  - c) *Tokenizing*, yaitu proses memecah sebuah kalimat menjadi kata.
- 3) Pembobotan
- Pada tahap ini akan dilakukan konversi data teks menjadi data vektor menggunakan *Vector Space Model*. Sebuah data *tweet*, dinyatakan dalam model *vector-space*. Semua kata yang terkandung dalam *tweet* didaftarkan, dinyatakan dengan  $\{f_1, f_2, \dots, f_m\}$  yang merupakan daftar kata. Setiap kata akan dihitung nilai TF-IDF sebagai representasi dari kata.
- 4) Pembelajaran dan Klasifikasi
- Pada tahap ini akan digunakan algoritma *clustering* DBSCAN untuk melakukan pengelompokan *tweet* ke dalam sejumlah *event*.
- 5) Evaluasi
- Pada tahap ini akan dilakukan uji kualitas hasil *cluster* dari kinerja DBSCAN menggunakan *overall similarity*.
- 6) *Postprocessing*
- Pada tahap ini akan dilakukan pengambilan informasi dari *tweet* yang terbentuk pada setiap *cluster*.

### 3.2. Metode Penelitian

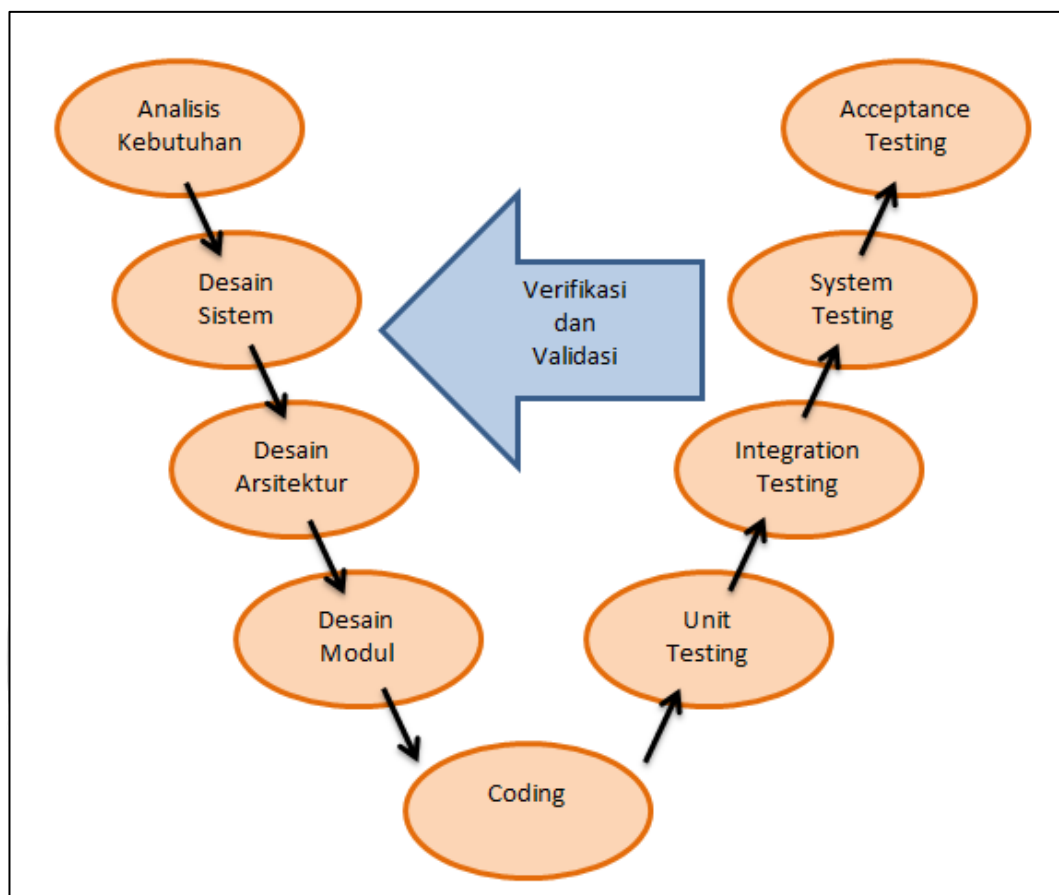
Untuk lebih jelasnya, metode penelitian yang dilakukan dijelaskan dalam sub-bab berikut:

#### 3.2.1. Proses Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data yang diambil dari Twitter melalui *Search* API di alamat: <http://search.twitter.com/search.json?q=banjir>. Diambil pada rentang bulan

Januari – Februari 2012 dan Januari 2013. Data hasil *crawl* ini dimasukkan ke *database* untuk diproses lebih lanjut.

### 3.2.2. Proses Pengembangan Perangkat Lunak



Gambar 3. 2 V-Model

Model proses pengembangan atau rekayasa perangkat lunak yang digunakan dalam penelitian ini adalah V Model. V Model merupakan perbaikan dari model *waterfall* (Balaji, 2012).

1. Analisis Kebutuhan, tahap pengumpulan data dan informasi terkait dengan metode yang akan digunakan dan berbagai kebutuhan penunjang lain untuk mendapatkan gambaran aplikasi. Pada tahap ini dilakukan pengumpulan data *tweet* dengan *keyword* banjir dan menentukan algoritma apa yang digunakan untuk melakukan *event detection*.

2. Desain Sistem, tahap pembuatan desain rekayasa sistem yang akan dibuat sesuai dengan analisa kebutuhan. Pada tahap ini dibuat desain dari sistem yang akan dikembangkan untuk *event detection* (lihat gambar 3.1).
3. Desain Arsitektur/Program, tahap pembuatan desain rekayasa struktur program, dan algoritma. Pada tahap ini ditentukan modul apa saja yang akan dibuat untuk melakukan *event detection*, yaitu modul *preprocessing*, modul pembobotan, modul, *clustering*, modul *postprocessing*.
4. Desain Modul, tahap pembuatan modul-modul program hasil implementasi dari tahap Desain Arsitektur/Program. Tahap ini membuat programmer dapat melakukan pemecahan unit-unit yang masing-masing unitnya memiliki fungsi yang sama. Pada tahap ini dibuat kerangka prosedur atau fungsi pada setiap modul.
5. *Coding*, tahap pembuatan aplikasi yang merupakan tindak lanjut dari perancangan program untuk mengolah data menjadi informasi berupa kode-kode program.
6. *Unit Testing*, tahap pengujian terhadap aplikasi yang telah dibuat. Jika terdapat kekurangan maka tahap ini berhubungan dengan tahap Desain Arsitektur/Program. Pada tahap ini diuji apakah ada *task* yang terlewat atau tidak ketika melakukan *event detection*.
7. *Integration Testing*, tahap melakukan integrasi dari modul-modul yang terpisah menjadi satu kesatuan yang utuh. Pada tahap ini setiap modul yang telah dibuat diuji fungsionalitasnya melalui antarmuka sistem.
8. *System Testing*, tahap pengujian sistem. Jika terdapat kekurangan maka tahap ini berhubungan dengan tahap Desain Sistem. Pada tahap ini dilakukan pengujian keseluruhan sistem dari mulai tahap *preprocessing* sampai *postprocessing* (ekstraksi informasi).
9. *Acceptance Testing*, tahap pengujian yang bersifat menyeluruh, yakni berupa simulasi terhadap data sebenarnya. Jika terdapat kekurangan maka tahap ini berhubungan dengan tahap Analisis Kebutuhan. Pada tahap ini dilakukan pengecekan terhadap data hasil *clustering* dan juga *event* yang terdeteksi.

### 3.3. Alat dan Bahan Penelitian

Alat yang digunakan dalam penelitian ini adalah seperangkat komputer yang dilengkapi perangkat keras dan perangkat lunak pendukung. Sedangkan bahan yang digunakan adalah data yang diambil melalui *microblogging* Twitter.

#### 3.3.1. Alat Penelitian

Dalam penelitian ini digunakan perangkat keras komputer dengan spesifikasi sebagai berikut:

1. *Processor* Intel Core 2 Solo
2. Memori 4 GB RAM
3. Hardisk berkapasitas 320 GB
4. Monitor 14” dengan resolusi 1366x768 pixel
5. Mouse dan Keyboard

Adapun perangkat lunak yang digunakan adalah

1. Microsoft Windows 7 Ultimate
2. Netbeans 6.8
3. Java SE 7
4. Notepad++

#### 3.3.2. Bahan Penelitian

Adapun bahan penelitian yang digunakan berupa data yang diambil dari Twitter dengan kata kunci ‘banjir’. Serta bahan lain berupa kumpulan kata *stopword* dan kumpulan kata sinonim yang diambil *library TweetMining* Berbahasa Indonesia yang dibuat oleh Yudi Wibisono di alamat: <http://code.google.com/p/id-tweetmining/>.