

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini. Analisa sentimen dilakukan untuk penilaian terhadap pendapat atau komentar orang terhadap beberapa topik sehingga didapatkan bahwa pendapat tersebut termasuk pendapat yang positif atau negatif. Penyampaian pendapat masyarakat pada masa kini semakin mudah dengan adanya berbagai sosial media, salah satunya yakni twitter. Twitter merupakan jejaring sosial yang memungkinkan penggunanya untuk mengirimkan pesan teks sejumlah 280 karakter.

Jenis *tweet* yang dituliskan oleh masyarakat terdapat berbagai topik atau tema yang di bahas, salah satu contohnya seperti pendapat terhadap operator seluler. Pendapat dari masyarakat mengenai operator seluler menjadi perlu untuk diketahui kelas topik dan sentimen yang terdapat pada *tweet* tersebut. Topik yang dibahas merupakan tema yang dihubungkan dengan operator seluler tersebut, sedangkan untuk sentimen merupakan nilai dari *tweet* yang membahas operator seluler yang dimaksud sehingga dapat diketahui nilai positif atau negatif yang terkandung dalam *tweet* tersebut.

3.2 Hasil Analisis

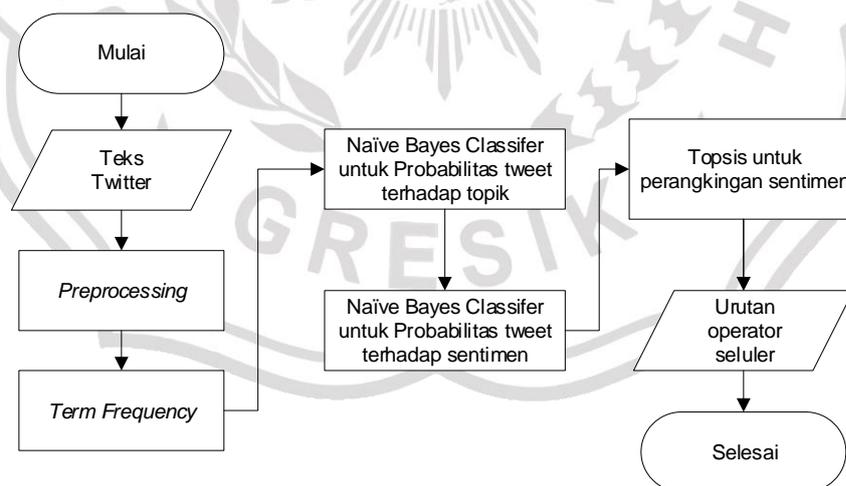
Hasil analisis yang dapat dilakukan dari pengklasifikasian topik dan sentimen *tweet* yang membahas operator seluler dapat melakukan klasifikasi jenis topik dan sentimen yang terkandung dalam *tweet* di sosial media twitter kemudian melakukan perbandingan operator seluler dari hasil pengklasifikasian topik dan sentimen. Pembuatan sistem klasifikasi topik dan sentimen operator seluler ini menggunakan metode *Naïve Bayes Classifier*. Penggunaan metode *Naïve Bayes Classifier* memerlukan data *tweet public* yang terlebih dahulu sudah memiliki label topik dan sentimen untuk dijadikan sebagai data training atau data acuan dalam menentukan kelas topik dan sentimen untuk *tweet* yang baru. *Tweet* yang akan dilakukan

klasifikasi akan diberi salah satu topik dari tiga topik yang dijadikan sebagai kelas topik yakni Sinyal, Internet, dan Paket. Penggunaan ketiga topik tersebut karena pentingnya hal tersebut dan juga banyak *tweet* yang membahas topik tersebut terlebih dahulu. Setelah diketahui topik dari *tweet* tersebut, maka *tweet* itu selanjutnya dilakukan klasifikasi kembali dengan *Naïve Bayes Classifier* untuk mendapatkan kelas sentimen positif atau negatif. Sehingga masing-masing *tweet* akan memiliki kelas topik dan sentimen kemudian dilanjutkan dengan menggunakan metode *TOPSIS* untuk perangkingan operator seluler dengan data hasil dari pengklasifikasian sentimen dan topik.

3.3 Perancang Sistem

3.3.1 Flowchart sistem

Sistem yang akan dibangun adalah aplikasi dengan konsep sistem temu kembali informasi yang dapat menganalisa sentimen terhadap operator seluler yang berada di media sosial *twitter* menggunakan metode *naïve bayes classifier* untuk proses klasifikasi *tweet*. Tujuan dari sistem adalah untuk memberikan informasi mengenai operator seluler kepada masyarakat yang akan di kelompokkan ke sentimen positif dan negatif dan topik sinyal, internet dan paket. Gambaran umum sistem yang akan dibangun seperti pada gambar 3.1.



Gambar 3. 1 Gambaran umum sistem yang akan dibangun

3.3.1.1 Teks Twitter

Pada penelitian ini tidak menggunakan teks yang berasal dari dokumen tetapi data yang akan diolah berasal dari *tweet* yang dituliskan oleh masyarakat di

sosial media *twitter*. Informasi *tweet* yang digunakan diambil menggunakan ekstensi *Web Scraper* pada *google chrome* kemudian masuk ke *twitter* dengan *keyword* pencariannya yakni nama operator seluler. Informasi *tweet* akan disimpan terlebih dahulu di *database* dan dikelompokkan berdasarkan *keyword* operator seluler yang digunakan. Data teks yang didapatkan dari sosial media *twitter* yang digunakan untuk data latih seperti pada tabel 3.1.

Tabel 3. 1 Contoh data latih teks *twitter*

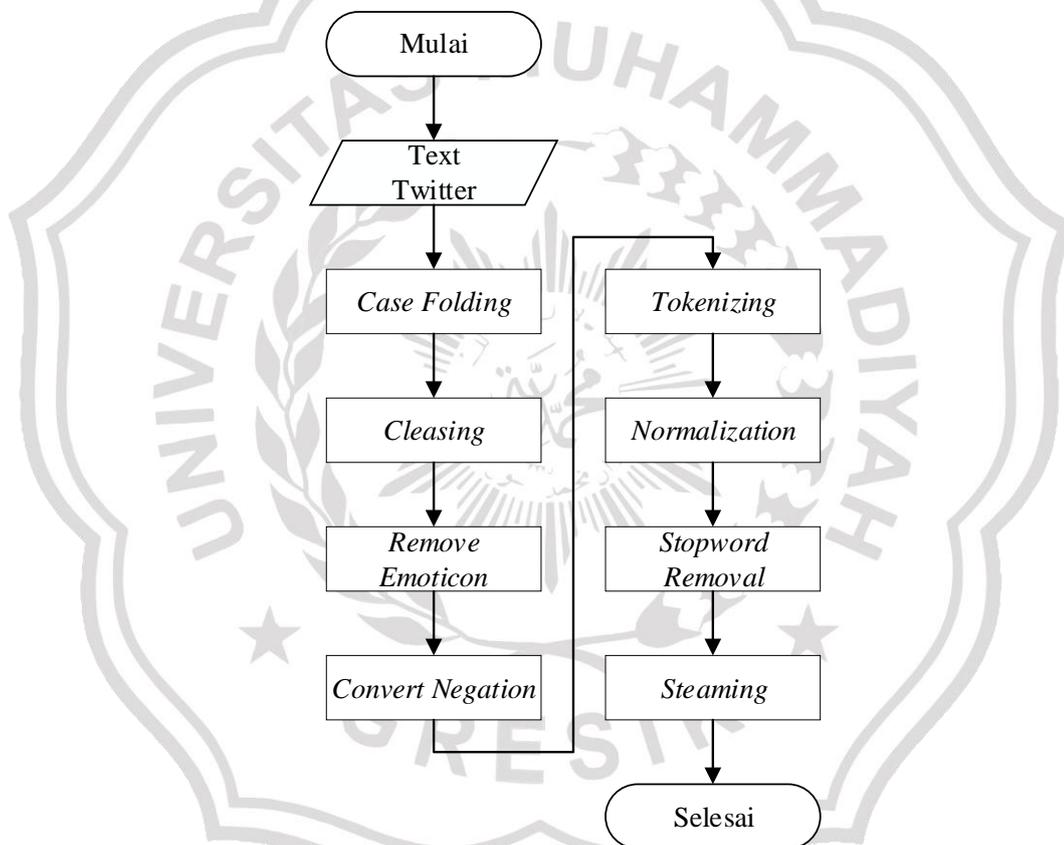
T	Operator Seluler	<i>Tweet</i>	Topik	Sentimen
T1	Telkomsel	@telkomsel tidak salah pilih, sinyal lancar di daerahku :)	Sinyal	Positif
T2	Telkomsel	Alhamdulillah pake @Telkomsel internet lancar jaya, walaupun paket na dua satengah giga	Internet	Positif
T3	Telkomsel	internet lancar mantep tanpa buffering nonton youtube @telkomsel	Internet	Positif
T4	Telkomsel	Ini @Telkomsel kenapa lagi? Dari tadi pagi kok gak bisa akses internet. Lemot banget	Internet	Negatif
T5	Telkomsel	harga kuota internet tambah mahal @telkomsel mengutamakan keutungan mulu . nyebelin telkomsel : (Paket	Negatif
T6	Telkomsel	@Fii_08 Semoga dia @Telkomsel bisa mempertahankan paket data murah <3	Paket	Positif
T7	Telkomsel	@Telkomsel kenapa jaringan internet saya lemot sekali min.. : (?	?

Dari tabel 3.1 diatas bahwa terdapat sepuluh teks *twitter* yang telah dilakukan *crawler*. Ketujuh teks tersebut terdiri dari tiga teks *twitter* yang digunakan sebagai data latih yakni T1, T4, T6 dan terdapat satu teks yang masih belum diketahui topik dan sentimennya yakni T7. Data latih didapatkan dengan membaca teks lengkap *twitter*. Cara menentukan topik dan sentimen yang digunakan sebagai data latih dengan memcocokkan dengan kumpulan kata positif, negatif, jika pada kumpulan kata-kata *tweet* tersebut diambil kategori sentimen yang dominan. Pada *tweet*

pertama terdapat kata tidaksalah dan lancar yang termasuk kedalam sentimen positif, Dikarenakan jumlah kategori terbanyak adalah positif maka teks twitter pertama termasuk kedalam kategori sentimen positif.

3.3.1.2 Preprocessing

Secara umum sistem ini memiliki beberapa tahapan proses, yang pertama adalah *preprocessing*. Tahap *preprocessing* mencakup berbagai proses seperti *case folding*, *cleasing*, *remove emoticon*, *convert negation*, *tokenizing*, *normalization*, *stop removal*, *stemming* dan *weighting*. Alur proses dalam tahap *preprocessing* seperti pada gambar 3.2



Gambar 3. 2 Flowchart tahap *preprocessing*

Pada tahap preprocessing ada beberapa proses yang akan dilakukan. Pertama yakni memasukan data inputan dimana pada tahap ini akan dimasukan berupa teks yang berasal dari sosial media *twitter*. Setelah teks *twitter* diinputkan maka teks tersebut akan dilakukan proses *case folding* yakni teks inputan dilakukan perubahan huruf ke huruf kecil (*lowercase*) untuk semua teks. Proses *cleasing* yakni

proses menghilangkan atribut. Kemudian proses *convert emoticon* yakni proses menghilangkan emoticon yang ada pada *tweet*. Proses *convert negation* adalah proses penggabungan kata negasi dengan kata selanjutnya. Proses *tokenizing* merupakan proses setelah *convert negation*, proses *tokenizing* merupakan proses memecah teks yang awalnya satu kalimat dirubah menjadi perkata. Proses *normalization* merupakan proses perubahan kata yang tidak sesuai dengan EYD. Proses *stopword removal* yakni melakukan pembuangan kata-kata yang tidak memiliki makna dan menghapus kata yang muncul berkali-kali sehingga kata yang muncul berkali-kali tidak mempengaruhi bobot kata tersebut. Setelah proses *stopword removal* maka teks inputan akan dilakukan proses selanjutnya yakni *stemming* dimana dilakukan pembuangan imbuhan pada kata sehingga menjadi kata dasar.

3.3.1.3 Case Folding

Tahap ini merupakan tahap dimana teks inputan yang berasal dari sosial media *twitter* pertama diproses. *case folding* merupakan proses penyamaan huruf teks. Pada data *tweet* terdapat kata yang berhuruf besar dan terdapat teks yang berhuruf kecil. Perbedaan bentuk huruf pada data *twitter* tersebut akan mengganggu proses selanjutnya maka dari itu diperlukan proses *case folding* untuk merubah teks *twitter* menjadi huruf kecil semua. Teks *twitter* pada tabel 3.1 yang akan diproses selanjutnya yakni T1, T4, T6, T7 yang masih tercampur antara huruf kapital dan huruf kecil akan dilakukan *case folding* sehingga teks *twitter* menjadi seperti pada tabel 3.2 berikut.

Tabel 3. 2 Teks *twitter* hasil *case folding*

<i>Tweet</i>
@telkomsel tidak salah pilih, sinyal lancar di daerahku :)
ini @Telkomsel kenapa lagi? dari td pagi kok gak bisa akses internet. lemot banget
@Fii_08 smg dia @Telkomsel bisa mempertahankan paket data murah <3
@Telkomsel knp jaringan internet saya lemot sekali min.. : (

3.3.1.4 *Cleasng*

Tahapan *cleasng* merupakan tahap pembersihan kata yang tidak berpengaruh sama sekali terhadap hasil klasifikasi sentimen. Komponen dokumen *tweet* memiliki berbagai atribut yang tidak berpengaruh terhadap sentimen, karena setiap *tweet* hampir semua memiliki atribut tersebut. Contoh dari atribut yang tidak penting tersebut adalah yaitu *mention* yang diawali dengan atribut ('@'), *hashtag* yang diawali dengan atribut ('#'), *link* yang diawali dengan atribut ('http', 'bit.ly') dan karakter simbol (~!@#\$%^&*()_+?<>.,?:{}[]|). Atribut yang tidak berpengaruh tersebut akan dihilangkan dari dokumen kemudian akan digantikan dengan karakter spasi. seperti pada tabel 3.3 berikut.

Tabel 3. 3 Teks *twitter* hasil *cleasng*

<i>Tweet</i>
tidak salah pilih, sinyal lancar di daerahku :)
ini knp lagi? dari tadi pagi kok gak bisa akses internet. lemot banget
smg dia bisa mempertahankan paket data murah <3
kenapa jaringan internet saya lemot sekali min.. : (

3.3.1.5 *Remove Emoticon*

Ketika sedang menulis status (*tweet*) seseorang kadang salah atau kurang tepat dalam penggunaan emoticon, entah disengaja atau tidak banyak yang melakukannya. Contohnya: Mereka hanya bisa memfitnah karena tidak bisa ketemu fakta buruk :), kata opini fitnah tapi emoticonnya senyum :), dengan begitu emoticon akan mengganggu dalam proses *Sentiment Analysis tweet*, jadi dalam proses ini emoticon dihapus atau diabaikan. seperti pada tabel 3.4 berikut.

Tabel 3. 4 Teks *twitter* hasil *remove emoticon*

<i>Tweet</i>
tidak salah pilih, sinyal lancar di daerahku
ini knp lagi? dari tadi pagi kok gak bisa akses internet. lemot banget
smg dia bisa mempertahankan paket data murah
kenapa jaringan internet saya lemot sekali min..

3.3.1.6 Convert Negation

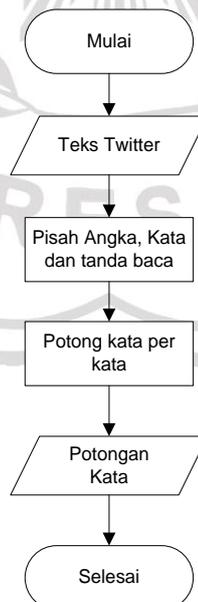
Tahap *convert negation* merupakan proses konversi kata – kata negasi yang terdapat pada suatu *tweet*. Kata negasi akan merubah makna sentimen suatu dokumen, sehingga kata negasi akan digabungkan dengan kata selanjutnya. Contoh kata negasi adalah ‘bukan’, ‘tidak’, ‘jangan’ dan lain sebagainya. seperti pada tabel 3.5 berikut.

Tabel 3. 5 Teks *twitter* hasil *convert negation*

<i>Tweet</i>
tidaksalah pilih, sinyal lancar di daerahku
ini knp lagi? dari tadi pagi kok gakbisa akses internet. lemot banget
smg dia bisa mempertahankan paket data murah
kenapa jaringan internet saya lemot sekali min..

3.3.1.7 Tokenizing

Tahap *tokenizing* merupakan tahap yang berfungsi untuk memisahkan antar kata yang ada didalam teks *twitter*. Proses ini menggunakan bantuan karakter spasi dalam pemisahan kata-nya. Pada sebuah teks *twitter* terdiri dari beberapa kata yang saling terhubung dan dipisahkan dengan spasi. Untuk mempermudah dalam pemrosesan sebuah teks maka masing-masing kata dalam sebuah kalimat tersebut harus dipisahkan. Alur proses *tokenizing* seperti pada gambar 3.3



Gambar 3. 3 Alur proses dalam *tokenizing*

Pada proses tokenizing tidak hanya kata yang akan dipisahkan, angka dan tanda baca juga dipisahkan. Pemisahan angka dan tanda baca harus dilakukan karena akan mampu mempengaruhi dalam penilaian terhadap sebuah kalimat atau kata tersebut. Teks *twitter* pada tabel 3.5 yang sudah dilakukan *convert negation* akan dilakukan *tokenizing* sehingga menjadi potongan kata. seperti pada tabel 3.6 seperti berikut.

Tabel 3. 6 Term hasil proses *tokenizing*

<i>Tweet</i>
tidaksalah pilih , sinyal lancar di daerahku
ini knp lagi ? dari tadi pagi kok gakbisa akses internet . lemot banget
smg dia bisa mempertahankan paket data murah
kenapa jaringan internet saya lemot sekali min . .

3.3.1.8 Normalization

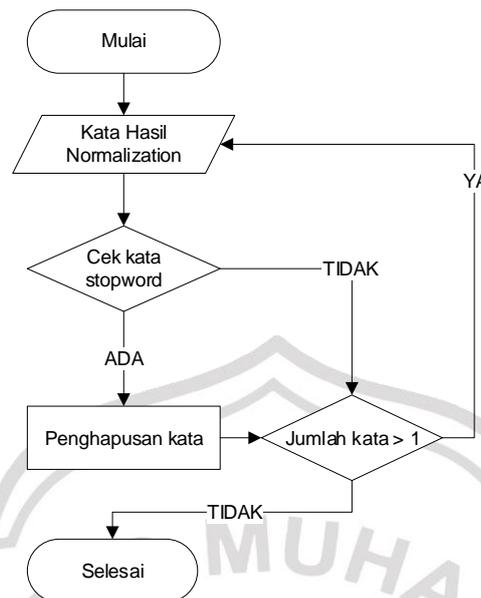
Pada tahap *normalization* ini dilakukan perubahan kata yang tidak sesuai dengan EYD, sehingga dapat mengurangi hasil sentimen dokumen. Tahap ini melakukan konversi kata singkatan. seperti pada tabel 3.7 seperti berikut.

Tabel 3. 7 Term hasil proses *normalization*

<i>Tweet</i>
tidaksalah pilih , sinyal lancar di daerahku
ini kenapa lagi ? dari tadi pagi kok tidakbisa akses internet . lemot banget
semoga dia bisa mempertahankan paket data murah
kenapa jaringan internet saya lemot sekali admin . .

3.3.1.9 Stopword Removal

Kalimat yang sudah melewati proses pemisahan kata melalui *tokenizing* selanjutnya akan dilakukan proses *stopword removal*. Proses dimana akan menghapus kata-kata yang tidak penting atau tidak memiliki makna seperti kata sambung. Pada proses *stopword removal* dibutuhkan sebuah *database* yang menyimpan daftar *stopword* yang digunakan sebagai acuan dalam penghapusan kata. Proses dalam *stopword removal* seperti pada gambar 3.4



Gambar 3. 4 Alur proses *stopword removal*

Kata hasil proses *normalization* digunakan sebagai inputan pada proses *stopword removal*. Setelah didapatkan kata tersebut maka tiap kata akan dilakukan pengecekan kata dengan data pembandingan yakni kata yg ada dalam stopword. Jika kata tersebut termasuk dalam kumpulan kata stopword maka kata tersebut akan dihapus. Setelah proses pengecekan kata stopword dilakukan maka dilakukan pengecekan kembali untuk jumlah kata. Jika dalam pengecekan ditemukan jumlah kata lebih dari satu maka kata tersebut akan dilakukan pengecekan stopword kembali. Hasil proses *stopword removal* yang didapatkan dari *term* pada tabel 3.7 seperti pada tabel 3.8 berikut.

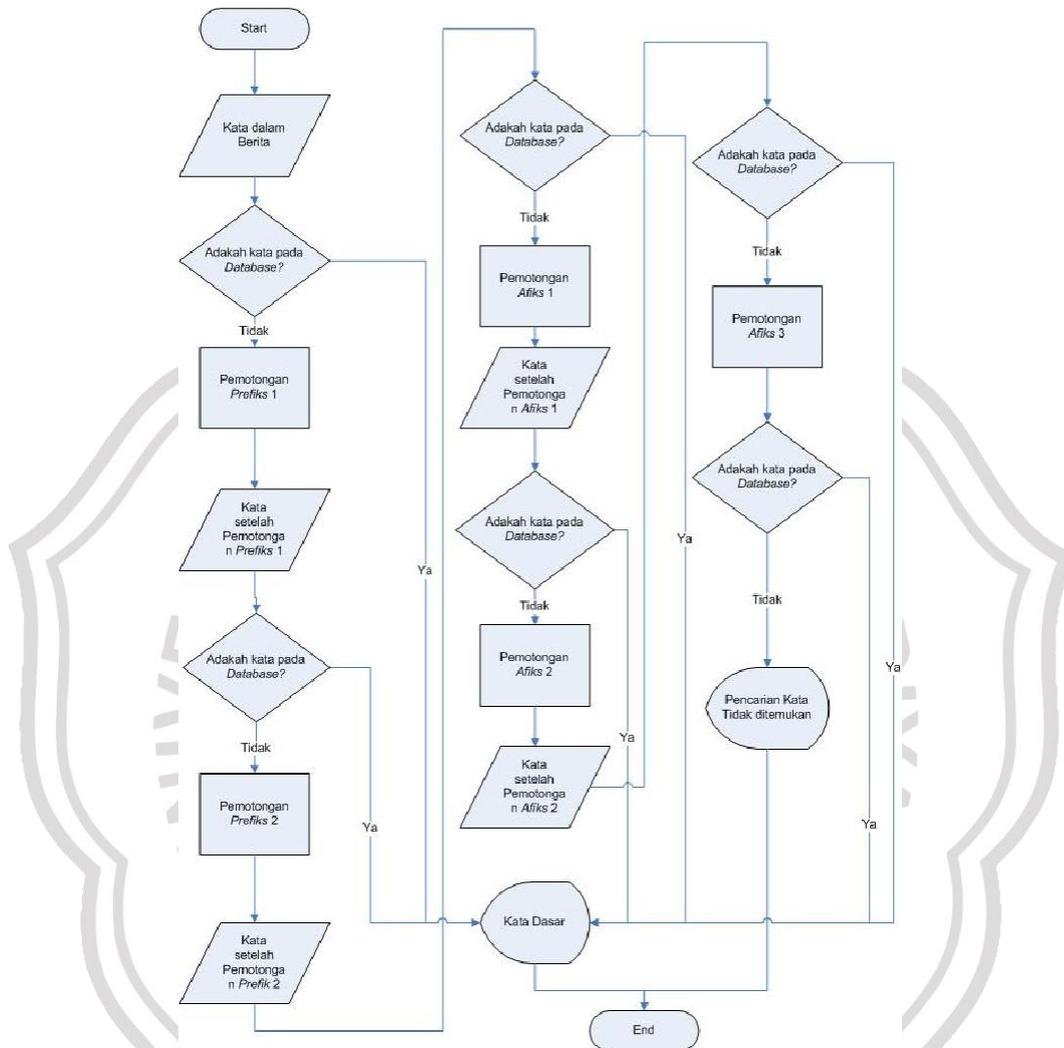
Tabel 3. 8 Term hasil proses *stopword removal*

<i>Tweet</i>
tidaksalah pilih sinyal lancar daerahku
pagi tidakbisa akses internet lemot banget
bisa mempertahankan paket data murah
jaringan internet lemot admin

3.3.1.10 *Stemming*

Proses setelah *stopword removal* adalah proses *stemming*. Proses *stemming* merupakan proses pencarian kata dasar terhadap sebuah kata yang telah dilakukan proses *stopword removal*. Pada proses ini kata akan dirubah menjadi kata dasar

sehingga kata yang memiliki imbuhan depan atau imbuhan belakang akan dihapus dan disesuaikan kata dasarnya. Proses *stemming* menggunakan metode Sastrawi. Alur proses dalam *stemming* seperti pada gambar 3.5



Gambar 3.5 Alur proses *stemming*

Proses *stemming* yang akan mengubah bentuk kata yang awalnya bukan kata dasar diubah menjadi kata dasar menggunakan metode Sastrawi. Hasil proses *stemming* yang dilakukan dengan menggunakan tabel 3.8 menghasilkan seperti pada tabel 3.9

Tabel 3.9 Term hasil proses *stemming* (korpus)

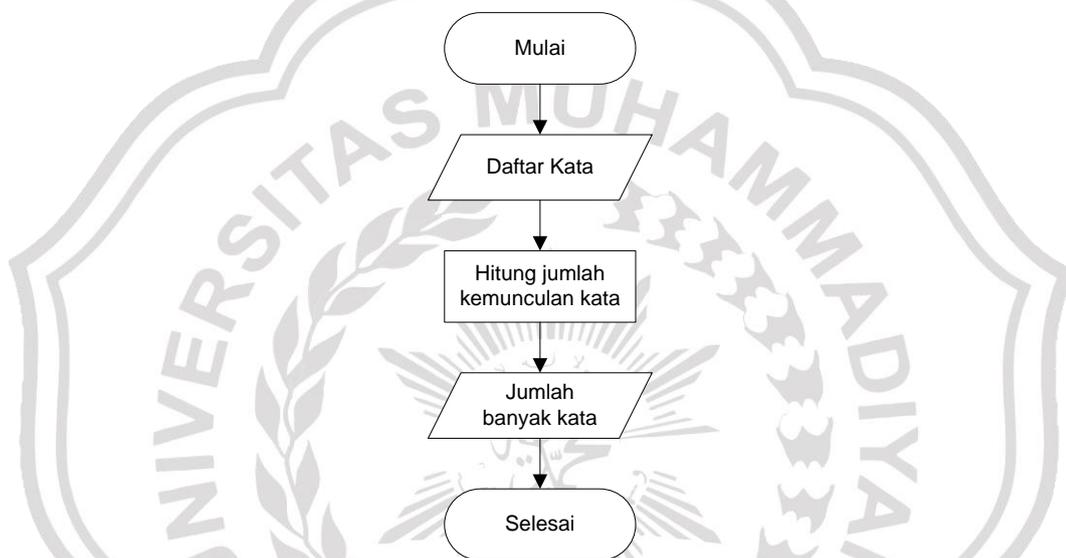
<i>Tweet</i>
tidaksalah pilih sinyal lancar daerah
pagi tidakbisa akses internet lemot banget

bisa tahan paket data murah
jaringan internet lemot admin

3.3.1.11 Perhitungan Analisis Sentimen (Representasi Data)

3.3.1.11.1 Fitur Kemunculan (*Term Frequency*)

Pembobotan kata dalam penelitian ini menggunakan perhitungan *term frequency* (TF). Perhitungan dalam pembobotan ini akan digunakan sebagai data pada penerapan metode *naïve bayes* untuk menentukan topik *twitter* dan penentuan hasil sentimen. Alur proses pembotoan kata seperti pada gambar 3.6



Gambar 3. 6 Alur proses perhitungan fitur kemunculan kata

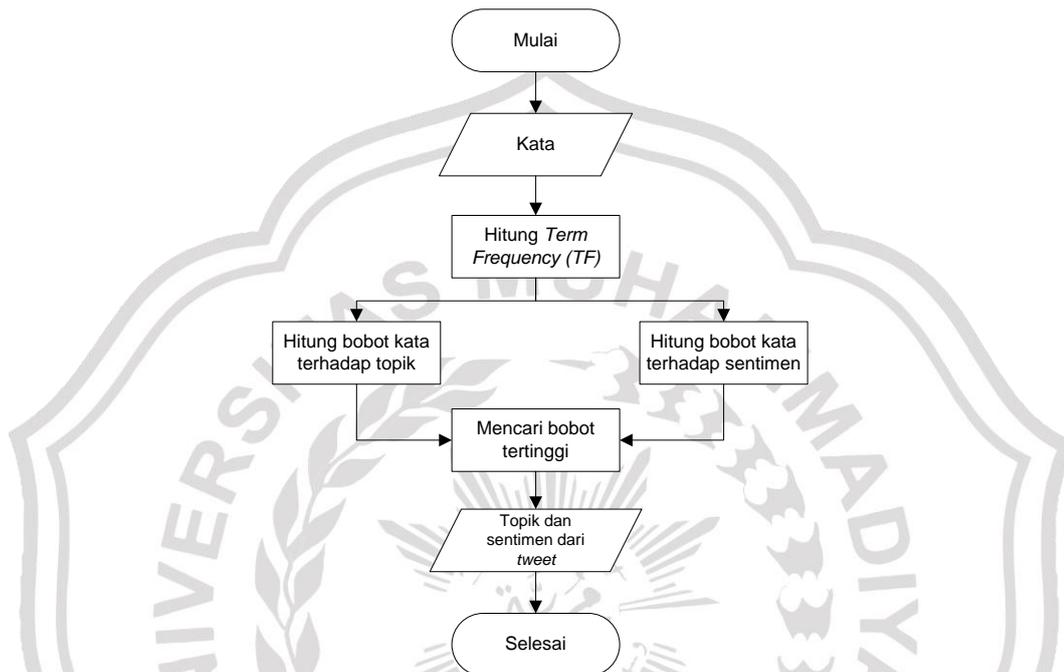
Proses pembobotan kata dengan perhitungan *term frequency* diawali dengan mengambil kata di *database* yang telah dilakukan *preprocessing*. Setelah didapatkan kata maka akan dihitung jumlah kemunculan atau frekuensi kata pada tiap *tweet*. Perhitungan *term* atau kata menggunakan kata hasil *preprocessing* dan menghasilkan bobot masing-masing kata seperti pada tabel 3.10 berikut.

Tabel 3. 10 term dan bobot hasil *term frequency*

Topik	Sentimen	<i>Tweet</i>
Sinyal	Positif	tidaksalah (1) pilih (1) sinyal (1) lancar (1) daerah (1)
Internet	Negatif	pagi (1) tidakbisa (1) akses (1) internet (1) lemot (1) banget (1)
Paket	Positif	bisa (1) tahan (1) paket (1) data (1) murah (1)
?	?	jaringan (1) internet (1) lemot (1) admin (1)

3.3.1.11.2 Perhitungan *Naïve Bayes Classifier*

Penelitian ini menggunakan metode *naïve bayes classifier* dalam proses pembobotan teks. Teks yang sudah melalui proses *preprocessing* akan dilakukan perhitungan untuk menentukan topik ataupun sentimen. Alur proses dalam perhitungan dengan *naïve bayes classifier* seperti padaa gambar 3.7



Gambar 3. 7 Alur proses perhitungan *naïve bayes classifier*

Perhitungan *naïve bayes classifier* digunakan untuk menentukan nilai teks tersebut dalam topik dan sentimen. Kata hasil *preprocessing* yang telah dihitung fitur kemunculannya seperti pada tabel 3.6 digunakan sebagai *input* awal. Setelah didapatkan fitur kemunculan kata maka akan ditentukan nilai bobot kata terhadap topik dan bobot kata terhadap sentimen. Nilai bobot pada topik dan sentimen yang diambil adalah nilai tertinggi dari beberapa kemungkinan yang telah dihitung.

3.3.1.11.3 Probabilitas Topik *Tweet*

Probabilitas topik *tweet* merupakan proses penentuan teks *twitter* yang dihitung untuk didapatkan nilai tertinggi dalam beberapa topik. Contoh kasus dalam penelitian ini menggunakan 4 teks *twitter* awal seperti pada tabel 3.1 yang telah dilakukan *preprocessing* dan masing-masing kata sudah memiliki fitur kemunculan seperti pada tabel 3.11.

Sinyal : tidaksalah (1), pilih (1), sinyal (1), lancar (1), daerah (1),

Internet : pagi (1), tidakbisa (1), akses (1), internet (1), lemot (1), banget (1)

Paket : bisa (1), tahan (1), paket (1), data (1), murah (1)

? : jaringan (1), internet (1), lemot (1), admin (1)

Tabel 3. 11 Fitur kemunculan kata

Teks	Tidaksalah	Pilih	Sinyal	Lancar	Daerah	Pagi
T1	1	1	1	1	1	0
T2	0	0	0	0	0	1
T3	0	0	0	0	0	0
T4	0	0	0	0	0	0

Teks	Tidakbisa	Akses	Internet	Lemot	Banget	Bisa	Tahan	Paket
T1	0	0	0	0	0	0	0	0
T2	1	1	1	1	1	0	0	0
T3	0	0	0	0	0	1	1	1
T4	0	0	1	1	0	0	0	0

Teks	Data	Murah	Jaringan	Admin
T1	1	1	0	0
T2	0	0	0	0
T3	0	0	0	0
T4	0	0	1	1

Setelah didapatkan jumlah kemunculan pada masing-masing kata maka selanjutnya menentukan nilai probabilitas awal dari topik dan masing-masing kata. Nantinya akan didapatkan bobot probabilitas pada topik $p(v_j)$ dan probabilitas pada tiap kata $p(w_k.v_j)$. Menentukan nilai probabilitas topik menggunakan rumus 2.7 dimana nilai $p(v_j)$ didapatkan dari jumlah *tweet* yang bertopik ci dibagi dengan keseluruhan *tweet* yang dijadikan sebagai data latih.

$$p(\text{sinyal}) = 1/3$$

$$p(\text{internet}) = 1/3$$

$$p(\text{paket}) = 1/3$$

Topik	$p(v_j)$	$p(w_k.v_j)$					
		Tahan	Paket	Data	Murah	Jaringan	Admin
Sinyal	1/3	1/24	1/24	1/24	1/24	1/24	1/24
Internet	1/3	1/24	1/24	1/24	1/24	1/24	1/24
Paket	1/3	2/24	2/24	2/24	2/24	1/24	1/24

$$\begin{aligned}
 p(\text{"sinyal"}|\text{"T4"}) &= p(\text{"sinyal"}) \times p(\text{"jaringan"}|\text{"sinyal"}) \times p(\text{"internet"}|\text{"sinyal"}) \times p(\text{"lemot"}|\text{"sinyal"}) \times p(\text{"admin"}|\text{"sinyal"}) \\
 &= 1/3 \times (1/24 \times 1/24 \times 1/24 \times 1/24) = 0.000001191
 \end{aligned}$$

$$\begin{aligned}
 p(\text{"internet"}|\text{"T4"}) &= p(\text{"internet"}) \times p(\text{"jaringan"}|\text{"internet"}) \times p(\text{"internet"}|\text{"internet"}) \times p(\text{"lemot"}|\text{"internet"}) \times p(\text{"admin"}|\text{"internet"}) \\
 &= 1/3 \times (1/24 \times 2/24 \times 2/24 \times 1/24) = 0.000004019
 \end{aligned}$$

$$\begin{aligned}
 p(\text{"paket"}|\text{"T4"}) &= p(\text{"paket"}) \times p(\text{"jaringan"}|\text{"paket"}) \times p(\text{"internet"}|\text{"paket"}) \times p(\text{"lemot"}|\text{"paket"}) \times p(\text{"admin"}|\text{"paket"}) \\
 &= 1/3 \times (1/24 \times 1/24 \times 1/24 \times 1/24) = 0.000001191
 \end{aligned}$$

Dengan melakukan perhitungan probabilitas teks terhadap topik dengan menggunakan metode *naïve bayes* maka dapat diketahui bahwa teks *twitter* T4 termasuk dalam kategori topik Internet karena nilai bobot probabilitas teks *twitter* T4 terhadap topik Internet lebih besar dari pada nilai bobot teks T4 terhadap topik lainnya.

3.3.1.11.4 Probabilitas Sentimen

Probabilitas sentimen merupakan proses penentuan teks *twitter* yang dihitung untuk didapatkan nilai tertinggi dari beberapa kategori dalam sentimen. Keempat teks yang digunakan untuk perhitungan probabilitas sentimen seperti pada tabel 3.1 yang terlebih dahulu sudah dilakukan *preprocessing* sampai dengan penghitungan fitur kemunculan (*term frequency*) seperti pada tabel 3.7

Positif : tidaksalah(1), pilih(1), sinyal(1), lancar(1), daerah(1), bisa(1), tahan(1), paket(1), data(1), murah(1)

Negatif : pagi(1), tidakbisa(1), akses(1), internet(1), lemot(1), banget(1)

Sentimen	$p(vk)$	$p(wk.vj)$					
		Tahan	Paket	Data	Murah	Jaringan	Admin
Positif	2/3	2/28	2/28	2/28	2/28	1/28	1/28
Negatif	1/3	1/24	1/24	1/24	1/24	1/24	1/24

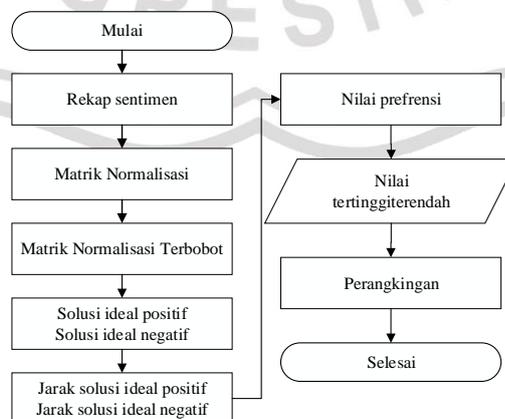
$$\begin{aligned}
 p(\text{"positif"}|\text{"T4"}) &= p(\text{"positif"}) \times p(\text{"jaringan"}|\text{"positif"}) \times \\
 & p(\text{"internet"}|\text{"positif"}) \times p(\text{"lemot"}|\text{"positif"}) \times p(\text{"admin"}|\text{"positif"}) \\
 &= 2/3 \times (1/28 \times 1/28 \times 1/28 \times 1/28) = 0.000001085
 \end{aligned}$$

$$\begin{aligned}
 p(\text{"negatif"}|\text{"T4"}) &= p(\text{"negatif"}) \times p(\text{"jaringan"}|\text{"negatif"}) \times \\
 & p(\text{"internet"}|\text{"negatif"}) \times p(\text{"lemot"}|\text{"negatif"}) \times p(\text{"admin"}|\text{"negatif"}) \\
 &= 1/3 \times (1/24 \times 2/24 \times 2/24 \times 1/24) = 0.000004019
 \end{aligned}$$

Dengan melakukan perhitungan probabilistik teks terhadap sentimen dengan menggunakan metode *naïve bayes classifier* maka dapat diketahui bahwa teks *twitter* T4 termasuk dalam kategori sentimen negatif karena nilai bobot probabilitas teks *twitter* T4 terhadap topik negatif lebih besar dari pada nilai bobot teks T4 terhadap topik lainnya.

3.3.1.11.5 Perhitungan TOPSIS

Pada penelitian ini penggunaan metode *TOPSIS* digunakan untuk menghitung nilai sentimen yang ada pada topik setiap operator seluler sehingga didapatkan urutan-urutan nama operator seluler yang mendapatkan nilai sentimen tertinggi sampai dengan sentimen terendah. Alur proses perhitungan dengan metode *TOPSIS* seperti pada gambar 3.8



Gambar 3. 8 alur proses perhitungan metode *TOPSIS*

Awal dari proses perhitungan metode *TOPSIS* adalah menentukan nilai bobot pada masing-masing topik. Penentuan nilai bobot topik dilakukan untuk mendapatkan bobot prioritas pada topik. Bobot prioritas dihitung dengan rumus (2.12)

Contoh kasus apabila diketahui jumlah *tweet* yang tergabung dalam topik sinyal terdapat 25 *tweet*, topik internet terdapat 30 *tweet*, topik paket terdapat 20 *tweet* dan jumlah keseluruhan *tweet* yakni 75.

$$\text{Bobot prioritas topik (Sinyal)} = 25 / 75 = 0.33$$

$$\text{Bobot prioritas topik (Internet)} = 30 / 75 = 0.40$$

$$\text{Bobot prioritas topik (Paket)} = 20 / 75 = 0.27$$

Detail hasil perhitungan bobot prioritas untuk perhitungan diatas seperti pada tabel 3.14

Tabel 3. 14 Bobot prioritas tiap topik

Topik	Jumlah	Bobot
Sinyal	25	0.33
Internet	30	0.40
Paket	20	0.27
JUMLAH	75	1.00

Nilai sentimen merupakan nilai yang didapatkan dari rekap jumlah sentimen dari operator seluler. Sentimen yang didapatkan dikelompokan berdasarkan topik dan sentimen-nya (positif dan negatif). Dimana dari tabel 3.1 terdapat tiga (3) jumlah *tweet* yang teridentifikasi sebagai topik internet untuk operator seluler Telkomsel, dari ketiga *tweet* tersebut dua (2) diantaranya termasuk sentimen positif dan terdapat satu (1) yang termasuk ke sentimen negatif. Dari jumlah *tweet* telkomsel untuk topik internet dan memiliki sentimen maka akan di hitung untuk menentukan operator seluler dengan sentimen terbaik. Dari tabel 3.1 terbentuk data pada tabel 3.15 untuk perhitungan metode *TOPSIS*

Tabel 3. 15 Rekap jumlah sentimen operator seluler

Nama	Sentimen	Kriteria		
		Sinyal	Internet	Paket
Telkomsel	Positif	5	2	4
	Negatif	0	1	3

Indosat	Positif	4	5	4
	Negatif	0	1	1
XL	Positif	4	4	2
	Negatif	1	1	3

Setelah didapatkan rekap sentimen pada masing-masing operator seluler, maka tahap selanjutnya adalah menghitung nilai sentimen tiap topik pada masing-masing operator seluler dengan melakukan kalkulasi pada tiap nilai sentimen. Kalkulasi dilakukan dengan cara mengurangi jumlah sentimen positif dengan nilai negatif.

$$Telkomsel(sinyal) = 5 - 0 = 5$$

$$Telkomsel(internet) = 2 - 1 = 1$$

$$Telkomsel(paket) = 4 - 3 = 1$$

Tabel 3. 16 rekap nilai sentimen operator seluler

Nama	Kriteria		
	Sinyal	Internet	Paket
Telkomsel	5	1	1
Indosat	4	4	3
XL	3	3	-1

Nilai sentimen seperti pada tabel 3.16 akan digunakan dalam perhitungan metode *TOPSIS* untuk mendapatkan operator seluler dengan sentimen terbaik. Sebelum proses dilanjutkan, dengan adanya ada yang minus maka perlu dilakukan normalisasi data terlebih dahulu. Normalisasi dilakukan menggunakan normalisasi min-max jangkauan [0, 1] dengan menggunakan rumus 2.10. Perhitungan normalisasi data dengan normalisasi min-max [0, 1] seperti berikut :

$$x(\text{telkomsel}, \text{sinyal}) = (5 - 3) / (5 - 3) = 1$$

$$x(\text{telkomsel}, \text{internet}) = (1 - 1) / (4 - 1) = 0$$

$$x(\text{telkomsel}, \text{paket}) = (1 - (-1)) / (3 - (-1)) = 0.5$$

$$x(\text{indosat}, \text{sinyal}) = (4 - 3) / (5 - 3) = 0.5$$

$$x(\text{indosat}, \text{internet}) = (4 - 1) / (4 - 1) = 1$$

$$x(\text{indosat}, \text{paket}) = (3 - (-1)) / (3 - (-1)) = 1$$

$$x(xl, \text{ sinyal}) = (3 - 3) / (5 - 3) = 0$$

$$x(xl, \text{ internet}) = (3 - 1) / (4 - 1) = 0.66667$$

$$x(xl, \text{ paket}) = (-1 - (-1)) / (3 - (-1)) = 0$$

Dari proses normalisasi maka didapatkan nilai sentimen pada masing - masing operator seluler seperti pada tabel 3.17

Tabel 3. 17 Hasil normalisasi min-max [0,1]

Nama	Kriteria		
	Sinyal	Internet	Paket
Telkomsel	1	0	0.5
Indosat	0.5	1	1
XL	0	0.66667	0

Setelah data nilai sentimen dilakukan normalisasi, maka tahap berikutnya pada perhitungan metode *TOPSIS* yakni menentukan kriteria , dimana untuk proses penentuan nilai data dilakukan dengan langkah pertama yang harus dilakukan sebelum menentukan pengelompokkan data yang nantinya digunakan sebagai penentuan nilai dari derajat keanggotaan berikut kriteria yang nantinya digunakan pada proses perhitungan :

Tabel 3. 18 Daftar kriteria berdasarkan ketentuan yang berlaku

No	Kriteria	Jenis Kriteria	Bobot
1.	Sinyal	Keuntungan (+)	0.33
2.	Internet	Keuntungan (+)	0.40
3.	Paket	Biaya (-)	0.27

Untuk proses berikutnya dilakukan dengan menentukan nilai matriks yang dilakukan dengan beberapa tahapan antara lain :

1. Pengkuadratan nilai kriteria
2. Matriks Pembagi
3. Normalisasi

Dengan rumus pada perhitungan persamaan 2.11, Berikut untuk hasil pengkuadratan kriteria

1. Telkomsel sinyal 1

$$\rightarrow 1^2 = 1$$

2. Indosat Harga 0.5

$$\rightarrow 0.5^2 = 0.25$$

3. XL sinyal 0

$$\rightarrow 0^2 = 0$$

Berikut untuk hasil pengkuadratan kriteria dapat dilihat pada tabel 3.19 :

Tabel 3. 19 Pengkuadratan Kriteria

Nama	Kriteria		
	Sinyal	Internet	Paket
Telkomsel	1	0	0.25
Indosat	0.25	1	1
XL	0	0.44444	0

Dari hasil kuadrat dilakukan proses penentuan nilai pembagi kriteria dengan menggunakan rumus menjumlahkan nilai dari keseluruhan data nilai kuadrat kemudian dilakukan proses pengakaran dari hasil jumlah tersebut sebagai berikut :

$$= \sqrt{k1^2 + k2^2 + k3^2}$$

$$= \sqrt{1.25}$$

$$= 1.11803$$

Dan seterusnya hingga ditemukan nilai pembagi dapat dilihat pada tabel 3.20 sebagai berikut :

Tabel 3. 20 Matriks Pembagi

Matriks Pembagi	Sinyal	Internet	Paket
	1.11803	1.2019	1.11803

Setelah ditemukan nilai pembagi maka dilakukan proses perhitungan normalisasi dari setiap kriteria dengan menggunakan matriks pembagi yang berhasil dihitung :

1. Telkomsel sinyal 1

$$R_{ij} = 1 / 1.11803$$

$$= 0.89443$$

2. Indosat sinyal 0.5

$$R_{ij} = 0.5 / 1.11803 \\ = 0.44721$$

3. XL sinyal 0

$$R_{ij} = 0 / 1.11803 \\ = 0$$

4. Dst hingga dilakukan proses perhitungan keseluruhan data setiap kriteria didapatkan hasil sebagai berikut dapat dilihat pada tabel 3.21 proses hasil normalisasi data :

Tabel 3. 21 Matrik Normalisasi

Nama	Kriteria		
	Sinyal	Internet	Paket
Telkomsel	0.89443	0	0.4472136
Indosat	0.44721	0.83205	0.8944272
XL	0	0.5547	0

Setelah ditentukan nilai normalisasi berikut untuk proses perhitungan dengan menggunakan data perhitungan bobot dari kriteria dan matriks normalisasi dengan rumus perhitungan persamaan 2.13, Berikut untuk detail perhitungan :

1. Telkomsel dengan normalisasi sinyal 0.89443

$$Y_{ij} = 0.89443 \times 0.33 \\ = 4.47214$$

2. Indosat dengan normalisasi sinyal 0.44721

$$Y_{ij} = 0.44721 \times 0.33 \\ = 2.23607$$

3. XL dengan normalisasi sinyal 0

$$Y_{ij} = 0 \times 0.33 \\ = 0$$

4. Dst hingga dilakukan proses perhitungan keseluruhan data setiap kriteria didapatkan hasil sebagai berikut, dapat dilihat pada tabel 3.22 proses hasil matriks normalisasi terbobot sebagai berikut :

Tabel 3. 22 Matrik Normalisasi Terbobot

Nama	Kriteria		
	Sinyal	Internet	Paket
Telkomsel	0.29814	0	0.119257
Indosat	0.14907	0.33282	0.2385139
XL	0	0.22188	0

Selanjutnya melakukan perhitungan matriks solusi ideal positif (A^+) dan matriks solusi ideal negatif (A^-) merupakan nilai minimum dan maximum dari setiap kriteria, nilai tersebut didapat dari proses perhitungan matriks normalisasi terbobot, berikut untuk rumus perhitungan persamaan 2.14, nilai dapat dilihat pada tabel 3.23.

Tabel 3. 23 ideal positif (A^+) Dan ideal positif (A^-)

	Sinyal	Internet	Paket
A^+	0.29814	0.33282	0
A^-	0	0	0.2385139

Selanjutnya menentukan jarak antara nilai setiap alternatif dengan matriks solusi ideal positif dan matriks solusi ideal negative, dimana untuk proses perhitungan digunakan sebagai penentuan nilai akhir dari proses metode, dengan menggunakan perhitungan persamaan 2.15. Berikut untuk detail perhitungan

1. Telkomsel

$$S_{i+} = \sqrt{(0.29814 - 0.29814)^2 + (0 - 0.33282)^2 + (0.119257 - 0)^2}$$

$$= 0.35354$$

2. Indosat

$$S_{i+} = \sqrt{(0.14907 - 0.29814)^2 + (0.33282 - 0.33282)^2 + (0.2385139 - 0)^2}$$

$$= 0.28127$$

3. XL

$$S_{i+} = \sqrt{(0 - 0.29814)^2 + (0.22188 - 0.33282)^2 + (0 - 0)^2}$$

$$= 0.31811$$

Dan seterusnya untuk matriks solusi ideal positif dan matriks solusi ideal negative dapat dilihat pada tabel 3.24 dan 3.25 berikut :

Tabel 3. 24 Jarak solusi ideal positif

Nama	S+
Telkomsel	0.35354
Indosat	0.28127
XL	0.31811

Dan untuk proses hasil matriks solusi ideal Negatif S_i^- sebagai berikut :

Tabel 3. 25 Jarak solusi ideal negatif

Nama	S-
Telkomsel	0.32111
Indosat	0.36468
XL	0.32576

Selanjutnya menentukan nilai preferensi untuk setiap alternatif (V_i), dengan perhitungan persamaan 2.16,

Berikut detail perhitungan nilai preferensi :

1. Telkomsel :

$$V_i = 0.3211 / (0.35354 + 0.32111) \\ = 0.4760$$

2. Indosat :

$$V_i = 0.36468 / (0.28127 + 0.36468) \\ = 0.5646$$

3. XL :

$$V_i = 0.32576 / (0.31811 + 0.32576) \\ = 0.5059$$

didapatkan hasil nilai preferensi dapat dilihat pada tabel 3.26 sebagai berikut :

Tabel 3.26 Proses perhitungan nilai preferensi

Nama	V_i
Telkomsel	0.4759638
Indosat	0.5645664
XL	0.5059375

Menentukan ranking dari setiap alternatif, dari nilai V_i terbesar sampai dengan V_i terkecil, Ranking operator seluler dapat dilihat pada tabel 3.27.

Tabel 3. 27 Hasil ranking nilai preferensi

Nama	Vi	Rangking
Telkomsel	0.4759638	3
Indosat	0.5645664	1
XL	0.5059375	2

Dari proses perhitungan nilai sentimen menggunakan metode *TOPSIS* diatas, didapatkan Vi sentimen pada masing – masing operator seluler dengan nilai tertinggi yakni Indosat nilai prefrensi 0.5645664 , XL nilai prefrensi 0.5059375 dan nilai terendah yakni Telkomsel nilai prefrensi 0.4759638. Maka kesimpulan didapatkan bahwa Indosat operator seluler yang terbaik dari Telkomsel dan XL.

3.3.2 Diagram Konteks

Diagram konteks merupakan diagram yang menunjukkan sebuah proses tunggal dalam sistem yang berhubungan dengan bagian yang terkait. Rangkaian diagram konteks yang digunakan pada penelitian ini seperti pada gambar 3.9

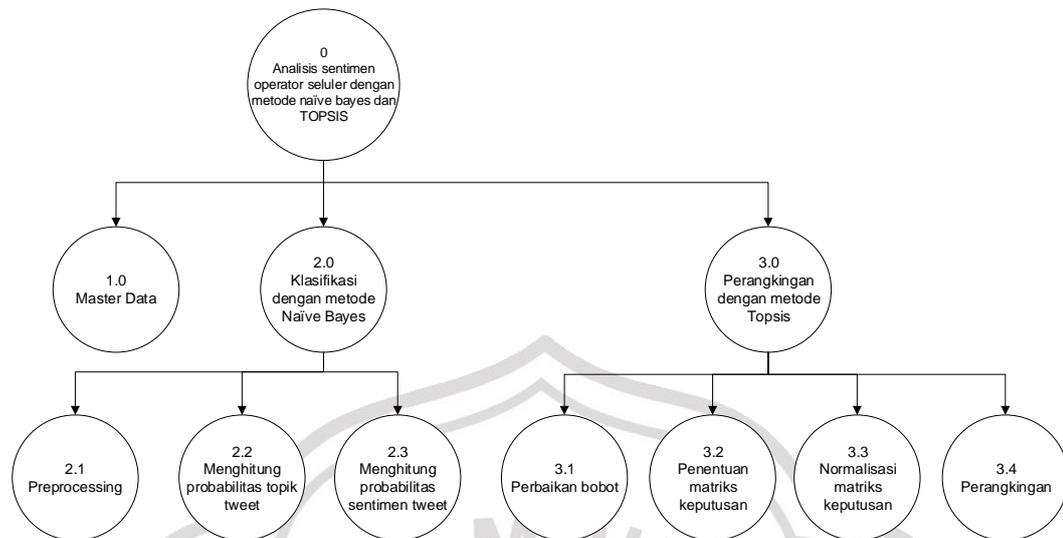


Gambar 3. 9 Diagram Konteks Analisis Sentimen pada Operator Seluler

Dari gambar 3.9 tersebut menggambarkan bahwa melibatkan dua pihak. Admin mengirimkan *input* berupa teks dari sosial media twitter mengenai masing-masing operator seluler yang digunakan sebagai data yang akan diproses. Setelah didapatkan hasil analisis maka *output* atau keluaran dari sistem berupa hasil perangkingan operator seluler berdasarkan sentimen yang didapatkan.

3.3.3 Diagram Berjenjang

Diagram berjenjang sangat diperlukan dalam perancangan semua proses yang ada. Diagram berjenjang merupakan penggambaran proses dari awal sampai ke level-level berikutnya. Dalam penelitian analisis sentimen ini mempunyai tiga level seperti pada gambar 3.10



Gambar 3. 10 Diagram berjenjang di analisis sentimen pada operator seluler

Berikut penjelasan gambar 3.10 berdasarkan kerangka diagram berjenjang diatas terlihat bahwa sistem yang dibuat terdiri dari tiga level

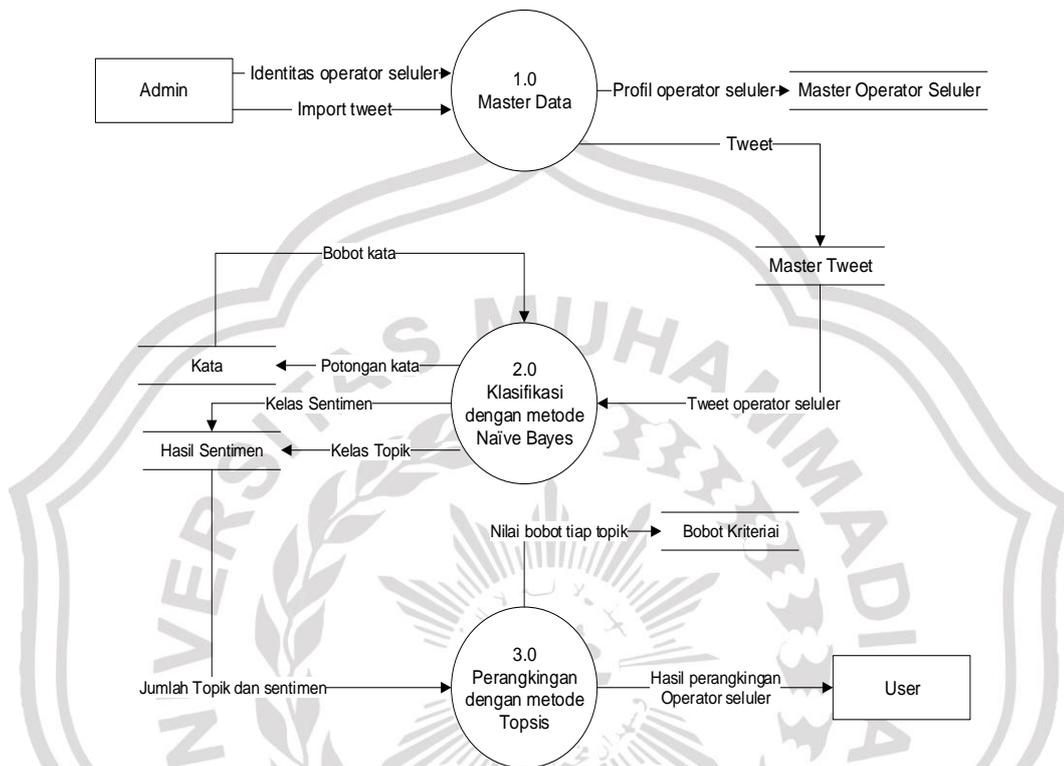
1. Top level : Analisis sentimen operator seluler dengan metode *Naïve Bayes* dan *TOPSIS*.
2. Level 0 : merupakan hasil *break down* dari proses keseluruhan dari analisis sentimen pada operator seluler menjadi beberapa sub proses yaitu :
 - a. Master data.
 - b. Klasifikasi dengan metode *Naïve Bayes*.
 - c. Perangkingan dengan metode *TOPSIS*.
3. Level 1 : merupakan sub proses dari beberapa proses pada level 0 dalam analisis sentimen pada operator seluler yang menggambarkan beberapa proses detail yaitu :
 - Hasil dari sub proses klasifikasi dengan metode *naïve bayes* :
 - a. *Preprocessing*.
 - b. Menghitung probabilitas topik *tweet*.
 - c. Menghitung probabilitas sentimen *tweet*.
 - Hasil dari sub proses perangkingan dengan metode *TOPSIS* :
 - a. Perbaikan bobot.
 - b. Penentuan matriks keputusan

c. Normalisasi matriks keputusan

d. Perangkingan

3.3.4 Data Flow Diagram

3.3.4.1 Data Flow Diagram level 0



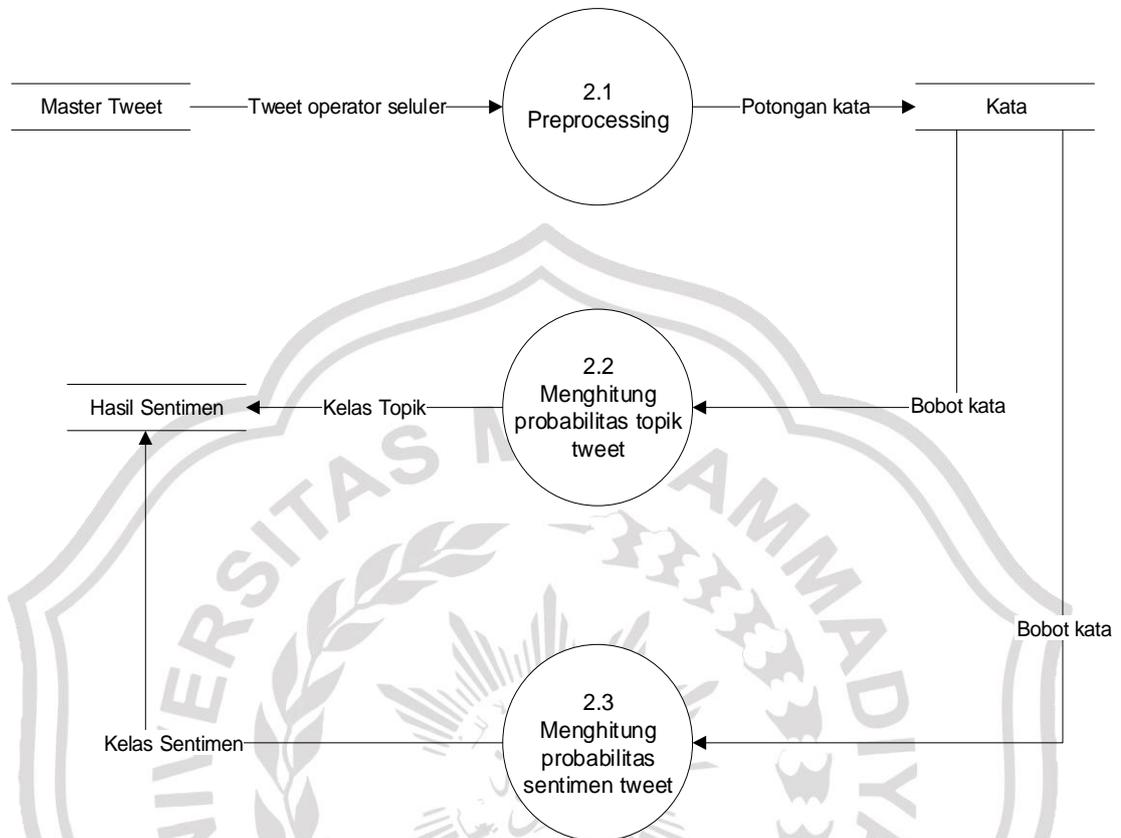
Gambar 3. 11 Data Flow Diagram level 0 analisis sentimen pada operator seluler

Berdasarkan pada gambar 3.11, dapat dijelaskan bahwa DFD level 0 menjelaskan beberapa proses yang terjadi. Beberapa proses yang ada pada DFD level 0 antara lain :

- a. Master data
- b. Klasifikasi dengan metode Naïve Bayes
- c. Perangkingan dengan metode TOPSIS

3.3.4.2 Data Flow Diagram level 1

1) DFD Level 1 Proses 2

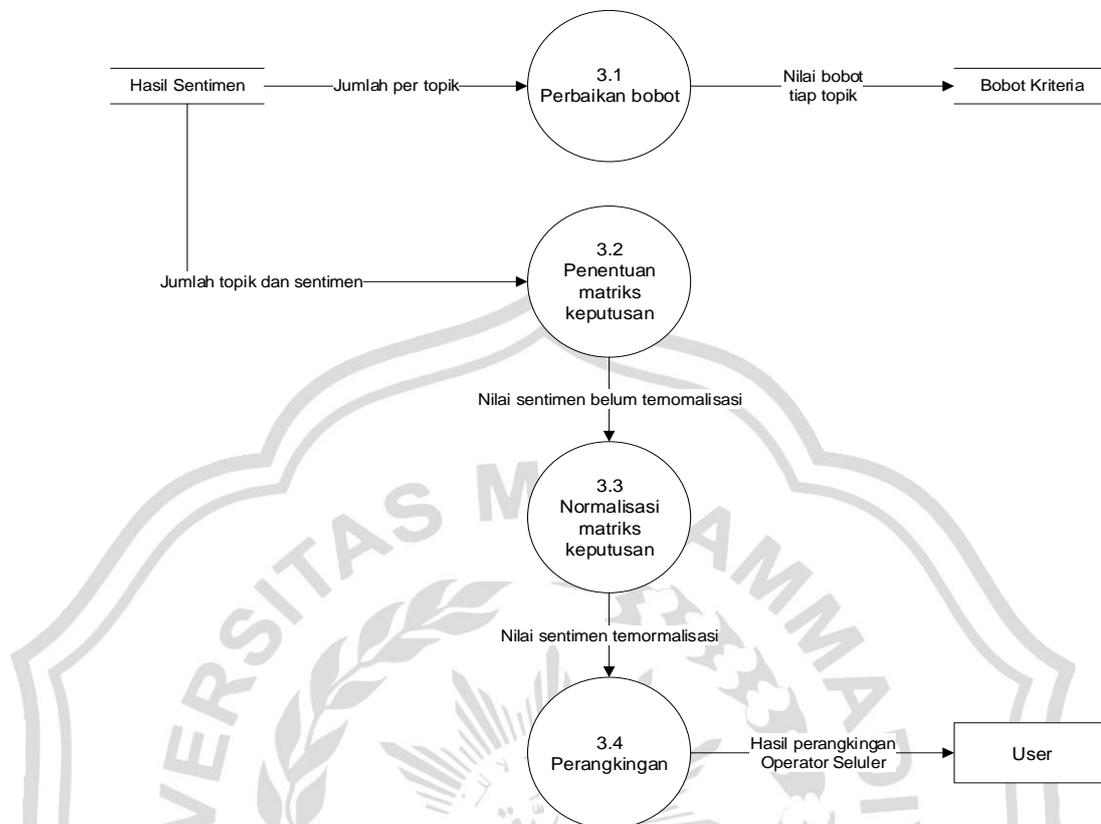


Gambar 3. 12 Data Flow Diagram level 1 untuk proses kedua analisis sentimen pada operator seluler

Berdasarkan pada gambar 3.12 dapat dijelaskan bahwa DFD Level 1 proses 2 menjelaskan tiga proses detail dari proses klasifikasi yakni :

- a. *Preprocessing.*
- b. Menghitung probabilitas topik *tweet.*
- c. Menghitung probabilitas sentimen *tweet.*

2) DFD Level 1 Proses 3



Gambar 3. 13 Data Flow Diagram level 1 untuk proses ketiga analisis sentimen pada operator seluler

Berdasarkan pada gambar 3.13 dapat dijelaskan bahwa DFD Level 1 proses 3 menjelaskan empat proses detail dari proses klasifikasi yakni :

- a. Perbaikan bobot.
- b. Penentuan matriks keputusan.
- c. Normalisasi matriks keputusan.
- d. Perangkingan

3.4 Perancangan Basis Data

Basis data adalah kumpulan file-file yang mempunyai kaitan antara satu file dengan file lain sehingga membentuk suatu bangunan data untuk menginformasikan suatu hasil dari sebuah proses. Berikut untuk struktur dan desain tabel dari *database* yang digunakan dalam proses pembuatan sistem analisis sentimen pada operator seluler di media sosial *twitter*.

3.4.1 Desain tabel

Desain tabel merupakan susunan dari tabel yang akan digunakan atau diimplementasikan kedalam database, dimana desain tabel memuat detail tipe data dan *primary key* serta *foreign key* dari tabel tersebut.

1. Tabel Master *Tweet*

Tabel master *tweet* digunakan sebagai master kalimat yang akan diproses, pada tabel master *tweet* merupakan awal dari sebuah proses analisis sentimen. Atribut yang ada pada tabel master *tweet* adalah *id_tweet*, *isi_tweet* dan *id_operator_seluler*. Skema tabel master *tweet* seperti pada tabel 3.28

Tabel 3. 28 tabel master *tweet*

Nama kolom	Tipe	Ukuran	Keterangan
<i>Id_tweet</i>	Int	15	<i>Primary key</i>
<i>Isi_tweet</i>	Text		
<i>Id_operator_seluler</i>	Int	15	<i>Foreign key</i>

2. Tabel Master Operator Seluler

Tabel master operator seluler digunakan sebagai master yang berisikan para operator-operator seluler yang akan diketahui hasil analisis sentimennya. Atribut yang ada pada tabel master operator seluler adalah *id_operator_seluler* dan *nama_operator*. Skema tabel master operator seluler seperti pada tabel 3.29

Tabel 3. 29 tabel master operator seluler

Nama kolom	Tipe	Ukuran	Keterangan
<i>Id_operator_seluler</i>	Int	15	<i>Primary key</i>
<i>Nama_operator</i>	Varchar	50	

3. Tabel kata *stopword*

Tabel kata *stopword* digunakan sebagai kata pembanding yang akan digunakan dalam penghapusan kata-kata yang tidak memiliki makna atau *stopword removal*. Atribut yang ada pada tabel kata *stopword* adalah

id_stopword dan kata_stopword. Skema tabel kata *stopword* seperti pada tabel 3.30

Tabel 3. 30 tabel kata *stopword*

Nama kolom	Tipe	Ukuran	Keterangan
Id_stopword	Int	15	<i>Primary key</i>
Kata_stopword	Varchar	100	

4. Tabel kata dasar

Tabel kata dasar digunakan sebagai kata pembanding untuk menentukan kata dasar pada proses *stemming*. Atribut yang ada pada tabel kata dasar adalah id_kata_dasar, kata_dasar dan tipe_kata. Skema tabel kata dasar seperti pada tabel 3.31

Tabel 3. 31 tabel kata dasar

Nama kolom	Tipe	Ukuran	Keterangan
Id_kata_dasar	Int	15	<i>Primary key</i>
Kata_dasar	Varchar	100	
Tipe_kata	Varchar	100	

5. Tabel kata

Tabel kata digunakan sebagai tempat untuk menyimpan kumpulan kata hasil preprocessing pada kalimat *twitter* yang akan digunakan dalam perhitungan bobot kata. Atribut yang ada pada tabel kata adalah id_kata, id_tweet, kata. Skema tabel kata seperti pada tabel 3.32

Tabel 3. 32 Skema pada tabel kata

Nama kolom	Tipe	Ukuran	Keterangan
Id_kata	Int	15	<i>Primary</i>
Id_tweet	Int	15	<i>Foreign key</i>
Kata	Varchar	100	

6. Tabel Sentimen

Tabel sentimen digunakan sebagai penyimpanan hasil pemrosesan analisis sentimen mulai dari pembobotan *tweet* terhadap topik maupun *tweet*

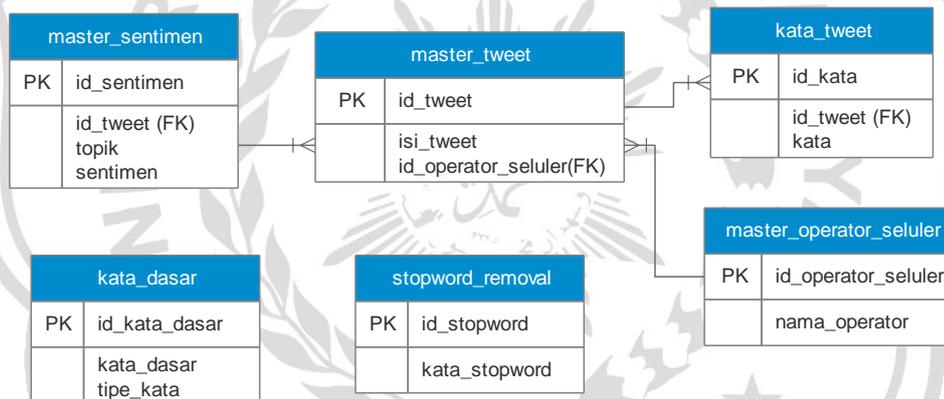
terhadap sentimen. Atribut yang ada pada tabel sentimen adalah id_sentimen, id_tweet, topik dan sentimen. Skema tabel sentimen seperti pada gambar 3.33

Tabel 3. 33 Skema tabel hasil sentimen

Nama kolom	Tipe	Ukuran	Keterangan
Id_sentimen	Int	15	<i>Primary key</i>
Id_tweet	Int	15	<i>Foreign key</i>
Topik	Varchar	30	
Sentimen	Varchar	30	

3.4.2 Entity Relationship Diagram

Entity relationship diagram (ERD) merupakan model konseptual yang menggambarkan hubungan antar tabel yang ada. ERD digunakan untuk memodelkan struktur data dan hubungan antar data. Desain *entity relationship diagram* pada pembuatan sistem analisis sentimen pada operator seluler seperti pada gambar 3.14



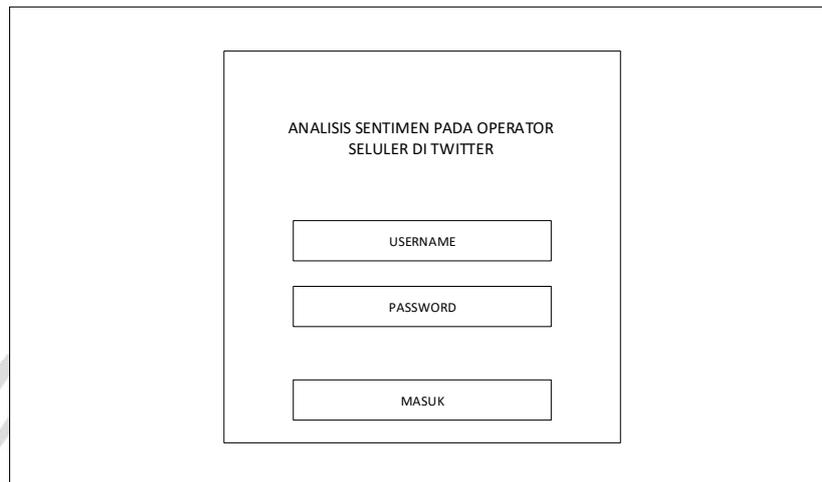
Gambar 3. 14 Entity relationship diagram dalam analisis sentimen operator seluler

3.5 Perancangan Antarmuka

Rancangan antarmuka (*interface*) berfungsi sebagai alat komunikasi antara sistem dengan pengguna. Antarmuka akan memberikan informasi berupa tampilan disertai dengan data-data yang diminta oleh pengguna. Dalam penelitian ini desain antarmuka dapat digunakan sebagai media pemasukan data *tweet* dan menampilkan hasil dari sentimen operator seluler.

3.5.1 Halaman *Login*

Halaman login merupakan halaman awal dari sistem. Admin harus memasukkan *username* dan *password* yang telah didaftarkan untuk masuk ke dalam sistem. Tampilan rancangan halaman login seperti pada gambar 3.15



ANALISIS SENTIMEN PADA OPERATOR
SELULER DI TWITTER

USERNAME

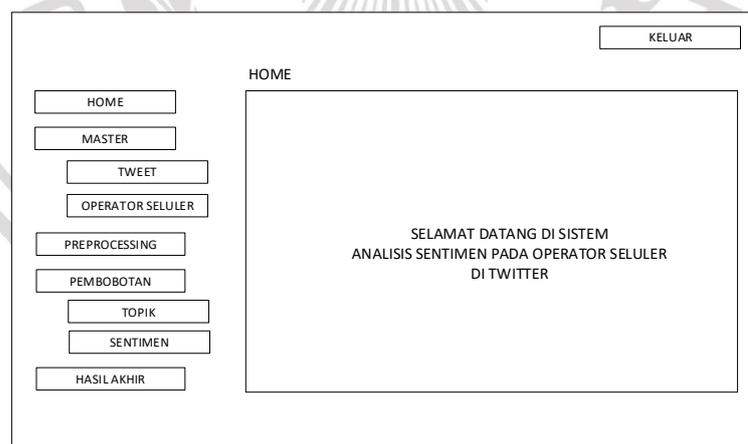
PASSWORD

MASUK

Gambar 3.15 Tampilan rancangan halaman *login*

3.5.2 Halaman *Home*

Halaman *home* merupakan halaman pertama yang muncul ketika pengguna berhasil *login*. Halaman ini berisikan informasi mengenai nama sistem dan kegunaannya. Tampilan rancangan halaman home seperti pada gambar 3.16



KELUAR

HOME

HOME

MASTER

TWEET

OPERATOR SELULER

PREPROCESSING

PEMBOBOTAN

TOPIK

SENTIMEN

HASIL AKHIR

SELAMAT DATANG DI SISTEM
ANALISIS SENTIMEN PADA OPERATOR SELULER
DI TWITTER

Gambar 3. 16 Tampilan rancangan halaman home

3.5.3 Halaman *Master Tweet*

Halaman master *tweet* berfungsi untuk menampilkan data-data yang didapatkan dari *tweet*. Halaman ini berisikan teks twitter, operator seluler, id twitter dan aksi untuk menghapus ataupun mengubah teks. Tampilan rancangan halaman master *tweet* seperti pada gambar 3.17

No	Operator Seluler	Id Twitter	Text Tweet	Aksi

Gambar 3. 17 Tampilan rancangan halaman master *tweet*

3.5.4 Halaman *Master Operator Seluler*

Halaman master operator seluler berfungsi untuk menampilkan data operator seluler yang ingin diketahui analisis sentimennya. Halaman ini berisikan nama operator serta aksi untuk menambahkan, mengubah atau menghapus daftar operator seluler. Tampilan rancangan halaman master operator seluler seperti pada gambar 3.18

No	Operator Seluler	Aksi

Gambar 3. 18 Tampilan rancangan halaman master operator seluler

3.5.5 Halaman Tambah *Tweet*

Halaman tambah *tweet* berfungsi untuk admin agar bisa menambahkan teks *tweet* baru yang masih belum diketahui topik dan hasil sentimennya. Halaman ini diberi akses untuk admin menambahkan id twitter, teks *tweet* dan operator seluler yang terkait dengan teks *tweet* tersebut. Tampilan rancangan halaman tambah *tweet* seperti pada gambar 3.19

Gambar 3. 19 Tampilan rancangan halaman tambah *tweet*

3.5.6 Halaman Tambah Operator Seluler

Halaman tambah operator seluler merupakan halaman yang bisa digunakan oleh admin untuk menambahkan operator seluler yang ingin diketahui analisa sentimennya. Admin bisa menuliskan operator seluler untuk disimpan didalam sistem. Tampilan rancangan halaman tambah operator seluler seperti pada gambar 3.20

Gambar 3. 20 Tampilan rancangan halaman tambah operator seluler

3.5.7 Halaman *Preprocessing*

Halaman *preprocessing* berfungsi untuk melakukan tahap atau proses *preprocessing*, dimana teks yang dimasukkan akan di proses sehingga didapatkan kata atau *term* dari teks tersebut. Tampilan rancangan halaman *preprocessing* seperti pada gambar 3.21

Gambar 3. 21 Tampilan rancangan halaman preprocessing

3.5.8 Halaman Pembobotan Topik

Halaman pembobotan topik akan menampilkan hasil proses perhitungan pembobotan *tweet* topik. Halaman ini menampilkan teks *tweet* serta hasil pembobotan pada teks tersebut. Tampilan rancangan halaman pembobotan topik seperti pada gambar 3.22

Gambar 3. 22 Tampilan rancangan halaman pembobotan topik

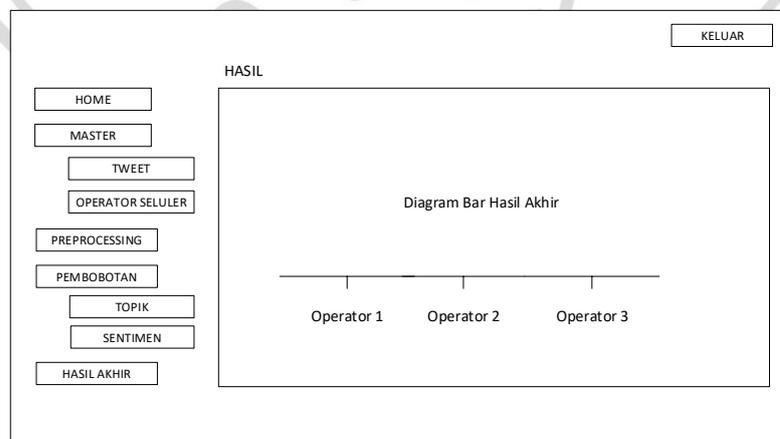
3.5.9 Halaman Pembobotan Sentimen

Halaman pembobotan ini merupakan tahap setelah dilakukannya pembobotan topik. Halaman ini akan menampilkan hasil pembobotan teks *tweet* sehingga didapatkan hasil sentimen dari teks tersebut. Tampilan rancangan halaman pembobotan sentimen seperti pada gambar 3.23

Gambar 3. 23 Tampilan rancangan halaman pembobotan sentimen

3.5.10 Halaman Hasil Sentimen

Halaman hasil sentimen akan menampilkan hasil akhir perhitungan dari data sentimen yang sudah didapatkan. Halaman hasil sentimen menampilkan nilai akhir sentimen setiap operator seluler serta menampilkan detail perhitungan dengan metode *TOPSIS*. Tampilan rancangan halaman hasil sentimen seperti pada gambar 3.24



Gambar 3. 24 Tampilan rancangan halaman hasil sentimen

3.6 Skenario Pengujian

Sebuah pengujian dalam pembuatan sistem analisis perlu dilakukan. Pengujian dalam hal ini digunakan untuk mengetahui tingkat akurasi dari perhitungan dengan metode *naïve bayes classifier*. Skenario pengujian pada penelitian ini menggunakan perhitungan *matriks confusion*. Skenario pengujian dalam penentuan analisis sentimen untuk mendapatkan nilai *accuracy*. Nilai *accuracy* merupakan tingkat kedekatan antara nilai prediksi dengan nilai aktual (Dataq, 2013).

Dari data yang telah diperoleh yakni 300 data latih dan 300 data uji akan menentukan topik dan sentimen dari ke 300 data uji tersebut. Penentuan topik dan sentimen menggunakan rumus (2.6, 2.7, 2.8) yakni *naïve bayes* setiap data uji akan dilakukan perhitungan dengan menghitung masing-masing bobot pada topik dan sentimen dengan bobot kata. Dari beberapa perhitungan pada beberapa topik maka topik dengan nilai tertinggi yang menjadi hasil dari data uji tersebut. Setelah diketahui topik dan sentimen maka akan menentukan nilai akurasi dengan rumus (2.9).

3.6.1 Evaluasi Sistem

Tahapan evaluasi berfungsi untuk menghitung seberapa baik sistem dalam mengidentifikasi *tweet* positif dan negatif yang diujikan di dalam sistem. Tahapan ini akan menggunakan persamaan yang sudah dijabarkan pada bagian (2.9,2.10,2.11,2.12). Sebagai contoh untuk tahapan evaluasi, digunakan data contoh pada Tabel 3.34.

Tabel 3. 34 Contoh nilai sentimen evaluasi sistem

No	Variable	Jumlah
1	True Positif (TP)	113
2	False Negatif (FN)	32
3	False Positif (FP)	2
4	True Negatif (TN)	3
Total		150

Dari Tabel 3.28 dapat dihitung nilai akurasi, *recall*, *precision*, dan *F-score* yang secara berurutan menggunakan persamaan (2.9), (2.10), (2.11), dan (2.12) sebagai berikut:

$$1. \text{ Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{116}{150} = 0.77$$

$$\begin{aligned}
 2. \text{ Recall} &= \frac{TP}{TP+FN} = \frac{113}{113+32} = 0.78 \\
 3. \text{ Precision} &= \frac{TP}{TP+FP} = \frac{113}{113+2} = 0.98 \\
 4. \text{ F-score} &= \frac{2(\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} = \frac{2(0.78 \times 0.98)}{0.78 + 0.98} = 0.87
 \end{aligned}$$

Oleh karena nilai akurasi, *recall*, *precision*, dan *F-score* dinyatakan dalam bentuk persentase, maka nilai yang sudah diperoleh akan dikalikan dengan 100%, sehingga nilai akurasi, *recall*, *precision*, dan *F-score* secara berurutan adalah 77%, 78%, 98%, dan 87%.

3.7 Spesifikasi Pembuatan Sistem

Pembuatan sistem untuk proses analisis sentimen pada operator seluler di media sosial twitter membutuhkan spesifikasi dalam pembuatannya. Pembuatan sistem membutuhkan spesifikasi dari perangkat lunak maupun perangkat keras. Kebutuhan perangkat lunak serta perangkat keras dari sistem sebagai berikut :

A. Kebutuhan Perangkat Lunak

Perangkat Lunak (*Software*) adalah program-program yang digunakan untuk menjalankan sistem perangkat keras, diantaranya adalah sistem operasi, bahasa pemrograman dan program aplikasi. Pembuatan sistem diperlukan perangkat-perangkat lunak yang sangat mendukung kinerja program, agar dapat mencapai hasil yang sempurna dari sistem tersebut. Perangkat lunak yang diperlukan dalam pembuatan aplikasi adalah sebagai berikut :

1. Sistem Operasi Windows (7,8,10)
2. Xampp Versi 5.0
3. Server *Database* MySQL
4. Sublime Text Editor

B. Kebutuhan Perangkat Keras

Sistem perangkat keras (*Hardware*) adalah komponen komponen pendukung kinerja dari sistem komputer. Komponen komponen yang dapat dipakai untuk menjalankan aplikasi penjualan adalah sebagai berikut:

1. Prosesor Intel Coleron CPU B816 1.60 Ghz
2. Memory RAM 2 GB
3. Monitor VGA 14 inch

4. Harddisk 500 GB
5. Keyboard
6. Mouse
7. Printer

