BAB 3
ANALISIS KEBUTUHAN IMPLEMENTASI ALGORITMA

3.1 Analisis Masalah

Analisis masalah adalah suatu gambaran masalah yang diangkat dalam penulisan skripsi tentang Analisis sentimen pengguna twitter pada akun resmi Samsung Indonesia dengan metode *Naive Bayes*. Analisis Sentimen termasuk kedalam kategorisasi teks yang mengelompokan teks-teks ke dalam label-label tertentu. Pada penelitian-penelitian sebelumnya analisis sentimen dimanfaatkan untuk evaluasi sebuah produk yang didapatkan dari *feedback* sebuah produk.

Samsung merupakan sebuah produk yang menggunakan Twitter sebagai salah satu sosial media untuk sarana promosi bisnis. Dengan adanya akun resmi Samsung Indonesia dan jumlah postingan yang banyak setiap harinya memungkinkan terdapat *feedback* dari costumer atau *follower* Samsung Indonesia sebagai ketersediaan senitmen yang dapat dimanfaatkan untuk evaluasi produk Samsung.

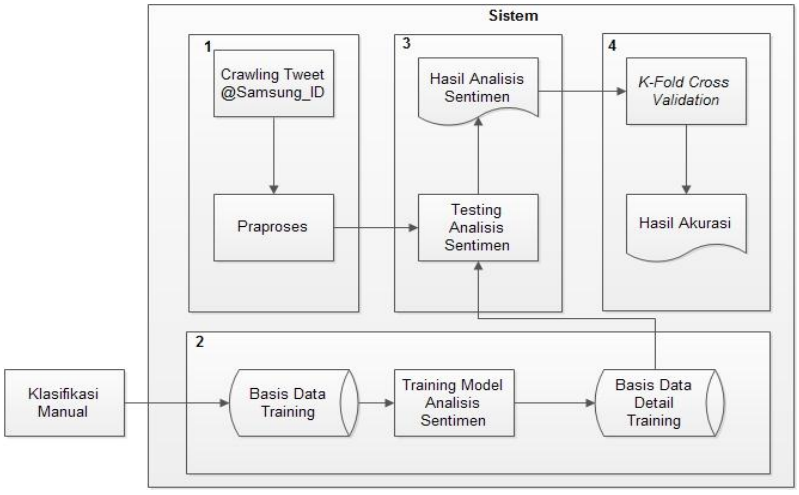
Namun permasalahan pada analisis sentimen adalah bagaimana melakukan pengklasifikasian sentimen pada data twitter yang naninya dapat dimanfaatkan untuk evaluasi sebuah produk dengan menggunakan metode *Naïve Bayes*.

3.2 Analisis Sistem

Analisis sistem didefinisikan sebagai penguraian dari sistem utama ke dalam sub-sub sistem dengan tujuan untuk mengidentifikasikan permasalahan-permasalahan yang ada dan kebutuhan-kebutuhan yang diperlukan agar dapat diusulkan dan diciptakan sistem baru yang lebih baik. Dalam mengimplementasikan metode *Naïve Bayes Classifier* akan dibangun sebuah prototype dengan gambaran sistem yang akan tertera pada Gambar 3.1.

Sistem yang akan dibangun memiliki 4 tahap yaitu praproses, training data, testing, dan *cross validation*. Untuk penjelasan setiap tahap dapat dilihat sebagai berikut:

1. Tahap pertama adalah tahap mengambil data twitter dengan crawling menggunakan stream API, kemudian data tersebut dilakukan proses *preprocessing* yang merupakan tahap dimana data yang akan digunakan sebagai data testing dibersihkan dari *noise* atau dari hal yang tidak mempunyai pengaruh dalam sentimen. *Preprocessing* yang dilakukan dalam tahap ini antara lain: tokenisasi, normalisasi fitur, *case folding*, hapus *stopword*,*Convert Emoticon*, *Convert* negasi.
2. Tahap yang kedua adalah tahap training data, pada tahap ini data training yang sudah dilakukan tahap *preprocessing* dan telah diklasifikasikan secara manual kedalam 2 kelas yaitu kelas sentimen positif, dan negatif disimpan kedalam *database* data training. Setelah itu *tweet* yang sudah diklasifikasikan secara manual dibuat model probabilistiknya yang nantinya akan disimpan kembali ke dalam database detail data training yang akan digunakan pada tahap testing sebagai model analisis sentimen.
3. Tahap yang ketiga yaitu tahap testing, pada tahap ini menggunakan data testing sebagai data masukan yang tentu saja sudah melalui tahap *preprocessing*. Data testing diproses sehingga dapat dilihat kelas sentimennya.
4. Tahap keempat yaitu tahap pengujian menggunakan metode *k-fold cross validation* dimana data dibagi sejumlah n-*fold* yang diinginkan,kemudian proses testing dan training dilakukan sebanyak n kali . Sampai hasil akurasi keluar, hasil akurasi tersebut didapat dari rata-rata akurasi pada setiap iterasi.



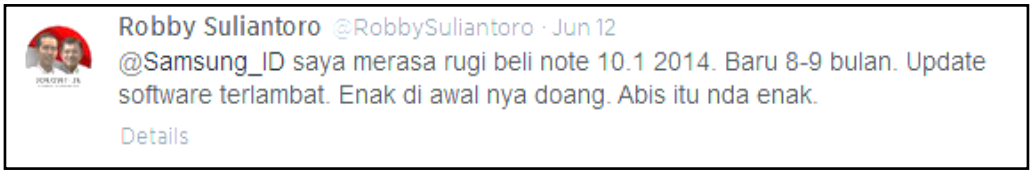
Gambar 3. 1 Gambaran Sistem

3.3 Analisis Data Masukan

Data masukan yang digunakan adalah data *tweet* dari akun twitter resmi Samsung Indonesia. Data *tweet* tersebut didapat dengan memanfaatkan fitur *API* (*Application Interface*) yang telah disediakan oleh twitter untuk mendapakan *API key*, dan *access token* yang nantinya digunakan untuk pengambilan data *tweet.* Data yang diambil adalah data *tweet* yang mengandung kata "Samsung_ID" atau data *tweet* yang terdapat pada akun resmi Samsung Indonesia "@Samsung_ID", data tersebut dapat dianggap mewakili sentimen dari pengguna atau *followers* Samsung Indonesia. Data berupa sebuah kalimat dengan panjang maksimal 140 karakter. Contoh dari setiap *tweet* yang mengandung sentimen pada akun "@Samsung_ID" terdapat pada Gambar 3. 2, Gambar 3. 3.



Gambar 3. 2*Tweet* dengan sentimen positif



Gambar 3. 3*Tweet* dengan sentimen negatif

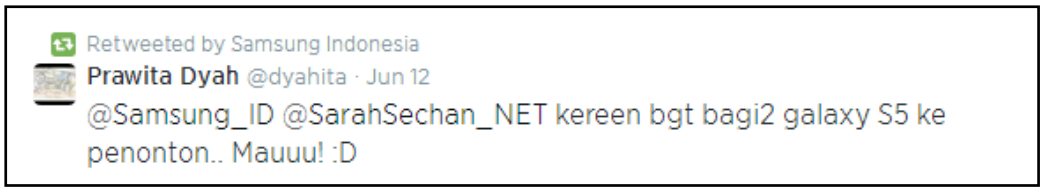
Data *tweet* yang diambil dari akun resmi Samsung Indonesia masih berupa data mentah dan terdapat noise atau ciri-ciri yang tidak mempunyai pengaruh pada klasifikasi sentimen seperti link, "@", *stopword, hashtag* yang ditandai dengan munculnya karakter "#". Contoh *tweet* mentah tersebut ada pada Gambar 3. 4.



Gambar 3. 4 Contoh data *tweet*

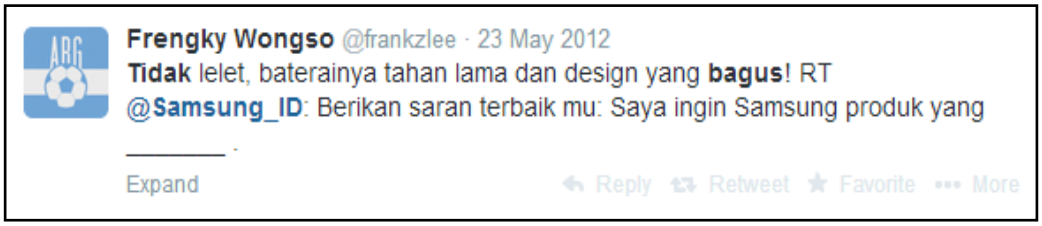
Selain itu, karateristik *tweet* yang dirasa mempunyai pengaruh pada penentuan sentimen adalah terdapat *emoticon* yang merupakan salah satu cara

untuk mengekspresikan ungkapan persetujuan atau pertidaksetujuan pada suatu kalimat atau *tweets.* Contoh *tweet* mentah yang terdapat *emoticon* ada pada Gambar 3. 5.



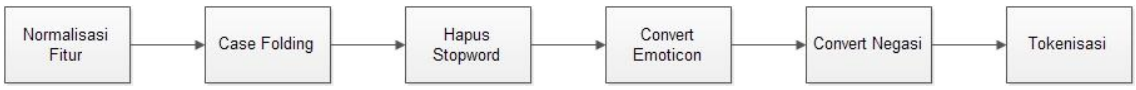
Gambar 3. 5 Contoh data *tweet* dengan *emoticon*

Karakteristik *tweet* selanjutnya yaitu ada terdapat kata negasi yang perlu diperhatikan dalam analisis sentimen karena dapat merubah nilai sentimen suatu *tweet*. Kata yang bersifat negasi seperti "tidak", "bukan", "ga", "jangan", "nggak", "tak", "tdk", dan "gak". Contohnya kata "lelet" yang merupakan kata dengan sentimen negatif, karena di depan kata "lelet" terdapat kata negasi "tidak" maka sentimen dari gabungan kata tersebut adalah positif. Contoh *tweet* yang terdapat kata negasi ada pada Gambar 3.6.



Gambar 3. 6 Contoh data tweet dengan negasi

Agar data-data *tweets* tersebut dapat dimanfaatkan dengan baik untuk mengklasifikasikan sentimen maka diperlukan proses *preprocessing* . Pada proses ini data tweet yang digunakan untuk data training data data testing dibersikan dari *noise* atau ciri-ciri yang tidak berpengaruh pada klasifikasi sentimen seperti link, "*@", "RT", stopword,* dan lain sebagainya. Gambaran proses *preprocessing* dapat dilihat pada Gambar 3. 7.

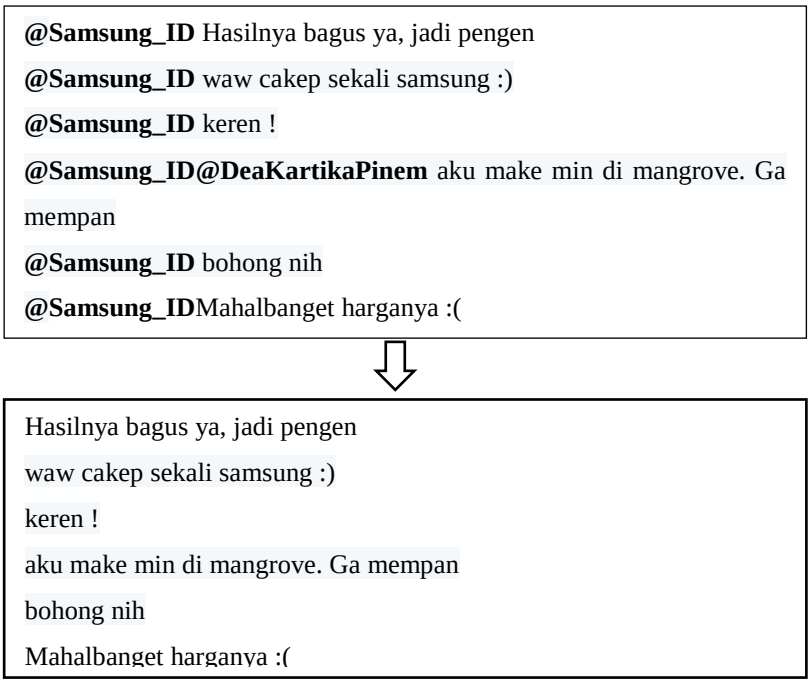


Gambar 3. 7 Tahapan *Preprocessing*

Proses *preprocessing* terdiri dari beberapa tahapan yaitu normalisasi fitur, *case folding*, hapus *stopwords*, *convert emoticon, convert* negasi dan tokenisasi. Penjelasan dari tahapan proses tersebut adalah sebagai berikut:

1. Normalisasi Fitur

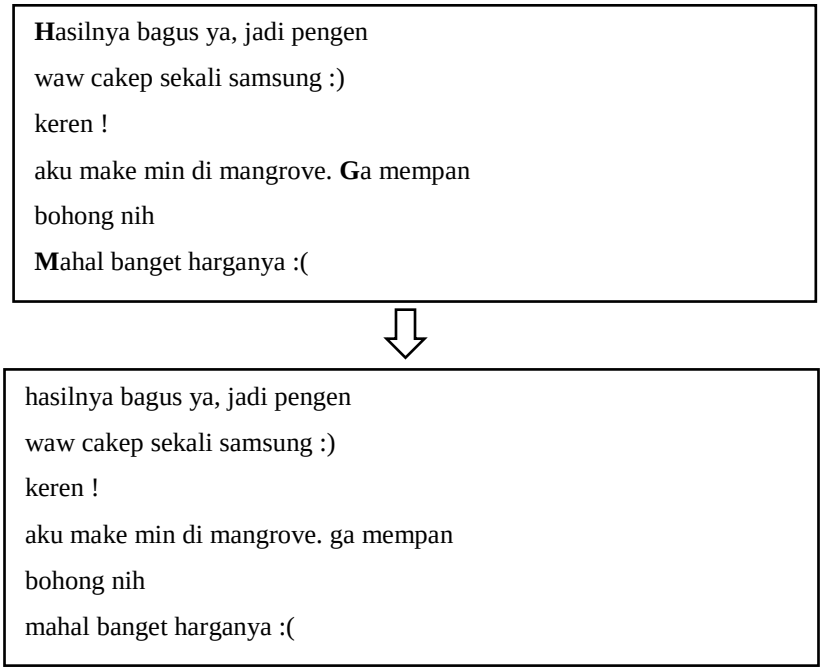
   Tweet yang terdapat pada akun resmi Samsung Indonesia memiliki berbagai komponen atau karakteristik tweet yang khas seperti "@" yang diidentifikasi sebagai komponen *username,* URL yang dikenal melalui operasi regular, hashtag yang menandakan kata sebagai topik yang sedang dibicarakan, dan "RT" yang diidentifikasi sebagai mengulang kembali *tweet* yang telah diposting. Komponen-komponen tersebut tidak memiliki pengaruh apapun terhadap sentimen, maka akan dibuang. Sebagai gambaran dari normalisasi fitur pada salah satu *tweet* akun resmi Samsung Indonesia terdapat pada Gambar 3.8.



Gambar 3. 8 Normalisasi Fitur

2. *Case Folding*

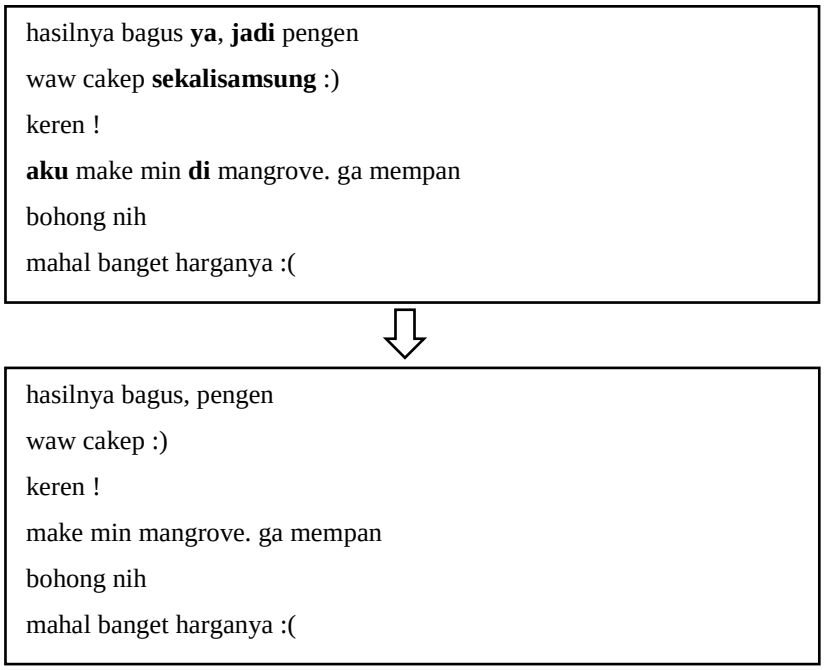
Pada proses *case folding* huruf besar atau *uppercase* yang terdapat pada *tweet* diubah menjadi *lowercase* atau huruf kecil. Sebagai gambaran dari proses *case folding* terdapat pada Gambar 3.9.



Gambar 3. 9 *Case Folding*

3. Hapus *Stopword*

Data *tweet* yang sudah melalui proses sebelumnya masih mengandung kata yang dianggap tidak dapat memberikan pengaruh dalam menentukan suatu kategori sentimen. Kata-kata tersebut dimasukkan kedalam daftar stopword yang biasanya berupa kata ganti orang, kata ganti penghubung, pronomial penunjuk, dan lain sebagainya. Jika terdapat kata yang tercantum pada daftar *stopword* maka kata tersebut dihilangkan,daftar *stopword* tersebutterdapat pada lampiran B. Gambaran dari proses Hapus *Stopword* terdapat pada Gambar 3.10.



Gambar 3. 10 Hapus Stopword

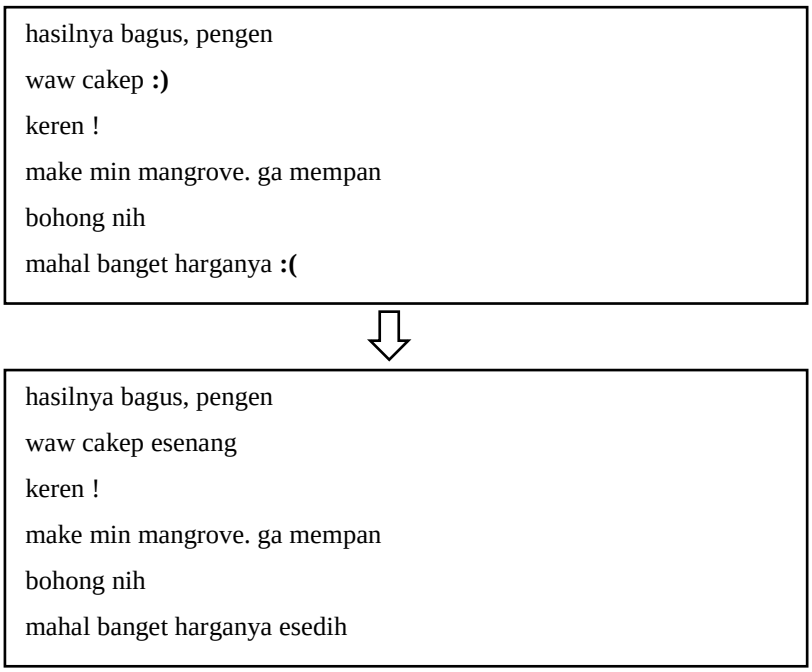
4. *Convert Emoticon*

Convert *emoticon* adalah proses mengkonversikan *emoticon* kedalam string yang sesuai dengan ekspresi *emoticon* itu sendiri. Convert *emoticon* dilakukan karena pada data *tweet* yang diambil dari Samsung Indonesia terdapat *emoticon* yang merupakan salah satu cara mengekspresikan persetujuan atau pertidaksetujuan dalam suatu *tweet*. Hal ini dirasa mempunyai pengaruh terhadap pengklasifikasian sentimen, oleh karena itu *convert emoticon* digunakan. Karena karakter ":)" termsuk kedalam *emoticon* senang , maka dikonversi kedalm string "esenang", begitupun dengan karakter ":(" termasuk kedalam *emoticon* sedih, maka dikonversi kedalam string "esedih". Gambaran dari proses convert *emoticon* terdapat pada Gambar 3.11.

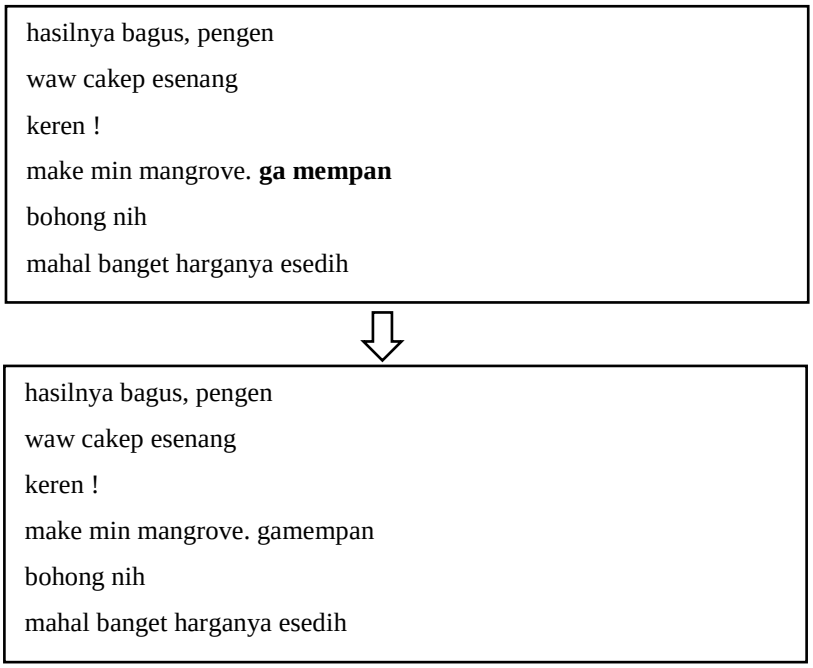
5. *Convert* Negasi

Convert negasi merupakan proses konversi kata-kata negasi yang terdapat pada suatu *tweet*, karena kata negasi mempunyai pengaruh dalam merubah nilai sentimen pada suatu *tweet*. Kata negasi yang terdapat pada suatu *tweet* akan dihilangkan, dan diberikan penanda . Jika terdapat kata negasi makan

akan disatukan dengan kata setelahnya. Kata-kata negasi tersebut meliputi kata “bkn”, “bukan”, “tida”, “tak”, “ga”, “enggak”, “g”, “jangan”, dan “tidak”. Gambaran dari proses *convert* negasi terdapat pada Gambar 3.12



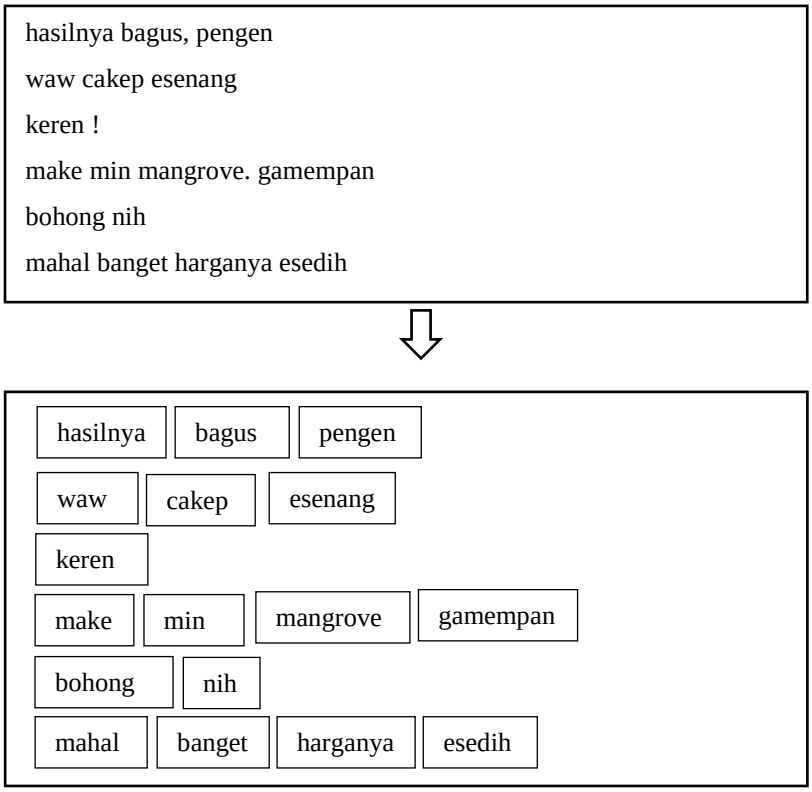
Gambar 3. 11*Convert Emoticon*



Gambar 3. 12*Convert* Negasi

6. Tokenisasi

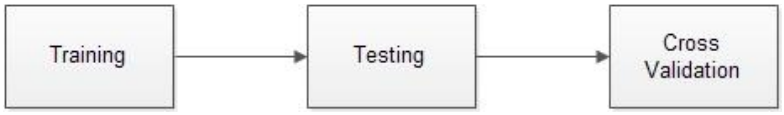
Pada proses tokenisasi setiap kata pada *tweet* dipisahkan, pada proses ini tahap yang dilakukan adalah memisahkan setiap kata yang dipisahkan oleh spasi, selanjutnya bagian *tweet* yang memiliki karakter selain alphabet, angka, dan garis bawah akan dipecah sesuai posisi karakter tersebut dan bagian yang hanya memiliki satu karakter non alphabet dan angka akan dibuang . Gambaran proses tokenisasi terdapat pada Gambar 3.13.



Gambar 3. 13 Tokenisasi

3.4 Analisis Metode / Algoritma

Tahapan-tahapan yang dilakukan untuk analisis sentimen dengan menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan sentimen digambarkan pada Gambar 3. 14.



Gambar 3. 14 Tahapan Analisis Sentimen dengan menggunakan *Naïve Bayes Classifier*

Tahapan pada Gambar 3. 14 terdiri dari tahapan training, testing, dan *cross validation* sebagai pengujian akurasi dari analisis sentimen. Berikut pemaparan lebih jelas tentang tahapan – tahapan yang akan dilakukan pada analisis sentimen dengan menggunakan *Naïve Bayes Classifier*.

3.4.1 Klasifikasi dengan Naïve Bayes

Pada pengklasifikasian menggunakan *naïve bayes* dibagi kedalam 2 proses, yaitu proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentiment dengan data testing atau data mentah yang baru. Berikut adalah algoritma klasifikasi sentimen menggunakan *Naïve Bayes Classifier*:

1. Proses Training
   a. Hitung p (c_i).
   b. Hitung p ($w_k|c_i$) untuk setiap kata w_k pada model.
2. Proses Testing
   a. Hitung $\prod_k$ p ($w_k|c$) × p (c)untuk setiap kategori.
   b. Tentukan kategori dengan nilai $\prod_k$ p ($w_k|c$) × p (c) maksimal.

Berikut merupakan contoh kasus dalam pengklasifikasian *tweet* menggunakan *Naïve Bayes Classifier*. *Tweet* atau dokumen yang akan diklasifikasikan sudah melalui tahap *preprocessing*.

1. Proses Training

Sebuah dokumen *trainer* yang diklasifikasikan secara manual dan dilakukan proses *preprocessing* sebelumnya adalah sebagai berikut:

a. Kata yang termasuk kedalam opini positif:
   Tweet1 : hasilnya bagus pengen
   Tweet2 : waw cakep esenang
   Tweet3 : keren

b. Kata yang termasuk kedalam opini negatif:

Tweet4 : make min mangrove gamempan

Tweet5 : bohong nih

Tweet6 : mahal banget harganya esedih

Gambaran dari contoh kasus ini terlihat pada Tabel 3. 1.

Tabel 3. 1 Gambaran Contoh Kasus

Tweet	Kategori	Fitur
Tweet1	Positif	hasilnya bagus pengen
Tweet2	Positif	waw cakep esenang
Tweet3	Positif	Keren
Tweet4	Negatif	make min mangrove gamempan
Tweet5	Negatif	bohong nih
Tweet6	Negatif	mahal banget harganya esedih

Dibentuk sebuah model probabilistik, dengan mengacu pada persamaan 2.3 dan persamaan 2.5 :

$$P(w_{bagus}\,|c_{positif}) = \frac{\#words(w_{bagus}, c_{positif}) + 1}{\#words(c_{positif}) + |V|} = \frac{1+1}{7+17} = \frac{2}{24}$$

$$P(w_{bagus}\,|c_{negatif}) = \frac{\#words(w_{bagus}, c_{negatif}) + 1}{\#words(c_{negatif}) + |V|} = \frac{0+1}{10+17} = \frac{1}{27}$$

Jika dibentuk sebuah tabel, maka hasil dari perhitungan probabilistik setiap kata pada data training terdapat pada Tabel 3. 2, Tabel 3.3.

Tabel 3. 2 Hasil Perhitungan Probabilitas Data Training (1)

Kategori	$P(c)$	$P(w_{kj} \mid c)$									
		hasilnya	bagus	pengen	waw	cakep	esenang	keren	make	min	mangrove
Positif	$\frac{1}{2}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
Negatif	$\frac{1}{2}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$

Tabel 3. 3 Hasil Perhitungan Probabilitas Data Training (2)

Kategori	$P(c)$	$P(w_{kj} \mid c)$						
		gamempan	bohong	sih	mahal	banget	harganya	esedih
Positif	$\frac{1}{2}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
Negatif	$\frac{1}{2}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$

Hasil perhitungan probabilitas tersebut digunakan sebagai model probabilistik yang nantinya digunakan sebagai penentuan kategori data tester. Contoh kasus data tester yang sudah melalu tahap *preprocessing* sebelumnya terdapat pada Tabel 3. 4.

Tabel 3. 4 Data Tester

Tweet	Kategori	Fitur
Tweet7	?	waw(2) keren bgt
Tweet8	?	ac jelek banget sih
Tweet9	?	keren banget

Untuk penentuan kategori pada tweet7 sebagai data tester 1, menggunakan persamaan 2.6 dengan perhitungan sebagai berikut:

$$P(\text{Tweet7}|C_{positif})$$
$$= P(w_{waw}|C_{positif}) \times P(w_{keren}|C_{positif}) \times P(w_{bgt}|C_{positif}) \times P(C_{positif})$$
$$= \frac{2}{24} \times \frac{2}{24} \times \frac{2}{24} \times 1 \times \frac{1}{2}$$
$$= 0.0002893518$$

$$P(\text{Tweet7}|C_{negatif})$$
$$= P(w_{waw}|C_{negatif}) \times P(w_{keren}|C_{negatif}) \times P(w_{bgt}|C_{negatif}) \times P(C_{negatif})$$
$$= \frac{1}{27} \times \frac{1}{27} \times \frac{1}{27} \times 1 \times \frac{1}{2}$$
$$= 0.0000254026$$

Sedangkan untuk perhitungan tweet8 sebagai file tester 2 sebagai berikut:

$$P(\text{Tweet8}|C_{positif})$$
$$= P(w_{ac}|C_{positif}) \times P(w_{jelek}|C_{positif}) \times P(w_{banget}|C_{positif}) \times$$
$$P(w_{sih}|C_{positif}) \times P(C_{positif})$$
$$= 1 \times 1 \times \frac{1}{24} \times 1 \times \frac{1}{2}$$
$$= 0.02083333334$$

$$P(\text{Tweet8}|C_{negatif})$$
$$= P(w_{ac}|C_{negatif}) \times P(w_{jelek}|C_{negatif}) \times P(w_{banget}|C_{negatif}) \times$$
$$P(w_{sih}|C_{negatif}) \times P(C_{negatif})$$
$$= 1 \times 1 \times \frac{2}{27} \times 1 \times \frac{1}{2}$$
$$= 0.037037037$$

Sedangkan untuk perhitungan tweet9 sebagai file tester 3 sebagai berikut:

$$P(\text{Tweet9}|C_{positif})$$
$$= P(w_{keren}|C_{positif}) \times P(w_{banget}|C_{positif}) \times P(C_{positif})$$
$$= \frac{2}{24} \times \frac{1}{24} \times \frac{1}{2}$$
$$= 0.0017362$$

$$P(\text{Tweet9}|C_{negatif})$$
$$= P(w_{keren}|C_{negatif}) \times P(w_{banget}|C_{negatif}) \times P(C_{negatif})$$
$$= \frac{1}{27} \times \frac{2}{27} \times \frac{1}{2}$$
$$= 0.0013717$$

Setelah menghitung probabilitas dari setiap tweet yang digunakan sebagai data tester didapatkan hasil yang terdapat pada Tabel 3. 5.

Tabel 3. 5 Nilai Probabilitas pada Data Tester

Tweet	Positif	Negatif
Tweet7	**0.0002893518**	0.0000254026
Tweet8	0.02083333334	**0.037037037**
Tweet9	**0.0017362**	0.0013717

Untuk dokumen tweet7 sebagai tester pertama dapat disimpulkan bahwa termasuk kedalam sentiment positif, dilihat dari nilai yang terbesar yaitu terdapat pada sentimen positif. Sedangkan untuk data tester kedua termasuk kedalam sentimen negatif karena dilihat dari hasil yang didapatkan, nilai dengan sentimen negatif lebih besar dibandingkan dengan sentimen positif dan file tester ketiga merupakan sentimen positif karena nilai yang terbesar yaitu pada perhitungan dengan sentimen positif.

3.4.2 *K-Fold cross validation*

Dalam melakukan pengujian dengan menggunakan *k-fold cross validation*, pembagian data traning dan data testing dilakukaan dengan mengambil k sama dengan 3. Pada kombinasi ini data dibagi menjadi 3 *subset* (S_1, S_2,S_3) dengan masing-masing *subset* memiliki anggota yang berbeda. Pada kombinasi ini proses identifikasi akan dilakukan 3 kali iterasi berdasarkan metode *k-fold cross validation.*

Jumlah data yang digunakan mengambil dari data *tweet* pada contoh kasussebelumnya sebanyak 8 data dan dibagi menjadi 3 *subset* sehingga pembagiannya pada masing-masing *subset* yang digunakan adalah *subset* pertama sebanyak 2 data, *subset* kedua sebanyak 3 data dan *subset* ketiga sebanyak 3 data. *Subset* tersebut terdiri dari masing-masing kategori sentimen. Data untuk percobaan yang dilakukan terdapat pada Tabel 3. 6.

Tabel 3. 6 Tabel Data *Tweet*

Subset	No	Data Tweet	Kategori
1	1.	keren	Positif
	2.	mahal banget harganya esedih	Negatif
	3.	keren banget	Positif
2	4.	hasilnya bagus pengen	Positif
	5.	ac jelek banget sih	Negatif
	6.	waw waw keren banget	Positif
3	7.	waw cakep esenang	Positif
	8.	make min mangrove gamempan	Negatif
	9.	bohong nih	Negatif

Percobaan terus dilakukan hingga setiap *subset* pernah dijadikan data testing, seperti pada Tabel 3. 7.

Tabel 3. 7 Susunan Data Training dan Data Testing

Fold	Training	Testing
Fold I	S2,S3	S1
Fold II	S3,S1	S2
Fold III	S2,S1	S3

1. Percobaan pertama (*fold* I) menggunakan 6 data sebagai data training yang yang terdapat pada *subset* 2, dan 3. *Subset* 1 yang berisi 3 data sebagai data testing. Perhitungan data training terdapat pada Tabel 3. 8, Tabel 3.9.

Tabel 3. 8 Hasil Perhitungan Probabilitas *Fold* I (1)

Kategori	$P(c)$	$P(w_{kj} \mid c)$								
		hasilnya	bagus	pengen	ac	jelek	banget	sih	waw	Keren
Positif	$\frac{1}{2}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	$\frac{4}{27}$	$\frac{2}{27}$
Negatif	$\frac{1}{2}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$

Tabel 3. 9 Hasil Perhitungan Probabilitas *Fold* I (2)

Kategori	$P(c)$	$P(w_{kj} \mid c)$							
		cakep	esenang	make	min	mangrove	gamempan	bohong	Nih
Positif	$\frac{1}{2}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$
Negatif	$\frac{1}{2}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$

Pada tahap selanjutnya yaitu menghitung data testing pada *fold* I yang sebelumnya sudah diklasifikasikan.

Data testing pertama : keren (positif)

$$P(\text{tweet1}|C_{positif})$$
$$= P(w_{keren}|C_{positif}) \times P(C_{positif})$$
$$= \frac{2}{27} \times \frac{1}{2}$$
$$= 0.037037$$

$$P(\text{tweet1}|C_{negatif})$$
$$= P(w_{keren}|C_{negatif}) \times P(C_{negatif})$$
$$= \frac{1}{27} \times \frac{1}{2}$$
$$= 0.0185185$$

Data tester kedua : mahal banget harganya esedih (negatif)

$$P(\text{tweet2}|C_{positif})$$
$$= P(w_{mahal}|C_{positif}) \times P(w_{banget}|C_{positif}) \times P(w_{harganya}|C_{positif}) \times$$
$$P(w_{esedih}|C_{positif}) \times P(C_{positif})$$
$$= 1 \times \frac{2}{27} \times 1 \times 1 \times \frac{1}{2}$$
$$= 0.037037$$

$$P(\text{tweet2}|C_{negatif})$$
$$= P(w_{mahal}|C_{negatif}) \times P(w_{banget}|C_{negatif}) \times P(w_{harganya}|C_{negatif}) \times P(w_{esedih}|C_{negatif}) \times P(C_{negatif})$$
$$= 1 \times \frac{2}{27} \times 1 \times 1 \times \frac{1}{2}$$
$$= 0.037037$$

Data tester ketuga : keren banget (positif)

$$P(\text{tweet3}|C_{positif})$$
$$= P(w_{keren}|C_{positif}) \times P(w_{banget}|C_{positif}) \times P(C_{positif})$$
$$= \frac{2}{27} \times \frac{2}{27} \times \frac{1}{2}$$
$$= 0.0027435$$

$$P(\text{tweet3}|C_{negatif})$$
$$= P(w_{keren}|C_{negatif}) \times P(w_{banget}|C_{negatif}) \times P(C_{negatif})$$
$$= \frac{1}{27} \times \frac{2}{27} \times \frac{1}{2}$$
$$= 0.0013717$$

Setelah dihitung probabilitas dari masing-masing *tweet* , hasil dari perhitungan tersebut selanjutnya dilihat nilai maksimal atau nilai terbesar sebagai kategori dari tweet tersebut. Hasil perhitungan tersebut terdapat pada Tabel 3.10.

Tabel 3. 10 Hasil Perhitungan Data Testing *Fold* I

Tweet	Positif	Negatif
keren	**0.037037**	0.0185185
mahal banget harganya esedih	0.0370370	0.0370370
keren banget	**0.0027435**	0.0013717

Berdasarkan Tabel 3.10, dapat diketahui tweet pertama masuk kedalam kategori positif karena nilai probabilitasnya lebih besar dari nilai probabilitas negatif, sedangkan untuk tweet kedua masuk kedalam kategori netral disebabkan oleh nilai probabilitas positif dan negatif sama besar, dan untuk tweet ketiga

masuk kedalam kategori positif karena nilai probabilitas sentimen positif lebih besar dari nilai probabilitas negatif . Jadi, yang sesuai dengan kategori awal yang telah di klasifikasikan sebanyak 1 tweet, dengan demikian maka prosentase kebenaran secara keseluruhan adalah (2/3) ×100 atau 66.67% .

2. Percobaan kedua (*fold* 2) menggunakan 6 data sebagai data training yang terdapat pada *subset* 3, dan 1. *Subset* 2 yang berisi 3 data sebagai data testing. Perhitungan data training terdapat pada Tabel 3. 11, Tabel 3.12.

Tabel 3. 11 Hasil Perhitungan Probabilitas *Fold* II (1)

Kategori	$P(c)$	$P(w_{kj} \mid c)$						
		keren	mahal	banget	harganya	esedih	waw	cakep
Positif	$\frac{1}{2}$	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$
Negatif	$\frac{1}{2}$	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{1}{24}$

Tabel 3. 12 Hasil Perhitungan Probabilitas *Fold* II (2)

Kategori	$P(c)$	$P(w_{kj} \mid c)$						
		esenang	make	min	mangrove	gamempan	bohong	nih
Positif	$\frac{1}{2}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$
Negatif	$\frac{1}{2}$	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$

Dengan perhitungan data testing terdapat pada Tabel 3. 13.

Tabel 3. 13 Hasil Perhitungan Data Testing *Fold* II

Tweet	Positif	Negatif
hasilnya bagus pengen	0.5	0.5
ac jelek banget sih	**0.05**	0.041667
waw waw keren banget	**0.000075**	0.00000301408

Berdasarkan Tabel 3. 13 dapat diketahui kategori yang sesuai dengan kategori awal yang telah diklasifikasikan sebanyak 1 tweet. Dengan demikian maka prosentase kebenaran secara keseluruhan adalah (1/3)*100 atau 33.33%.

3. Percobaan ketiga (*fold* 3) menggunakan 6 data sebagai data training yang berisi *subset* 2, dan 1. *Subset* 3 yang berisi 3 data sebagai data testing. Perhitungan data training terdapat pada Tabel 3. 14.

Tabel 3. 14 Hasil Perhitungan Probabilitas *Fold* III

Kategori	$P(c)$	$P(w_{kj} \mid c)$											
		keren	mahal	banget	harganya	esedih	hasilnya	bagus	pengen	ac	jekel	sih	waw
Positif	$\frac{2}{3}$	$\frac{4}{22}$	$\frac{1}{22}$	$\frac{3}{22}$	$\frac{1}{22}$	$\frac{1}{22}$	$\frac{2}{22}$	$\frac{2}{22}$	$\frac{2}{22}$	$\frac{1}{22}$	$\frac{1}{22}$	$\frac{1}{22}$	$\frac{3}{22}$
Negatif	$\frac{1}{3}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

Dengan perhitungan data testing terdapat pada Tabel 3. 15.

Tabel 3. 15 Hasil Perhitungan Data Testing *Fold* III

Tweet	Positif	Negatif
waw cakep esenang	**0.09091**	0.01667
make min mangrove gamempan	**0.66667**	0.33333
bohong nih	**0.66667**	0.33333

Berdasarkan Tabel 3. 15 dapat diketahui kategori yang sesuai dengan kategori awal yang telah di klasifikasikan hanya 1 tweet. Dengan demikian maka prosentase kebenaran secara keseluruhan adalah (1/3)*100 atau 33.33%. Setelah menghitung akurasi sebanyak 3 iterasi, selanjutnya menghitung rata-rata akurasi. Perhitungan akurasi dari proses *cross validation* terdapat pada Tabel 3. 16.

Tabel 3. 16 Hasil Perhitungan Akurasi

Fold	Akurasi
I	66.67 %
II	33.33 %
III	33.33 %
Rata-rata akurasi	44.44 %

Berdasarkan Tabel 3.16 dapat dilihat rata-rata akurasi sebesar 44.44% sebagai akurasi final dari implementasi analisis sentimen dengan metode *Naïve Bayes Classifier*. Nilai tersebut menandakan akurasi yang kurang baik, ini disebabkan oleh data training yang digunakan sangat kecil, karena metode *Naïve*

Bayes Classifier secara umum semakin besar data training maka akan semakin baik kinerja model klasifikasi suatu sistem [15].

3.5 Analisis Kebutuhan Implementasi Metode

Analisis kebutuhan implementsi metode mecangkup analisis non fungsional yang akan digunakan untuk keperluan implementasi metode pada penelitian ini. Analisis non fungsional itu sendiri adalah sebuah langkah untuk menganalisis sumber daya yang akan digunakan perangkat lunak yang dibangun. Analisis non fungsional yang dilakukan dibagi dalam 3 tahap, yaitu:

1. Analisis Kebutuhan Perangkat Keras (*Hardware*)
2. Analisis Kebutuhan Perangkat Lunak (*Software*)
3. Analisis Kebutuhan Perangkat Pikir (*Brainware*)

3.5.1 Analisis Kebutuhan Perangkat Keras (*Hardware*)

Analisis kebutuhan perangkat keras (*Hardware*) pada penelitian ini merupakan kebutuhan perangkat lunak yang digunakan dalam pembangunan perangkat lunak ini. Adapun perangkat lunak yang digunakan sebagai berikut:

1. *Processor* 2.13GHz
2. *RAM* 1024MB
3. Monitor dengan resolusi 1366x768 pixels
4. Keyboard Strandard
5. Mouse Standard

3.5.2 Analisis Kebutuhan Perangkat Lunak (*Software*)

Analisis kebutuhan perangkat lunak (*Software*) pada penelitian ini merupakan kebutuhan perangkat lunak yang digunakan dalam pembangunan perangkat lunak ini. Adapun perangkat lunak yang digunakan sebagai berikut:

1. Sistem Operasi Windows 8 Strater 32 bit.
2. Bahasa pemograman PHP.
3. Web server XAMPP v3.1.0.3.1.0
4. Code editor berupa SublimeText3Portable.

5. DBMS : MySQL 5.5.27
6. Adobe Photoshop digunakan sebagai perancangan UI (*User Interface*)
7. Web Browser: Google Chrome 34.0.1847.137

3.5.3 Analisis Kebutuhan Perangkat Pikir (*Brainware*)

Kebutuhan perangkat pikir (*Brainware*) untuk implementasi metode yang dibahas dalam penelitian ini dikelompokkan menjadi lima kategori yaitu *user knowledge and experience, user jobs and tasks, user physical characteristic* dan *user tools*. Penjelasan untuk masing-masing kategori sebagai berikut:

1. Pengetahuan dan Pengalaman Pengguna (*User Knowledge and Experience)*

Pengetahuan dan pengalaman pengguna yang ditargetkan dalam pembangunan perangkat lunakini dapat dilihat padaTabel 3. 17.

Tabel 3. 17 Pengetahuan dan Pengalaman Pengguna

Aspek	Pengguna
Education Level	SMA sampai Perguruan tinggi (D3,S1,S2,S3)
Reading Level	Sedang Sampai Tinggi
Computer Literacy	*moderate experience till highly technical*
Task Experience	Rendah sampai Tinggi
System Experience	Rendah sampai Tinggi
Application Experience	Rendah sampai Tinggi
Use Of Other System	No
Native Language or Culture	Dominan Indonesia

2. Tugas dan Kebutuhan Pengguna (*User Jobs and Tasks)*

Tugas dan pekerjaan pengguna yang ditargetkan dalam pembangunan perangkat lunak ini dapat dilihat pada Tabel 3. 18.

Tabel 3. 18 Tugas dan Kebutuhan Pengguna

Aspek	Pengguna
Frequency Of Use	*Continual*
Primary Training	Disediakan menu petunjuk
Task or Need Importance	Sedang sampai Tinggi
Task Structure	Sedang sampai Tinggi
Job Categories	No
Life Style	No

3. Karakteristik Pengguna(*User Physical Characteristic*)

Karakter fisik pengguna yang ditargetkan dalam pembangunan perangkat lunakini dapat dilihat pada Tabel 3. 19.

Tabel 3. 19 Karakteristik Pengguna

Aspek	Pengguna
Handedness	Keduanya dalam kondisi baik
Gender	Laki-laki dan perempuan
Age	20 tahun ke atas
Disabilities	Tidak cacat tubuh

4. *User Tool*

Ketika pengguna menggunakan perangkat lunak ini peralatan yang dominan digunakan adalah mouse.

3.6 Analisis Kebutuhan Fungsional

Dalam langkah ini dilakukan penentuan entitas-entitas yang diperlukan untuk implementasi metode dalam penelitian ini, baik entitas internal maupun entitas eksternal, data yang mengalir, serta prosedur-prosedur yang bisa dilakukan masing-masing entitas seperti diagram konteks, diagram alir data, spesifikasi proses dan kamus data. Kebutuhan-kebutuhan fungsional yang dibutuhkan untuk mengimplementasikan metode dalam penelitian ini terlampir pada lampiran A.

3.7 Perancangan Sistem

Perancangan sistem merupakan bagian dari metodologi pembangunan suatu perangkat lunak yang dilakukan setelah melalui tahapan analisis. Pada bagian ini akan dijelaskan perancangan sistem yang diajukan untuk mengimplementasikan metode pada analisis sentimen. Perancangan sistem yang dibuat terdiri dari perancangan struktur tabel, perancangan menu, perancangan antarmuka, dan jaringan semantik. Perancangan-perancangan tersebut terlampir pada lampiran B.