

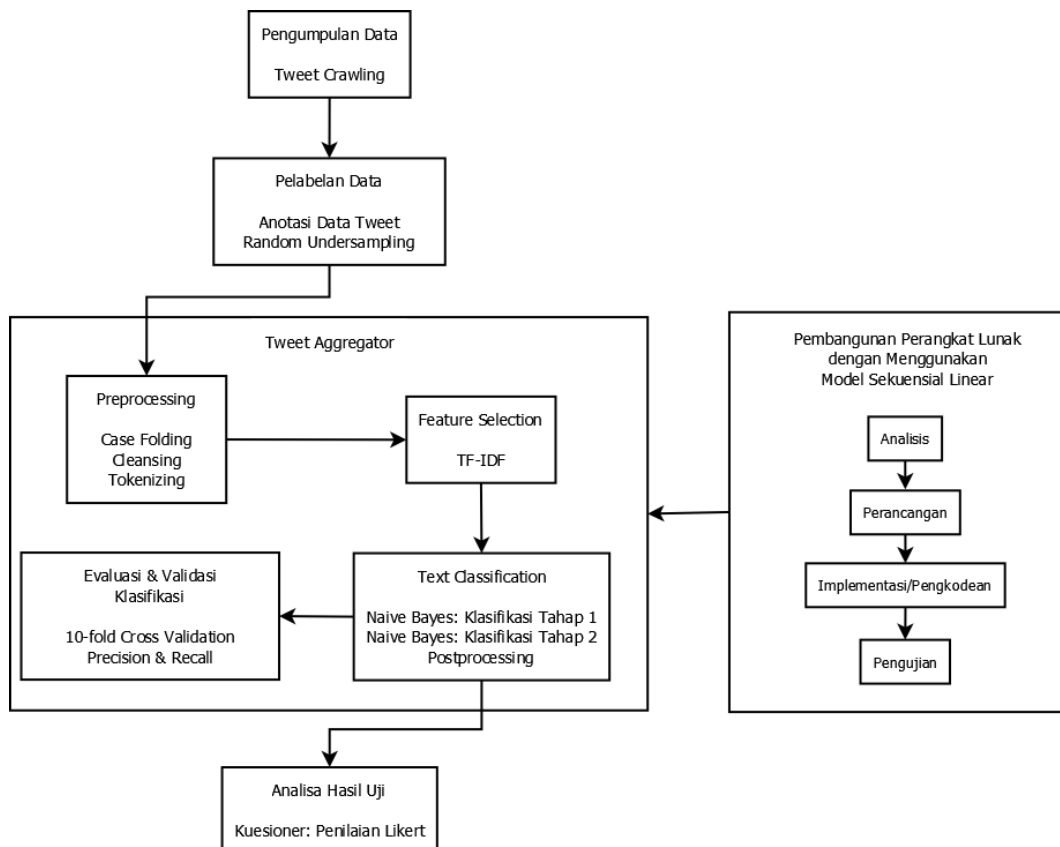
BAB III

METODOLOGI PENELITIAN

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

3.1 Desain Penelitian

Desain penelitian merupakan tahapan yang akan dilakukan penulis untuk memberikan gambaran serta kemudahan dalam melakukan penelitian. Tahapan penelitian yang digunakan dijabarkan pada Gambar 3.1.



Gambar 1.1 Desain Penelitian

Berikut penjelasan dari desain penelitian:

1. Pengumpulan data

Data yang digunakan dalam penelitian ini diambil dari situs jejaring sosial Twitter yang dinamakan *tweet*. *Tweet* yang saat ini penulis dapatkan sekitar 61.620 *tweet* dengan cara melakukan *crawling* dengan memanfaatkan Twitter API. Data diambil dari bulan Februari 2015 sampai Agustus 2015.

Tweet yang diambil merupakan *tweet* pribadi dari beberapa *user* yang berkompeten di masing-masing bidang kategori berita. Kategori berita ini, antara lain: politik dan hukum, bisnis, ekonomi, agama, teknologi, kenegaraan, internasional, pendidikan, kesehatan, olahraga, dan hiburan.

2. Pelabelan

Beberapa data yang telah dikumpulkan dibagi menjadi data *training* dan data *testing*. Data ini dikelompokkan secara manual menurut kelas masing-masing oleh seorang *supervisor* (orang yang mendefinisikan kelas dan label pada *training document*). Data ini nantinya digunakan untuk membentuk fungsi klasifikasi.

Data yang telah dilabeli secara manual ini di-*sampling* dengan menggunakan teknik *random undersampling* untuk mendapatkan jumlah data yang seimbang di setiap kelasnya.

3. *Preprocessing*

Proses yang dilakukan adalah *preprocessing* dasar yang meliputi: (Feinerer & Hornik, 2008)

- 1) *Case Folding*, merupakan proses penyeragaman bentuk semua kata menjadi *lowercase*.
- 2) *Cleansing*, merupakan proses pembersihan data teks dari hal-hal yang tidak diperlukan
- 3) *Tokenizing*, merupakan proses pemecahan *string* berdasarkan tiap kata yang menyusunnya.
- 4) *Stopword Removal*, merupakan penghapusan *stopwords* (kata-kata yang tidak deskriptif).

5)Penyamaan Sinonim, merupakan penyamaan kata-kata dengan sinonimnya.

Proses ini kemudian dimodifikasi oleh peneliti untuk disesuaikan dengan sumber data *tweet* yaitu pada proses *cleansing* ditambahkan proses sebagai berikut:

- a. Menghapus *username* yang muncul pada setiap *tweet* (direpresentasikan sebagai @username) dihapus.
- b. Menghapus tanda “RT” yang muncul di setiap *tweet* dihapus (RT merupakan tanda bahwa sebuah *tweet* merupakan hasil *retweet*).
- c. Menghapus format *hashtag* yang muncul pada setiap *tweet* (direpresentasikan sebagai #hashtag) dihapus.
- d. Menghapus penghilangan tanda baca.
- e. Menghapus kemunculan angka atau *Clean Number*.

4. *Feature Selection*

Pada tahap ini dilakukan proses seleksi fitur dari kumpulan term yang didapat dari *preprocessing*. *Feature selection* dilakukan dengan memberi bobot pada masing-masing *term* menggunakan metode *Term Frequency * Inverse Document Frequency* (TF*IDF). Fitur yang dipilih merupakan *term* dengan bobot yang lebih besar.

5. *Text Classification*

Pada tahap ini dilakukan proses pengelompokkan *tweet* berdasarkan kelas-kelas yang telah ditentukan dengan menggunakan algoritma *Naïve Bayes*. Proses pengelompokkan dilakukan sebanyak 2 tahap. Tahap pertama untuk mengelompokkan *tweet* berita dan bukan berita, kemudian tahap kedua dilakukan untuk mengelompokkan *tweet* ke dalam kelas yang telah ditentukan, yaitu: politik dan hukum, bisnis, ekonomi, agama, teknologi, kenegaraan, internasional, pendidikan, kesehatan, olahraga, dan hiburan.

Tweet hasil pengklasifikasian dibersihkan dari *username*, tanda *retweet* (RT), dan format *hash tag* untuk menghasilkan *tweet* yang lebih merepresentasikan informasi berita.

6. Evaluasi dan Validasi Klasifikasi

Pada tahap ini dilakukan pengujian kualitas dari hasil dari *classification* dengan menggunakan metode *k-fold cross validation*. Penulis menggunakan *10-fold cross validation* dengan membagi data yang telah dianotasikan menjadi 10 *subset* dan melakukan proses *training* dan *testing* secara berulang-ulang pada data yang telah dibagi tersebut. Untuk mengukur tingkat ketepatan dan keberhasilan hasil klasifikasi dilakukan perhitungan *precision* dan *recall* untuk masing-masing kelas.

7. Analisa Hasil Uji

Pada tahap ini dilakukan pengujian aplikasi untuk mengetahui tingkat kepuasan pengguna terhadap aplikasi yang dibangun. Pengujian ini dilakukan dengan menyebarkan kuesioner dan dinilai dengan menggunakan perhitungan skala *Likert*.

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 dengan menggunakan Twitter API. *Tweet* diambil pada rentang bulan Februari 2015 sampai bulan Agustus 2015. *Tweet* ini diambil dari 100 *user* yang kompeten di masing-masing kategori klasifikasi. Kategori-kategori tersebut, antara lain: politik dan hukum, bisnis, ekonomi, agama, teknologi, kenegaraan, internasional, pendidikan, kesehatan, olahraga, dan hiburan. Sedangkan *user* yang berkompeten didapatkan dari Ensiklopedi Tokoh Indonesia Online pada tautan <http://www.tokohindonesia.com/> dan daftar wartawan terkenal Indonesia.

3.2.2. Proses Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang penulis gunakan pada penelitian ini adalah *Linear Sequential Model* yang disebut juga dengan *classic*

life cycle atau *waterfall model*. *Linear sequential model* merupakan proses pengembangan perangkat lunak yang sistematis dengan pendekatan-pendekatan sekuensial. Metode ini merupakan pengembangan dari *conventional engineering cycle* dan terdiri dari beberapa aktivitas, sebagai berikut: (Pressman, 2001)

1) *Pemodelan Sistem / Information Engineering*

Tahap penyusunan *requirements* untuk keseluruhan elemen sistem dan mengalokasikan *subset* untuk setiap *requirements* tersebut. Pemodelan ini melibatkan proses analisis dan perancangan.

2) *Analisis Kebutuhan Perangkat Lunak*

Karena proses pengumpulan *requirements* terfokus pada *software*, maka *software engineer* (“*analyst*”) harus mengerti domain informasi dari *software* tersebut, seperti: fungsi, sifat, performa, dan *interface*-nya.

Pada tahap ini dilakukan pengumpulan akun *user* yang memiliki kompetensi di bidang berita, kemudian *tweet* dari akun tersebut ditarik menggunakan Twitter API. Data *tweet* ini akan menjadi masukan dari sistem. Kemudian ditentukan modul-modul pembangun sistem, serta hasil *output* dari sistem.

3) *Perancangan*

Proses *design* menerjemahkan *requirement* menjadi representasi dari *software*. Proses ini terdiri dari beberapa tahap yang terfokus pada atribut dari program, yaitu:

- a. *Data structure*
- b. *Software architecture*
- c. *Interface representations*
- d. *Procedural (algorithmic) detail*

Design ini didokumentasikan dan menjadi bagian dari *software configuration*.

Pada proses *design* penulis merancang dengan menggunakan *Unified Modelling Language* (UML).

Pada tahap ini dibuat desain dari sistem yang akan dikembangkan dalam bentuk *Entity Relationship Diagram* (ERD), UML, serta rancangan antar muka sistem.

4) Pengkodean

Merupakan proses penerjemahan *design* menjadi bentuk yang dapat dibaca oleh komputer (kode program).

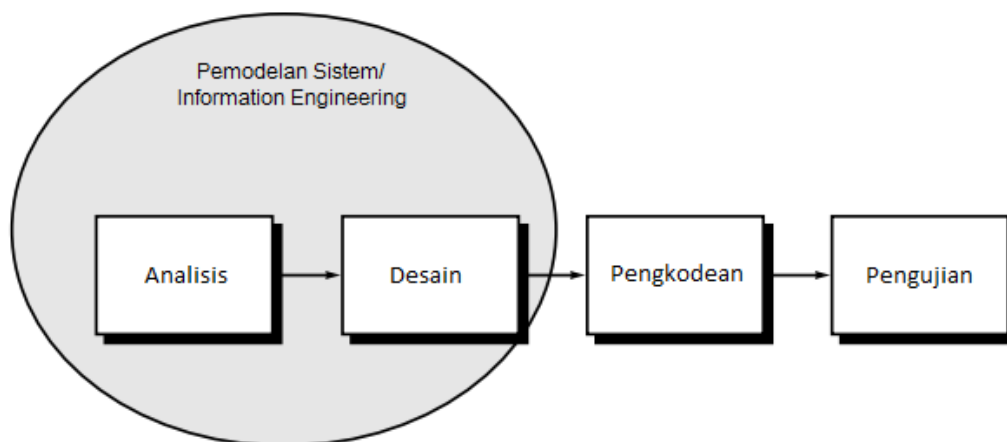
Pada tahap ini dilakukan pengkodean dari semua modul yang sudah didefinisikan sebelumnya menggunakan bahasa pemrograman Java.

5) Pengujian

Setelah proses pengkodean, dilakukan proses *testing*. Proses *testing* terfokus pada *logical internal* dari *software* dan *functional external*, yaitu dengan mengetes semua fungsi untuk menemukan *error* dan memastikan *input* yang telah ditetapkan akan menghasilkan *output* yang sesuai.

Pada tahap ini dilakukan pengujian pada setiap modul untuk memastikan sistem berjalan dengan baik.

Gambaran dari tahapan *Linear sequential model* dapat dilihat pada Gambar 3.2.



Gambar 1.2 *Linear Sequential Model*

sumber: Software Engineering: A Practitioner's Approach Fifth Edition, 2001, hlm.: 29)

3.3 Alat dan Bahan Penelitian

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

Rany Kasman, 2015

PEMBANGUNAN TWEET AGGREGATOR DENGAN MENGGUNAKAN METODE NAÏVE BAYES

Universitas Pendidikan Indonesia | \.upi.edu perpustakaan.upi.edu

3.3.1. Alat Penelitian

Perangkat keras komputer yang digunakan dalam penelitian ini memiliki spesifikasi sebagai berikut:

1. Processor Intel® Core™ i5
2. Memori 4 GB RAM
3. *Hardisk* berkapasitas 720 GB
4. Monitor 14” dengan resolusi 1366x768 pixel
5. *Mouse* dan *keyboard*

Adapun perangkat lunak yang digunakan adalah:

1. Microsoft Windows 8.1 Pro
2. Eclipse 4.3
3. Java EE
4. Java Runtime Environment (JRE 8)
5. Apache Tomcat 8.0
6. Xampp 1.8.3
7. Navicat Premium
8. Sublime Text 2
9. Google Chrome

3.3.2. Bahan Penelitian

Adapun bahan penelitian yang digunakan merupakan data yang diambil dari Twitter. Data yang diambil merupakan *tweet* pribadi dari beberapa *user* yang berkompeten masing-masing kategori berita yang telah ditentukan sebelumnya. Serta bahan lain berupa kumpulan *stopwords* yang diambil dari hasil penelitian yang dilakukan oleh Fadillah Z Talla dengan judul “*A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*” dan kumpulan kata sinonim yang diambil dari *library Tweet Mining* berbahasa Indonesia yang dibuat oleh Yudi Wibisono yang dapat diperoleh pada tautan: <http://www.code.google.com/p/id-tweetmining/>.