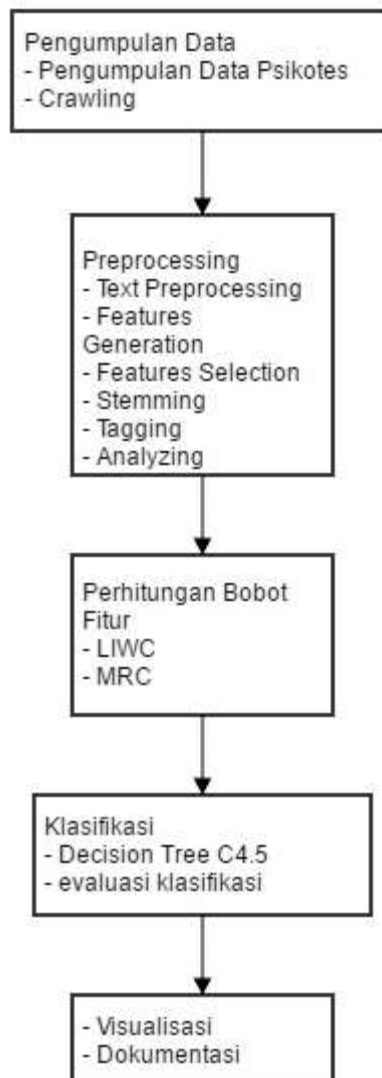


## BAB III METODE PENELITIAN

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 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 dua tahapan yaitu pengumpulan data psikotes dan *crawling* data *tweet* dari Twitter, berikut tahapannya.

- a. Pengumpulan data hasil psikotes, yaitu proses pengumpulan data yang didapat dengan mengambil data hasil psikotes peserta didik dari x-foundation dalam *Entrepreneur Development Scholarship for Youth* yang selanjutnya digunakan dalam mengetahui bagaimana sifat dan karakter dari peserta didik tersebut dan dijadikan sebagai acuan dalam prediksi kepribadian dalam penelitian ini.
- b. Pengumpulan data Twitter, yaitu proses pengumpulan data yang didapat dari media sosial Twitter dengan mengambil kurang lebih 158 akun pengguna Twitter (akun sesuai dengan data hasil psikotes). Pengambilan data akun tersebut dilakukan dengan memanfaatkan REST API Twitter. Data tersebut akan dijadikan sebagai dataset dalam penelitian ini.

## 2. *Preprocessing*

Berikut adalah tahapan yang dilakukan dalam *preprocessing*.

### a. *Text Preprocessing*

Dalam *text processing* dilakukan dua tahapan yaitu *text cleanup* dan *case folding*, *text cleanup* adalah proses pembersihan data teks yang tidak diperlukan seperti menghilangkan url, menghilangkan *hashtag* (#) dan pembuangan *stopwords*. *Case folding* adalah penyeragaman huruf menjadi *lowercase*, serta penghapusan angka dan tanda baca sehingga teks hanya berisi huruf latin a sampai z.

### b. *Features Generation*

*Features generation* adalah proses pengolahan teks dengan cara memisahkan setiap kata atau yang sering disebut *tokenizing*.

### c. *Features Selection*

Tahap ini merupakan tahap mengambil kata-kata penting dari hasil token. Dapat menggunakan algoritma *stop list* (membuang kata yang kurang penting) atau *word list* (menyimpan kata penting).

### d. *Stemming*

Tahap *stemming* adalah tahap mencari root kata dari tiap kata hasil *filtering*.

### e. *Tagging*

Tahap *tagging* adalah tahap mencari bentuk awal/root dari kata hasil *steaming*, dapat juga dengan melakukan penyeragaman makna dari kata-kata yang termasuk dalam daftar kata sinonim bahasa Indonesia.

f. *Analyzing*

Tahap *analyzing* merupakan tahap penentuan seberapa jauh keterhubungan antar kata-kata antar dokumen yang ada.

3. Perhitungan Bobot Fitur

Pada tahapan ini menggunakan fitur gabungan dari *Linguistic Inquiry and Word Count (LIWC) tool* dan *MRC Psycholinguistic Database*. Dengan cara menghitung nilai rata-rata kemunculan dari fitur-fitur tersebut. Nilai rata-rata didapat dari seberapa banyak kemunculan dari fitur-fitur tersebut dibagi dengan jumlah total dari *tweet* pengguna. Tabel 3.1 menjelaskan fitur apa saja yang digunakan dalam penelitian ini.

Tabel 3.1 Fitur Gabungan LIWC dan MRC dalam Penelitian

No	Fitur	No	Fitur
1	<b>Mention:</b> jumlah @ dalam <i>post</i>	18	<b>Swear words:</b> jumlah kata-kata kasar dalam <i>post</i>
2	<b>Hashtag:</b> jumlah # dalam <i>post</i>	19	<b>Family words:</b> jumlah kata-kata yang menunjukkan hubungan kekeluargaan
3	<b>Emoticon positif:</b> jumlah <i>emoticon positif</i> dalam <i>post</i> seperti (^_^)	20	<b>Friend words:</b> jumlah kata-kata yang menunjukkan hubungan pertemanan
4	<b>Emoticon negative:</b> jumlah <i>emoticon negative</i> dalam <i>post</i> seperti (-_-)	21	<b>Emosi senang:</b> jumlah kata-kata yang menunjukkan emosi senang
5	<b>Koma:</b> jumlah tanda koma (,) dalam <i>post</i>	22	<b>Emosi marah:</b> jumlah kata-kata yang menunjukkan emosi marah
6	<b>Titik dua:</b> jumlah titik dua (:)	23	<b>Causation words:</b> jumlah kata-kata yang menunjukkan sebab akibat
7	<b>Titik koma:</b> jumlah titik koma (;) dalam <i>post</i>	24	<b>Hear words:</b> jumlah kata-kata yang berhubungan dengan ekspresi mendengar dalam <i>post</i>
8	<b>Kata ganti orang pertama:</b> jumlah kata ganti orang pertama dalam <i>post</i>	25	<b>Motion words:</b> jumlah kata-kata yang menunjukkan ekspresi bergerak seperti memanjat, berenang, berlari
9	<b>Kata ganti orang kedua:</b> jumlah kata ganti orang kedua dalam <i>post</i>	26	<b>Space words:</b> jumlah kata-kata yang menunjukkan ukuran benda atau jarak dalam <i>post</i>
10	<b>Kata ganti orang ketiga:</b> jumlah kata ganti orang ketiga dalam <i>post</i>	27	<b>Time words:</b> jumlah kata-kata yang menunjukkan keterangan waktu dalam <i>post</i>

11	<b>Preposisi:</b> jumlah preposisi dalam <i>post</i>	28	<b>Money words:</b> jumlah kata-kata yang berhubungan dengan uang
12	<b>long words:</b> jumlah kata yang lebih panjang dari 6 huruf pada <i>post</i>	29	<b>Death words:</b> jumlah kata-kata yang berhubungan dengan kematian
No	Fitur	No	Fitur
13	<b>mean word frequency:</b> rata-rata sederhana dari frekuensi kata pada <i>post</i>	30	<b>Assent words:</b> jumlah kata-kata yang menunjukkan ekspresi setuju
14	<b>type/token ratio:</b> perhitungan rata-rata banyaknya jumlah kata yang telah digunakan dalam <i>post</i>	31	<b>url:</b> jumlah url dalam <i>post</i>
15	<b>Konjungsi:</b> jumlah kata-kata konjungsi dalam <i>post</i>	32	<b>Protected:</b> status akun pengguna apakah di kunci ( <i>protected</i> ) atau tidak
16	<b>Adverb:</b> jumlah kata keterangan dalam <i>post</i>	33	<b>List:</b> jumlah list dari akun pengguna
17	<b>Negasi:</b> jumlah kata negasi seperti (tidak, tanpa, bukan) dalam <i>post</i>	34	<b>Status:</b> jumlah status yang telah di <i>post</i> oleh pengguna

#### 4. Klasifikasi

Berikut merupakan tahapan dari proses klasifikasi :

##### a. *Decision Tree C4.5*

Pada tahapan ini digunakan algoritma klasifikasi *Decision Tree C4.5* untuk mengelompokkan fitur LIWC dan MRC yang bertujuan untuk mendapatkan pola serta nilai akurasi yang terbaik dalam memprediksi tingkat emosi pengguna.

##### b. Evaluasi Klasifikasi

Pada tahapan ini dilakukan proses evaluasi klasifikasi

#### 5. Visualisasi

Pada tahapan ini digunakan visualisasi untuk menampilkan prediksi kepribadian dari pengguna dengan cara memasukkan *username* dari akun pengguna tersebut.

## 3.2 Metode Penelitian

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

### 3.2.1 Metode Pengumpulan Data

Pada penelitian ini menggunakan dua tahapan pengumpulan data, yaitu pengumpulan data hasil psikotes dan pengumpulan data akun Twitter. Pengumpulan data hasil psikotes, yaitu proses pengumpulan data yang didapat dengan mengambil data hasil psikotes peserta

Nur Hadi Kusumo, 2016

PREDIKSI TINGKAT KEMATANGAN EMOSIONAL SESEORANG MELALUI AKTIVITAS DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA C4.5

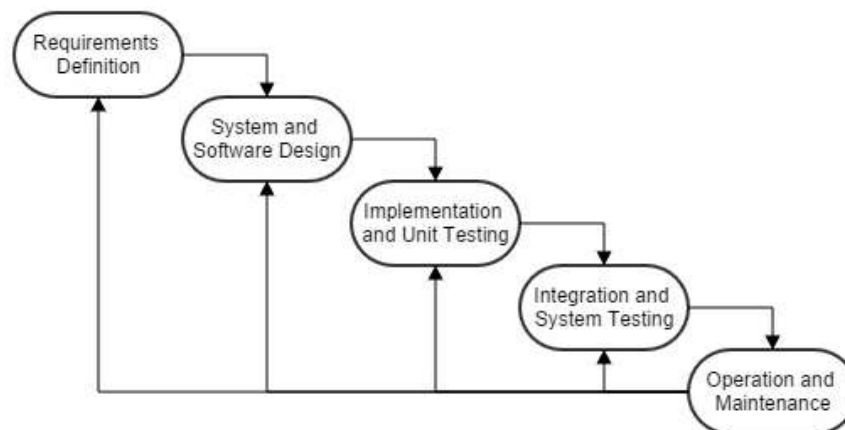
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

didik dari Mien R. Uno Foundation dalam *Entrepreneur Development Scholarship for Youth* (Envoy) yang selanjutnya digunakan dalam mengetahui bagaimana sifat dan karakter dari peserta didik tersebut dan dijadikan sebagai acuan dalam prediksi kepribadian dalam penelitian ini. Pada data ini terdapat sebuah hasil psikotes yang berupa alat ukur mengenai kematangan emosi dengan rentang nilai 1 sampai 5, semakin besar nilainya maka semakin baik pula tingkat kematangan emosinya dan data nilai ini yang menjadi acuan pada penelitian ini.

Pengumpulan data Twitter, yaitu proses pengumpulan data yang didapat dari media sosial Twitter dengan mengambil 158 akun pengguna Twitter (akun sesuai dengan data hasil psikotes), sebenarnya data dalam hasil psikotes tersebut sebanyak 197 orang, akan tetapi dalam penelitian ini hanya mengambil 158 akun dikarenakan ada beberapa orang yang tidak memiliki akun twitter atau akun twitternya tidak aktif. Akun dari pengguna Twitter yang diambil merupakan akun pengguna yang tersebar di 4 kota besar di Pulau Jawa, yaitu Bandung, Semarang, Yogyakarta dan Surabaya. Bahasa yang digunakan setiap akun pengguna juga cukup variatif, yaitu Bahasa Indonesia, Bahasa Sunda, Bahasa Jawa dan Bahasa Inggris. Rentang usia akun pengguna Twitter pada penelitian ini antara usia 19-25 tahun. Pengambilan data akun tersebut dilakukan dengan memanfaatkan REST API Twitter, data yang didapat berbentuk format JSON yang selanjutnya dimasukkan ke dalam *database* untuk proses lebih lanjut. Data tersebut dijadikan sebagai dataset dalam penelitian ini.

### **3.2.2 Metode Pengembangan Perangkat Lunak**

Metode pendekatan perangkat lunak yang digunakan dalam pengembangan perangkat lunak ini adalah pendekatan model *waterfall*. Model *waterfall* pertama kali diperkenalkan oleh Royce pada tahun 1970 (Petersen, dkk. 2009) dengan tujuh tahapan beruntun dan memiliki *feedback loop* antar tahapan jika diperlukan. Model ini mengalami beberapa perubahan diantaranya adalah perubahan tahapan dari tujuh menjadi lima tahapan (Pressman, 2010). Pembangunan perangkat lunak dalam penelitian ini menggunakan model *waterfall* terbaru yaitu versi Sommerville 2011. Dalam model *waterfall* Sommerville terdapat kemungkinan untuk kembali ke tahap sebelumnya apabila terjadi kesalahan atau perbaikan, dimana alur prosesnya seperti gambar 3.2.



Gambar 3.2 Model Waterfall (Sommerville, 2011)

a. *Requirements Definition*

Pada tahap ini merupakan tahapan analisis untuk menentukan kebutuhan, batasan dan tujuan dari perangkat lunak. Hal ini kemudian didefinisikan secara rinci dan terbentuk sebagai spesifikasi sistem. Pada tahap ini dilakukan pengumpulan data *tweet* dari akun pengguna yang telah ditentukan sebelumnya dan menentukan algoritma apa yang akan digunakan pada penelitian ini.

b. *System and Software Design*

*System design* adalah proses perancangan perangkat keras maupun perangkat lunak yang dibutuhkan untuk menunjang sistem yang akan dibangun. *Software design* adalah proses perancangan yang melibatkan dan menggambarkan dasar sistem serta hubungan satu sama lain. Pada tahap ini dibuat desain dari sistem yang akan dikembangkan yaitu sistem prediksi tingkat kematangan emosional.

c. *Implementation and Unit Testing*

Pada tahap ini, *system design* yang telah dibangun kemudian diimplementasikan dalam bentuk unit program, kemudian dilakukan *testing* untuk memastikan unit program tersebut berjalan dengan baik.

d. *Integration and System Testing*

Setelah unit program berjalan dengan baik dan lolos *testing* maka dilanjutkan dengan mengintegrasikan setiap unit program sehingga sesuai dengan sistem yang diinginkan. Setelah sistem sudah terbentuk lalu dilakukan *testing* ulang pada sistem untuk memastikan sistem dapat berjalan dengan baik secara keseluruhan.

e. *Operating and Maintenance*

Pada tahap ini merupakan tahap dimana sistem sudah dijalankan, pada tahap ini juga dapat melakukan perbaikan ketika terjadi kesalahan atau *error* yang tidak ditemukan sebelumnya. Perbaikan juga dapat dilakukan jika sistem memerlukan kebutuhan baru.

### 3.3 Alat dan Bahan Penelitian

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

#### 3.3.1 Alat Penelitian

Dalam penelitian ini, peneliti menggunakan beberapa alat bantu penunjang baik berupa perangkat keras maupun perangkat lunak. Perangkat keras yang digunakan dalam pembangunan sistem ini adalah laptop dengan spesifikasi sebagai berikut:

- a. *Processor* Intel® Core™ i3 CPU
- b. RAM 4 Ghz
- c. *Hardisk* 500 GB HDD
- d. *Mouse*

Sementara itu Perangkat lunak yang digunakan dalam pembangunan sistem ini adalah sebagai berikut:

- a. Sistem Operasi Microsoft Windows 10 Pro 64-bit
- b. JDK (Java Development Kit)
- c. JRE (Java Runtime Environment)
- d. XAMPP 3.2.1
- e. Google Chrome
- f. Notepad++
- g. Chrome Gliffy
- h. Navicat Premium
- i. MySQL

#### 3.3.2 Bahan Penelitian

Bahan penelitian yang digunakan pada penelitian ini adalah *tweet* yang didapat dari *crawling* data akun pengguna yang telah mengikuti psikotes x-foundation dalam *Entrepreneur Development Scholarship for Youth* dari tahun 2013 sampai tahun 2015.

Nur Hadi Kusumo, 2016

**PREDIKSI TINGKAT KEMATANGAN EMOSIONAL SESEORANG MELALUI AKTIVITAS DI MEDIA SOSIAL  
TWITTER MENGGUNAKAN ALGORITMA C4.5**

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

Nur Hadi Kusumo, 2016

*PREDIKSI TINGKAT KEMATANGAN EMOSIONAL SESEORANG MELALUI AKTIVITAS DI MEDIA SOSIAL  
TWITTER MENGGUNAKAN ALGORITMA C4.5*

Universitas Pendidikan Indonesia | [repository.upi.edu](http://repository.upi.edu) | [perpustakaan.upi.edu](http://perpustakaan.upi.edu)