

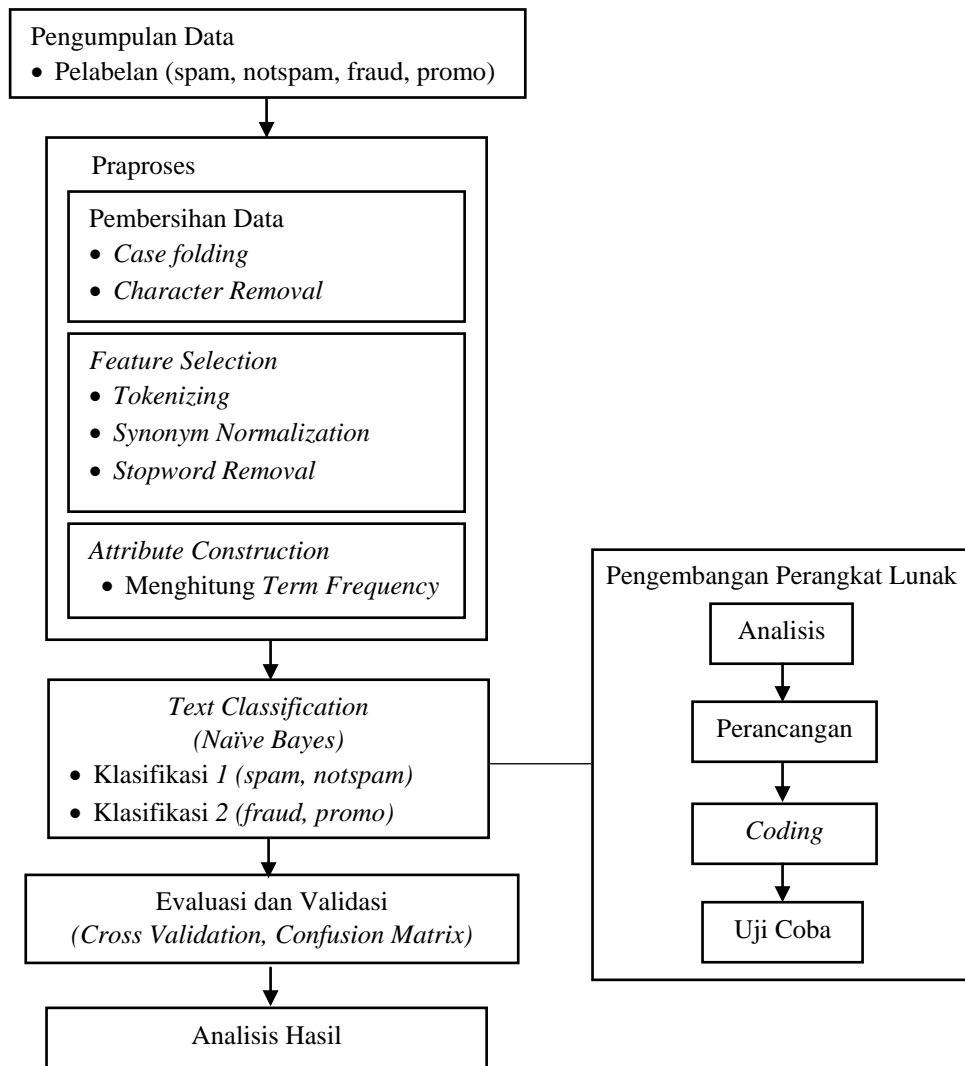
BAB III

METODOLOGI PENELITIAN

Pada bab ini berisi penjelasan langkah-langkah yang akan dilakukan dalam penelitian terdiri dari desain penelitian, metode penelitian, dan alat dan bahan penelitian.

3.1 Desain Penelitian

Desain penelitian merupakan rancangan tahapan yang akan dilakukan dalam penelitian. Berikut ini proses desain penelitian seperti pada gambar 3.1.



Gambar 3.1 Desain Penelitian

Adapun penjelasan mengenai desain penelitian sebagai berikut:

1. Pengumpulan Data

Pengumpulan data dilakukan secara *crowdsourcing* dari para pengguna telepon seluler yang memiliki data SMS spam ataupun notspam. Lalu, studi literatur dilakukan untuk mempelajari perilaku SMS spam. SMS spam dapat dikirimkan secara *broadcast* melalui layanan SMS Gateway yang disediakan provider seluler. Serta memahami karakteristik pesan penipuan dan promosi. Setelah data terkumpul dilakukan pelabelan.

Pelabelan dilakukan oleh seorang supervisor (orang yang memberi label terhadap data, sesuai kelas yang sudah didefinisikan). Data ini nantinya akan digunakan untuk membentuk model untuk melakukan klasifikasi. Adapun kelas yang didefinisikan pada tahap pelabelan ini adalah spam dan notspam untuk data training pertama, sedangkan pada data training kedua yaitu fraud dan promo.

2. Praproses

Tahap praproses terdiri dari proses pembersihan data dan *feature selection* berikut penjelasan masing-masing proses.

a. Pembersihan data

Adapun proses yang dilakukan pada data *cleansing* diantaranya *case folding* dan *character removal*. Proses pertama dilakukan *case folding* dengan mengubah semua karakter huruf menjadi *lowercase* (huruf kecil). Selanjutnya melakukan *character removal* dengan menghilangkan semua karakter yang dianggap tidak diperlukan, yaitu semua karakter selain huruf (a-z).

b. Feature Selection

- 1) *Tokenizing* dilakukan untuk mendapatkan kata-kata dari dokumen, yang merupakan cikal bakal atribut.
- 2) *Synonym Normalization* merupakan tahapan melakukan normalisasi terhadap kata yang memiliki makna yang sama. Proses ini dilakukan agar memperkecil

dimensi data dan mendapatkan attribut yang berkualitas. Proses dilakukan dengan melakukan *filter* setiap kata dengan *database* yang telah disediakan.

- 3) *Stopword Removal* menghilangkan kata yang tidak mempengaruhi kelas klasifikasi, seperti kata hubung, kata sandang, kata ganti, dll. Proses dilakukan dengan melakukan *filter* setiap kata dengan *database* yang telah disediakan.

c. *Attribute Constuction*

Attribute Constuction dapat dilakukan dengan menghitung *term frequency* atau jumlah suatu kata yang terdapat dalam satu dokumen.

3. *Text Classification*

Pada tahap ini dilakukan dua tahap klasifikasi. Klasifikasi pertama mengklasifikasi SMS merupakan kelas spam atau notspam. Jika klasifikasi pertama merupakan spam, maka data akan melakukan klasifikasi kedua menentukan jenis spam yaitu fraud atau promo. Jika klasifikasi pertama merupakan kelas notspam maka, proses selesai dengan hasil kelas notspam.

4. Evaluasi dan Validasi

Evaluasi dan validasi dilakukan menggunakan metode *k-fold cross validation*. Data dianotasikan menjadi sebanyak k subset. Proses pengujian *k-fold cross validation* akan dilakukan sebanyak k kali proses *training* dan k kali proses *testing*. Dari setiap pengulangan didapatkan hasil akurasi, *recall*, *precision*, dan *f-measure*. Pada tahap ini penulis menggunakan *10-fold*.

5. Analisis Hasil

Pada tahap ini dilakukan analisis untuk penarikan kesimpulan dari seluruh proses penelitian.

6. Pembangunan Perangkat Lunak

Pembangunan perangkat lunak terdiri dari beberapa proses, yaitu analisis, perancangan, *coding*, dan uji coba.

3.2 Metode Penelitian

Adapun metode yang digunakan dalam penelitian ini dibagi menjadi dua, yaitu metode pengumpulan data dan metode pengembangan perangkat lunak.

3.2.1 Metode Pengumpulan Data

1. *Crowdsourcing*

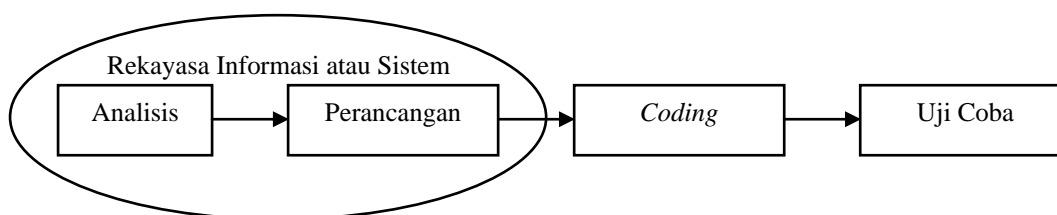
Pengumpulan data dilakukan menggunakan metode *crowdsourcing*. *Crowdsourcing* merupakan metode pengumpulan data memanfaatkan keramaian. *Crowdsourcing* dilakukan dengan memanfaatkan para pengguna telepon seluler yang memiliki SMS spam ataupun notspam.

2. Studi Literatur

Studi literatur merupakan proses mendapatkan teori yang relevan dengan kasus permasalahan. Studi literatur dapat dilakukan terhadap informasi keilmuan mengenai teknik-teknik terkait yang didapatkan dari buku-buku, jurnal, dan artikel ilmiah.

3.2.2 Metode Pengembangan Perangkat Lunak

Pengembangan perangkat lunak dilakukan menggunakan metode *Linear Sequential Model*. *Linear Sequential Model* memiliki proses yang sistematis dengan pendekatan sekuensial. Dimulai dari analisis, perancangan, *coding*, dan pengujian. Metode ini merupakan metode klasik yang juga disebut *life cycle* atau *waterfall*. Berikut ini proses *Linear Sequential Model* pada gambar 3.2 (Pressman, 2001).



Gambar 3.2 *Linear Sequential Model*

Sumber: (Pressman, 2001)

1. Rekayasa Informasi atau Sistem

Ferdila Rahmi, 2016

APLIKASI SMS SPAM FILTERING PADA ANDROID MENGGUNAKAN ALGORITMA NAÏVE BAYES
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Sistem merupakan satu kesatuan yang terdiri dari beberapa elemen saling berhubungan. Sistem perangkat lunak terdiri dari beberapa elemen seperti perangkat keras, pengguna, dan *database*. Suatu perangkat lunak merupakan bagian dari sistem bisnis, yang bekerja membangun kebutuhan informasi untuk semua bagian sistem. Tahap ini *developer* (orang yang membangun perangkat lunak) menentukan permasalahan dan menemukan solusi dengan merekayasa suatu perangkat lunak.

2. Analisis

Tahap analisis menghasilkan spesifikasi kebutuhan perangkat lunak. Sebelumnya, dilakukan analisis untuk dapat mendefinisikan kebutuhan. Analisis pengembangan perangkat lunak dapat dilakukan berdasarkan *information domain*, fungsi, perilaku, *performance*, dan antarmuka dari sistem.

3. Perancangan

Tahap perancangan akan menggambarkan representasi sistem berdasarkan spesifikasi kebutuhan yang telah dikumpulkan. Representasi dapat berupa diagram, *pseudocode*, atau *mockup* yang menjadi acuan ketika tahap *coding*. Desain perangkat lunak dibagi menjadi empat bagian yaitu, struktur data, arsitektur perangkat lunak, representasi antarmuka, dan detail prosedural (algoritma).

4. *Coding* atau Implementasi

Tahap *coding* merupakan tahapan menterjemahkan hasil desain ke dalam bahasa mesin.

5. Uji Coba

Setelah perangkat lunak dibangun dilakukan pengujian. Pengujian fokus pada *logical internal* dan *functional external*, yaitu dengan melakukan pengujian pada semua fungsi untuk menemukan *error* dan memastikan input yang telah ditetapkan memberikan output yang diharapkan.

3.3 Alat dan Bahan Penelitian

Adapun alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

3.3.1 Alat Penelitian

Alat yang digunakan dalam penelitian ini adalah seperangkat komputer dan seperangkat *mobile device* android. Adapun spesifikasi perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

- PC
 1. Processor AMD E-450
 2. Memori 4 GB
 3. Hardisk 300 GB
 4. Monitor 14 inch
 5. Mouse and Keyboard
- *Mobile device* Android
 1. CPU ARM Cortex-A53 1.2GHz Quad-Core
 2. Memori 1 GB
 3. *Storage* 8 GB
 4. LCD 5 inch

Sedangkan spesifikasi perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

- PC
 1. Microsoft Windows 8 Pro-32 bit
 2. Java JDK 1.7.0_07
 3. IntelliJ IDEA
 4. Android Studio
 5. XAMPP 5.6.12
 6. Navicat Premium Trial version
 7. Sqlite Browser

Ferdila Rahmi, 2016

APLIKASI SMS SPAM FILTERING PADA ANDROID MENGGUNAKAN ALGORITMA NAÏVE BAYES
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

8. Microsoft Office 2013
 9. Dia Diagram Editor
 10. Balsamiq Trial
- *Mobile Device* Android 5.02 (Lollipop)

3.3.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini berupa *dataset* yang terdiri dari data *training*, data *stopword*, dan data sinonim:

1. Data *Training*

Jumlah data *training* yang digunakan sebanyak 1140 data, terdiri dari 570 kelas notspam dan 570 kelas spam. Kelas spam yang terdiri dari 325 kelas fraud 245 kelas promo. Data yang diasumsikan sudah dalam keadaan random pada *database*. Data *training* digunakan untuk membangun model dengan algoritma *Naïve Bayes* dan validasi menggunakan *k-fold cross validation*.

2. *Stopword*

Kumpulan data *stopword*.

3. Data sinonim

Kumpulan data sinonim sebagian 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/>.