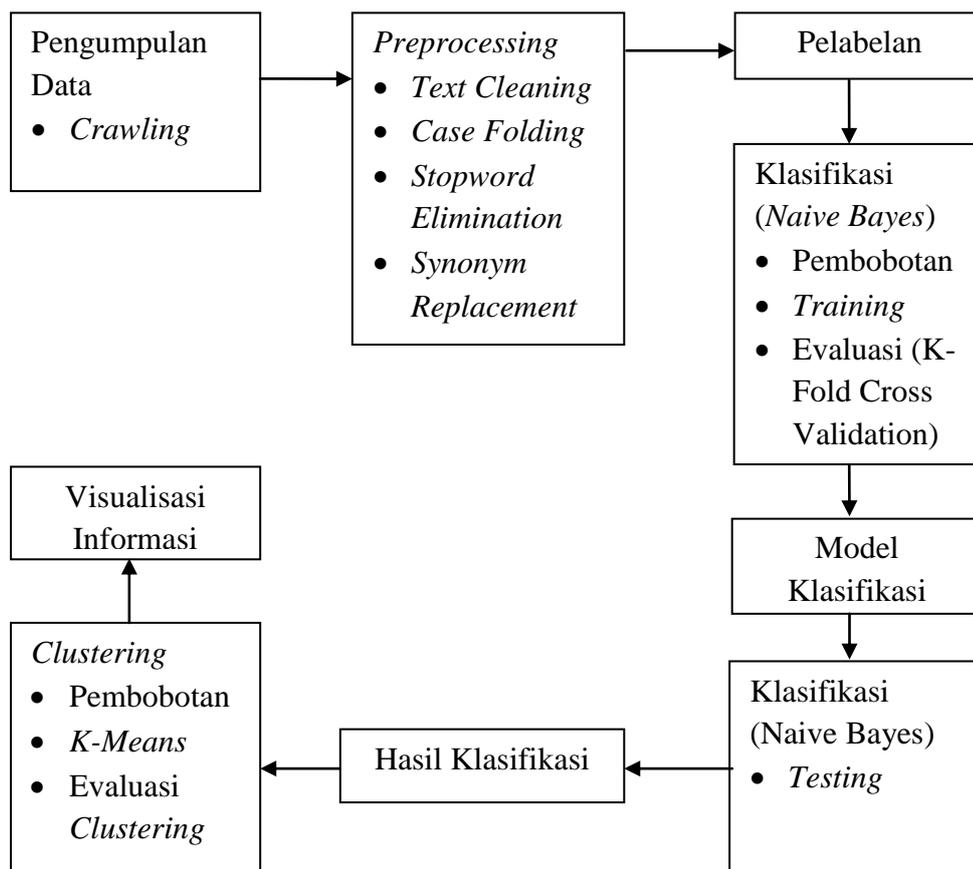


BAB III METODOLOGI PENELITIAN

1.1. Rancangan Penelitian

Rancangan penelitian adalah rencana dan struktur penyelidikan yang disusun sedemikian rupa sehingga penelitian akan memperoleh jawaban untuk pertanyaan-pertanyaan penelitiannya (Kerlinger, 1990: 483). Desain penelitian merupakan tahapan-tahapan yang akan dilakukan oleh penulis dalam penelitian. Berikut tahapan penelitian yang akan dilakukan:



Gambar 3.1 Desain Penelitian

Penjelasan lebih rinci dari rancangan penelitian diatas adalah sebagai berikut:

1) Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari layanan jejaring sosial berbasis lokasi Twitter. Data yang diperoleh merupakan kumpulan *tweet* yang didapatkan dengan menggunakan Twitter API dengan *keywords* nama-nama pusat perbelanjaan yang terdapat di kota Bandung, adapun rinciannya adalah sebagai berikut:

2) *Preprocessing*

Proses yang dilakukan dalam tahapan ini adalah sebagai berikut:

- a) *Text Cleaning*, yaitu menghilangkan tanda baca, menghilangkan URL dan string-string khusus pada Twitter seperti *user* (@username), tanda *retweet* (RT), dan *hashtag* (#hashtag).
- b) *Case Folding*, yaitu untuk menyeragamkan bentuk huruf.
- c) *Stopword Elimination*, yaitu proses pembuangan kata-kata yang dianggap tidak memiliki *value*, seperti imbuhan, akhiran, dan kata sambung.
- d) *Synonym Replacement*, yaitu proses penggantian kata ke dalam bentuk sinonimnya.

Hasil yang diperoleh dari tahapan ini adalah kumpulan *tweet* yang sudah bersih dari hal-hal yang tidak diperlukan, sebagaimana penjelasan di atas.

- 3) Pelabelan adalah tahapan dimana *tweet* diberi label yang nantinya akan digunakan pada proses *training* di tahap klasifikasi. Terdapat tiga label yang disediakan yaitu *aktivitas_lokasi* untuk *tweet* yang berisi informasi mengenai aktivitas dan lokasi, *non_aktivitas_non_lokasi* untuk *tweet* yang tidak berisi informasi mengenai aktivitas dan lokasi, dan *lokasi_non_aktivitas* untuk yang hanya berisi informasi mengenai lokasi saja. Pelabelan dilakukan dengan menggunakan program aplikasi Microsoft Excel. Hasil dari tahapan ini adalah kumpulan *tweet* yang memiliki label.

4) Klasifikasi

Proses klasifikasi dibedakan menjadi dua proses yaitu:

- a. *Training*, proses ini digunakan untuk melatih algoritma klasifikasi yang digunakan yaitu algoritma Naive Bayes agar mampu melakukan prosesnya sesuai dengan yang diharapkan. Pada tahap ini pertama-tama akan dilakukan proses pembobotan terhadap kumpulan *tweet* hasil pelabelan menggunakan perhitungan TF-IDF dengan hanya menghitung TF (*term frequency*)-nya saja dan dilanjutkan dengan perhitungan *Add-One (Laplace Smoothing)*. Selanjutnya akan dihasilkan model klasifikasi yang nantinya digunakan pada tahap *testing*. Pada tahap ini juga akan dilakukan proses validasi dan evaluasi terhadap kinerja model klasifikasi yang dihasilkan dengan menghitung nilai F-Measure menggunakan teknik *K-Fold Cross Validation*. Perhitungan nilai F-Measure melibatkan perhitungan nilai *Precision* dan *Recall*.
- b. *Testing*, proses ini dilakukan untuk melakukan pengklasifikasian terhadap dataset dengan memanfaatkan model klasifikasi yang dihasilkan pada proses *training*. Hasil pada tahap ini adalah kumpulan *tweet* yang telah diklasifikasikan ke dalam kelas aktivitas_lokasi yang akan digunakan pada tahapan berikutnya yaitu *Clustering*.

5) *Clustering*

Pada tahap ini data hasil klasifikasi pertama-tama dilakukan pembobotan menggunakan perhitungan TF-IDF dengan hanya menghitung TF (*term frequency*)-nya saja. Selanjutnya dilakukan proses *clustering* dengan algoritma K-Means untuk mengelompokkan data sesuai dengan informasi aktivitas yang ada di dalamnya. Pada tahap ini juga sekaligus dilakukan proses validasi dan evaluasi kualitas hasil dari *clustering* dengan algoritma *K-Means*. Validasi dan evaluasi menggunakan perhitungan nilai *purity*.

6) Visualisasi Informasi

Pada tahap ini informasi yang dihasilkan dari proses *clustering* selanjutnya akan divisualisasikan ke dalam format yang mudah dimengerti oleh para pengguna.

1.2. Metode Penelitian

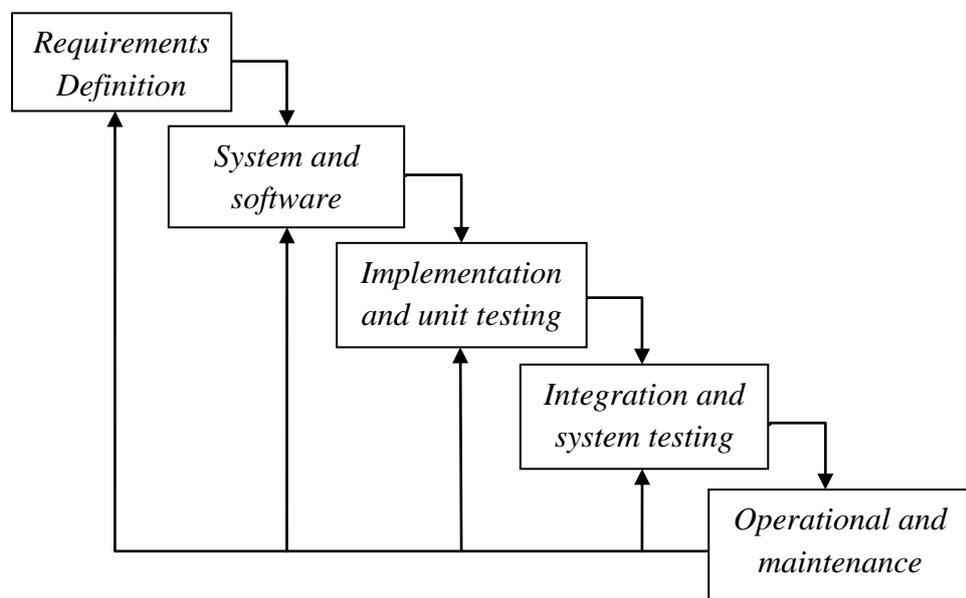
Metode penelitian yang digunakan dalam penelitian ini dibedakan menjadi dua, diantaranya dijelaskan sebagai berikut:

1.2.1. Proses Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari layanan jejaring sosial berbasis lokasi Twitter dengan memanfaatkan *Streaming* API melalui alamat: <https://stream.twitter.com/1.1/statuses/filter.json>.

1.2.2. Proses Pengembangan Perangkat Lunak

Untuk keperluan perancangan, dibutuhkan sebuah metode sebagai acuan proses pengembangan perangkat lunak. Maka dari itu metode perancangan yang digunakan adalah Waterfall model (Pressman, 2010).



Gambar 3.2 Model Waterfall

1) *Requirement Definiton*

Mengumpulkan kebutuhan secara lengkap kemudian dianalisis dan didefinisikan kebutuhan yang harus dipenuhi oleh software yang akan dibangun.

2) *System and Software Design*

Proses pencarian kebutuhan diintensifkan dan difokuskan pada *software*. Bertujuan untuk memudahkan dalam pemahaman terhadap proses yang terjadi, menjelaskan alur sistem dalam perangkat lunak tersebut. Pada proses ini terdapat empat atribut diantaranya struktur data, arsitektur perangkat lunak, representasi antarmuka, dan algoritma prosedural.

3) *Implementation and Unit Testing*

Desain program diterjemahkan ke dalam kode-kode dengan menggunakan bahasa pemrograman yang sudah ditentukan. Program yang dibangun langsung diuji tiap unit.

4) *Integration and System Testing*

Tahap ini merupakan implementasi dari tahap design yang secara teknis nantinya dikerjakan oleh *programmer*. Penyatuan unit-unit program kemudian diuji secara keseluruhan (*system testing*).

5) *Operation and Maintenance*

Tahap ini adalah tahap pemastian *software* bebas dari *error*, dan hasilnya harus benar-benar sesuai dengan kebutuhan yang sudah didefinisikan sebelumnya, terakhir dilakukan pemeliharaan *software*.

1.3. Alat dan Badan Penelitian

Alat yang digunakan dalam penelitian ini adalah seperangkat komputer beserta perangkat lunaknya. Sedangkan bahan penelitian yang digunakan merupakan data yang diperoleh dari layanan jejaring sosial berbasis lokasi Twitter yaitu berupa kumpulan *tweet*.

1.3.1. Alat Penelitian

Alat penelitian yang digunakan dibagi ke dalam dua jenis, yaitu perangkat keras dan perangkat lunak.

- 1) Perangkat Keras
 - a) Intel® Pentium® Dual-Core *processor* T 4500 (2.3GHz, 1MB L2 *Cache*, 800MHz FSB)
 - b) RAM 4 GB
 - c) Harddisk 500 GB
 - d) *Monitor* Resolusi 1366 x 768, 64 bit *colour quality*
 - e) *Mouse* dan *Keyboard*

- 2) Perangkat Lunak
 - a) Windows 8 Profesional
 - b) Netbeans 8.0
 - c) Java SE 7
 - d) XAMPP 1.7.1
 - e) Notepad++

1.3.2. Bahan Penelitian

Bahan penelitian yang digunakan dalam penelitian ini adalah data kumpulan *tweet* yang diperoleh dari Twitter melalui *Streaming API* dengan kumpulan kata kunci (*key word*) berupa nama-nama lokasi yang telah ditentukan sebelumnya. Selain itu digunakan juga data kumpulan sinonim dan *stopword*.