

## **BAB III**

### **ANALISIS DAN PERANCANGAN SISTEM**

#### **3.1 Analisis Sistem**

Sistem temu kembali informasi (*information retrieval*) adalah sistem yang dapat menemukan kembali (*retrieve*) informasi yang relevan dengan kata kunci (*keyword*) yang diinginkan oleh pengunannya. Penerapan sistem temu kembali informasi pada umumnya digunakan untuk pencarian dokumen juga bisa digunakan dalam analisa sentimen. Analisa sentimen merupakan penilaian terhadap pendapat atau komentar orang terhadap beberapa topik sehingga didapatkan bahwa pendapat tersebut termasuk pendapat yang positif, netral 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 tokoh politik. Pendapat dari masyarakat mengenai tokoh politik menjadi perlu untuk diketahui kelas topik dan sentimen yang terdapat pada *tweet* tersebut. Topik yang dibahas merupakan tema yang dihubungkan dengan tokoh politik tersebut, sedangkan untuk sentimen merupakan nilai dari *tweet* yang membahas tokoh politik yang dimaksud sehingga dapat diketahui nilai positif, netral atau negatif yang terkandung dalam *tweet* tersebut.

#### **3.2 Hasil Analisis**

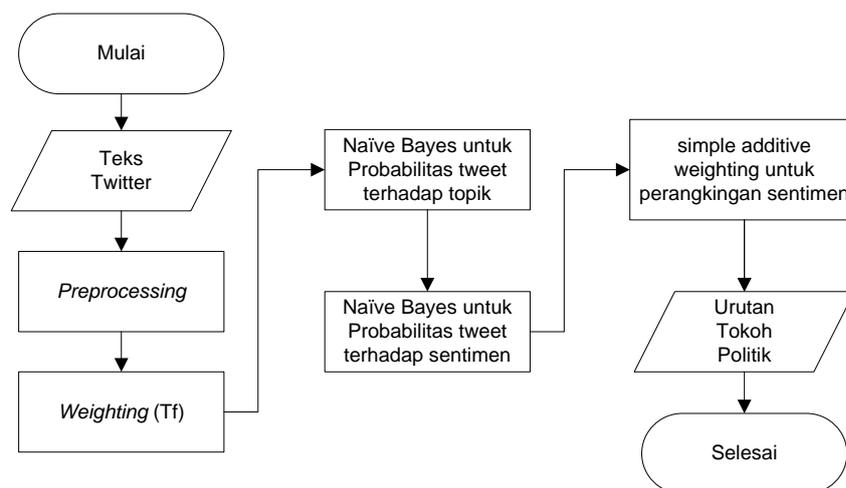
Hasil analisis yang dapat dilakukan dari pengklasifikasian topik dan sentimen *tweet* yang membahas tokoh politik dapat melakukan klasifikasi jenis topik dan sentimen yang terkandung dalam *tweet* di sosial media twitter. Pembuatan sistem klasifikasi topik dan sentimen tokoh politik ini menggunakan metode *Naïve Bayes*. Penggunaan metode *Naïve Bayes* memerlukan data *tweet* yang terlebih dahulu sudah memiliki label topik dan sentimen untuk dijadikan sebagai data latih atau data acuan dalam menentukan kelas topik dan sentimen untuk *tweet* yang

baru. *Tweet* yang akan dilakukan klasifikasi akan diberi salah satu topik dari lima topik yang dijadikan sebagai kelas topik yakni Integritas, Program Kerja, Perilaku, Kinerja dan Kapabilitas. Penggunaan kelima 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* untuk mendapatkan kelas sentimen positif, netral atau negatif. Sehingga masing-masing *tweet* akan memiliki kelas topik dan sentimen yang dapat digunakan untuk mengukur tingkat elektabilitas dari tokoh politik bersangkutan.

### 3.3 Perancangan Sistem

#### 3.3.1 Flowchart sistem

Sistem yang akan dibangun adalah aplikasi dengan konsep sistem temu kembali informasi yang dapat menganalisa sentimen terhadap tokoh politik yang berada di media sosial *twitter* menggunakan metode *naïve bayes* untuk proses klasifikasi tweet. Tujuan dari sistem adalah untuk memberikan informasi mengenai sentimen masyarakat terhadap tokoh politik yang akan di kelompokkan ke sentimen positif, negatif atau netral. 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 pembahasan sistem temu kembali informasi atau *information retrieval* biasanya sebuah inputan data adalah sebuah teks dokumen, tetapi 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 API (*Application Programming Interface*) yang sudah disiapkan oleh media sosial *twitter* dengan *keyword* pencariannya yakni nama tokoh politik. Informasi *tweet* akan disimpan terlebih dahulu di *database* dan dikelompokan berdasarkan *keyword* tokoh politik 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

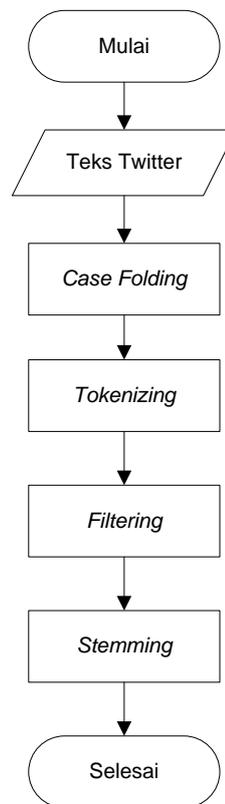
T	Tokoh	<i>Tweet</i>	Topik	Sentimen
T1	Joko Widodo	Apakah semua janji-janjinya pada masyarakat terwujudkan ? sudah pasti terwujud janji-janjinya yang pernah disampaikan ke masyarakat	Integritas	Positif
T2	Joko Widodo	Belum meratanya pembangunan menjadi bukti, bahwa gagalnya pembangunan daerah tertinggal.	Proker	Negatif
T3	Joko Widodo	Salah satu ruas jalan di sulawesi selatan yang rusak telah selesai diperbaiki, langkah bagus untuk mengantisipasi arus mudik	Proker	Positif
T4	Joko Widodo	Pendukung tidak ada yang memviralkan harga sembako melonjak tinggi, walau mendekati lebaran tapi sembako tetap tinggi	Proker	Negatif
T5	Joko Widodo	Pembukaan workshop dan inovasi desa menjadi awal kebangkitan usaha desa	Proker	Positif
T6	Joko Widodo	Seluruh desa di kaltim dan kaltara sudah berlistrik. Terima kasih pak	Proker	Positif
T7	Joko Widodo	Tutur bahasa yang pantas dalam berbicara, logat bicara tidak asal, pantas dalam bertingkah	Perilaku	Netral
T8	Joko	Sebuah pekerjaan bukan hanya	Kinerja	Positif

	Widodo	wacana saja tapi harus berhasil terwujud. Bekerja menjadikan semuanya terwujud, wujud keberhasilan dan kesuksesan. Jangan hanya wacana saja yaa.		
T9	Joko Widodo	Tak pandai-pandai, bodoh, saya ragu dia pernah lulus sekolah	Kapabilitas	Negatif
T10	Joko Widodo	Kesuksesan dalam pembangunan menghasilkan kesejahteraan. Semoga sukses dan terus berhasil	?	?

Dari tabel 3.1 diatas bahwa terdapat sepuluh teks twitter yang telah dilakukan *crawler*. Kesepuluh teks tersebut terdiri dari lima teks twitter yang digunakan sebagai data latih yakni T1, T2, T7, T8, T9 dan terdapat satu teks yang masih belum diketahui topik dan sentimennya yakni T10. Data latih didapatkan dengan membaca teks lengkap twitter dan melakukan perbandingan untuk menentukan topik dan sentimen dengan acuan basis data *opinion word* pada tabel 2.5, 2.6 dan 2.7. Cara menentukan topik dan sentimen yang digunakan sebagai data latih dengan mencocokkan dengan kumpulan kata positif, netral dan negatif, jika pada kumpulan kata-kata *tweet* tersebut diambil kategori sentimen yang dominan. Pada *tweet* pertama terdapat kata janji dan wujud yang termasuk kedalam sentimen positif, sedangkan terdapat kata masyarakat yang termasuk kedalam kategori netral. 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*, *tokenizing*, *filtering*, *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 *tokenizing* merupakan proses setelah *case folding*, proses *tokenizing* merupakan proses memecah teks yang awalnya satu kalimat dirubah menjadi perkata. Proses *filtering* yakni melakukan pembuangan kata-kata yang tidak memilik makna dan menghapus kata yang muncul berkali-kali sehingga kata yang muncul berkali-kali tidak mempengaruhi bobot kata tersebut. Setelah proses *filtering* 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, T2, T7, T8, T9, T10 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>
apakah semua janji-janjinya pada masyarakat terwujudkan ? sudah pasti terwujud janji-janjinya yang pernah disampaikan ke masyarakat
belum meratanya pembangunan menjadi bukti, bahwa gagalnya pembangunan daerah tertinggal.
tutur bahasa yang pantas dalam berbicara, logat bicara tidak asal, pantas dalam bertingkah
sebuah pekerjaan bukan hanya wacana saja tapi harus berhasil terwujud. bekerja menjadikan semuanya terwujud, wujud keberhasilan dan kesuksesan. jangan hanya wacana saja yaa.
tak pandai-pandai, bodoh, saya ragu dia pernah lulus sekolah
kesuksesan dalam pembangunan menghasilkan kesejahteraan. semoga sukses dan terus berhasil

### 3.3.1.4 *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 3.2 yang sudah dilakukan *case folding* akan dilakukan *tokenizing* sehingga menjadi potongan kata seperti pada tabel 3.3 seperti berikut.

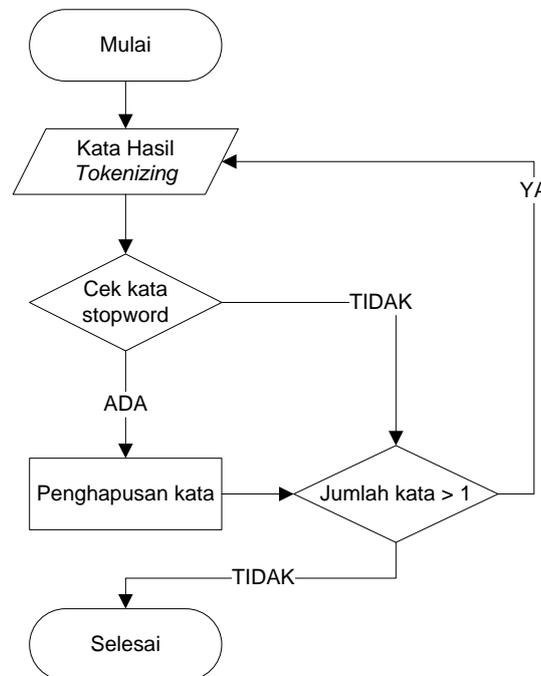
**Tabel 3. 3** Term hasil proses *tokenizing*

<i>Tweet</i>
apakah   semua   janji   -   janjinya   pada   masyarakat   terwujudkan   ?   sudah   pasti   terwujud   janji   -   janjinya   yang   pernah   disampaikan   ke   masyarakat
belum   meratanya   pembangunan   menjadi   bukti   ,   bahwa   gagalnya   pembangunan   daerah   tertinggal.
tutur   bahasa   yang   pantas   dalam   berbicara   ,   logat   bicara   tidak   asal   ,   pantas   dalam   bertingkah
sebuah   pekerjaan   bukan   hanya   wacana   saja   tapi   harus   berhasil   terwujud   .   bekerja   menjadikan   semuanya   terwujud   ,   wujud   keberhasilan   dan   kesuksesan   .   jangan   hanya   wacana   saja   yaa.
tak   pandai   -   pandai   ,   bodoh   ,   saya   ragu   dia   pernah   lulus   sekolah

kesuksesan | dalam | pembangunan | menghasilkan | kesejahteraan | . | semoga | sukses | dan | terus | berhasil

### 3.3.1.5 Filtering

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



**Gambar 3. 4** Alur proses *filtering*

Kata hasil proses *tokenizing* digunakan sebagai inputan pada proses *filtering*. Setelah didapatkan kata tersebut maka tiap kata akan dilakukan pengecekan kata dengan data pembanding yakni kata yg ada dalam stopwords. Jika kata tersebut termasuk dalam kumpulan kata stopwords maka kata tersebut akan dihapus. Setelah proses pengecekan kata stopwords dilakukan maka dilakukan pengecekan kembali untuk jumlah kata. Jika dalam pengecekan ditemukan jumlah kata lebih dari satu maka kata tersebut akan dilakukan pengecekan stopwords

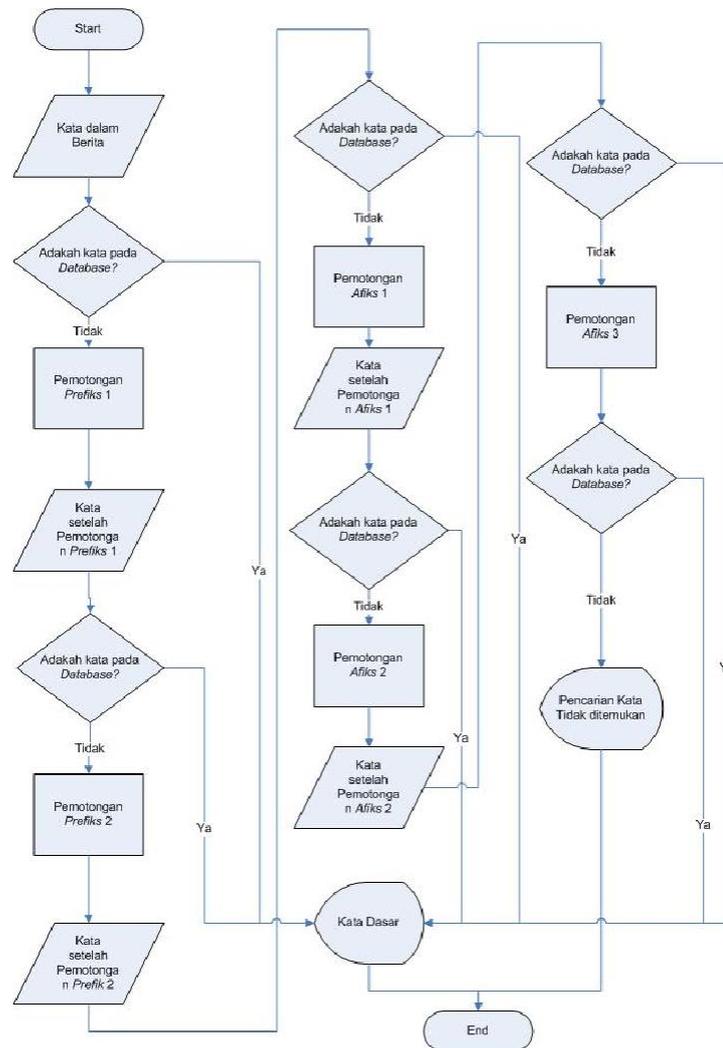
kembali. Hasil proses *filtering* yang didapatkan dari *term* pada tabel 3.3 seperti pada tabel 3.4 berikut.

**Tabel 3. 4** Term hasil proses filtering

<i>Tweet</i>
janji   janjinya   masyarakat   terwujudkan   terwujud   janji   janjinya   masyarakat
meratanya   pembangunan   bukti   gagalnya   pembangunan   tertinggal.
bahasa   pantas   berbicara   bicara   asal   pantas   bertingkah
pekerjaan   wacana   berhasil   terwujud   bekerja   terwujud   wujud   keberhasilan   kesuksesan   wacana
pandai   pandai   bodoh   ragu   lulus   sekolah
kesuksesan   pembangunan   menghasilkan   kesejahteraan   sukses   berhasil

### 3.3.1.6 Stemming

Proses setelah *filtering* adalah proses *stemming*. Proses *stemming* merupakan proses pencarian kata dasar terhadap sebuah kata yang telah dilakukan proses *filtering*. 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 Nazief-Adriani. 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 nazief adriani. Hasil proses stemming yang dilakukan dengan menggunakan tabel 3.4 menghasilkan seperti pada tabel 3.5

**Tabel 3. 5** Term hasil proses *stemming* (korpus)

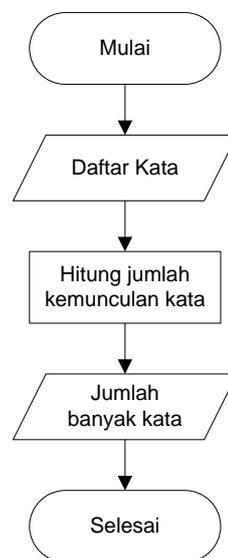
<i>Tweet</i>
janji   janji   masyarakat   wujud   wujud   janji   janji   masyarakat
rata   bangun   bukti   gagal   bangun   tinggal.
bahasa   pantas   bicara   bicara   asal   pantas   tingkah
kerja   wacana   hasil   wujud   kerja   wujud   wujud   hasil   sukses   wacana
pandai   pandai   bodoh   ragu   lulus   sekolah

sukses | bangun | hasil | sejahtera | sukses | hasil

### 3.3.1.7 Perhitungan Analisis Sentimen (Representasi Data)

#### 3.3.1.7.1 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 3.6 berikut.

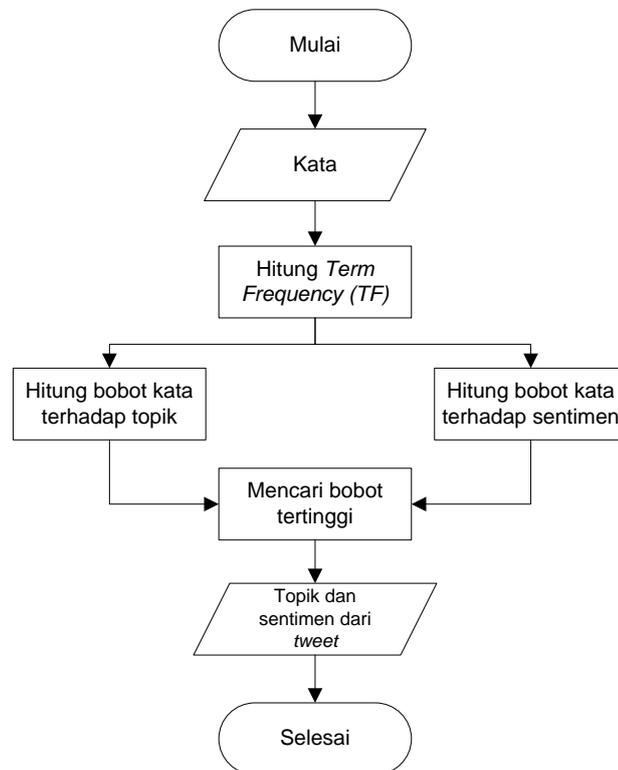
**Tabel 3. 6** term dan bobot hasil *term frequency*

Topik	Sentimen	<i>Tweet</i>
Integritas	Positif	janji (4)   masyarakat (2)   wujud (2)
Proker	Negatif	rata (1)   bangun (2)   bukti (1)   gagal (1)   tinggal (1).
Perilaku	Netral	bahasa (1)   pantas (2)   bicara (2)   asal (1)   tingkah (1)
Kinerja	Positif	kerja (2)   wacana (2)   hasil (2)   wujud (3)   sukses

		(2)
Kapabilitas	Negatif	pandai (2)   bodoh (1)   ragu (1)   lulus (1)   sekolah (1)
?	?	sukses (2)   bangun (1)   hasil (2)   sejahtera (1)

### 3.3.1.7.2 Perhitungan *Naïve Bayes*

Penelitian ini menggunakan metode *naïve bayes* 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* seperti pada gambar 3.7



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

Perhitungan *naïve bayes* 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.

### 3.3.1.7.3 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 6 teks *twitter* awal seperti pada tabel 3.1 yang telah dilakukan *preprocessing* dan masing-masing kata sudah memiliki fitur kemunculan seperti pada tabel 3.6.

Integritas : janji(4), wujud(2), masyarakat(2)

Proker : bangun(2), rata(1), bukti(1), gagal(1), tinggal(1)

Perilaku : bahasa(1), bicara(2), pantas(2), asal(1), tingkah(1)

Kinerja : kerja(2), wujud(3), hasil(2), wacana(2), sukses(1)

Kapabilitas : pandai(2), bodoh(1), ragu(1), lulus(1), sekolah(1)

? : sukses(2), bangun(1), sejahtera(1), hasil(2)

**Tabel 3. 7** Fitur kemunculan kata

Teks	Janji	Wujud	Masyarakat	Bangun	Rata	Bukti	Gagal	Tinggal
T1	4	2	2	0	0	0	0	0
T2	0	0	0	2	1	1	1	1
T3	0	0	0	0	0	0	0	0
T4	0	3	0	0	0	0	0	0
T5	0	0	0	0	0	0	0	0
T6	0	0	0	1	0	0	0	0

Teks	Bahasa	Bicara	Pantas	Asal	Tingkah	Kerja	Hasil	Wacana
T1	0	0	0	0	0	0	0	0
T2	0	0	0	0	0	0	0	0
T3	1	2	2	1	1	0	0	0
T4	0	0	0	0	0	2	2	1
T5	0	0	0	0	0	0	0	0
T6	0	0	0	0	0	2	2	0

Teks	Sukses	Pandai	Bodoh	Ragu	Lulus	Sekolah	Sejahtera
T1	0	0	0	0	0	0	0
T2	0	0	0	0	0	0	0
T3	0	0	0	0	0	0	0
T4	1	0	0	0	0	0	0
T5	0	2	1	1	1	1	0

T6	2	0	0	0	0	0	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  $p(ci)$  dan probabilitas pada tiap kata  $p(wkj.ci)$ . Menentukan nilai probabilitas topik menggunakan rumus 2.1 dimana nilai  $p(ci)$  didapatkan dari jumlah tweet yang bertopik  $ci$  dibagi dengan keseluruhan tweet yang dijadikan sebagai data latih.

$$p(integritas) = 1/5$$

$$p(proker) = 1/5$$

$$p(perilaku) = 1/5$$

$$p(kinerja) = 1/5$$

$$p(kapabilitas) = 1/5$$

Dimana pada  $p(integritas)$  didapatkan nilai  $1/5$  yakni satu (1) dari jumlah tweet yang sudah bertopik integritas dan lima (5) dari jumlah keseluruhan tweet yang dijadikan data latih.

Nilai  $p(wkj.ci)$  merupakan nilai probabilitas awal dari masing-masing kata. Untuk mendapatkan nilai  $p(wkj.ci)$  menggunakan rumus 2.2 dimana frekuensi kemunculan kata akan ditambahkan satu (1), penambahan nilai satu (1) dilakukan agar hasil perhitungan tidak menjadi nol. Setelah didapatkan frekuensi kemunculan kata dan ditambahkan satu maka akan dibagi dengan jumlah pada masing-masing topik dan ditambahkan dengan semua kata yang ada.

Diketahui bahwa jumlah kemunculan kata janji pada topik integritas berjumlah 4, kata janji pada topik proker berjumlah 0, kata janji pada topik perilaku berjumlah 0, kata janji pada topik kinerja berjumlah 0 dan kata janji pada topik kapabilitas berjumlah 0. Jumlah keseluruhan term terdapat 23 term, sedangkan semua jumlah fitur kemunculan kata pada topik integritas terdapat 8, jumlah itu didapatkan dari kata janji sebanyak 4, wujud sebanyak 2, masyarakat sebanyak 2. Jumlah kemunculan kata pada masing-masing topik didapatkan dari tabel 3.7

$$p(janji, integritas) = \frac{4+1}{8+23} = \frac{5}{31}$$

$$p(\text{janji, proker}) = \frac{0+1}{6+23} = \frac{1}{29}$$

$$p(\text{janji, perilaku}) = \frac{0+1}{7+23} = \frac{1}{30}$$

$$p(\text{janji, kinerja}) = \frac{0+1}{10+23} = \frac{1}{32}$$

$$p(\text{janji, kapabilitas}) = \frac{0+1}{6+23} = \frac{1}{29}$$

**Tabel 3. 8** Bobot probabilitas naïve bayes untuk topik

Topik	$p(ci)$	$p(wkj.ci)$						
		Janji	Wujud	Masyarakat	Bangun	Rata	Bukti	Gagal
Integritas	1/5	5/31	3/31	3/31	1/31	1/31	1/31	1/31
Proker	1/5	1/29	1/29	1/29	3/29	2/29	2/29	2/29
Perilaku	1/5	1/30	1/30	1/30	1/30	1/30	1/30	1/30
Kinerja	1/5	1/32	4/32	1/32	1/32	1/32	1/32	1/32
Kapabilitas	1/5	1/29	1/29	1/29	1/29	1/29	1/29	1/29

Topik	$p(ci)$	$p(wkj.ci)$						
		Tinggal	Bahasa	Bicara	Pantas	Asal	Tingkah	Kerja
Integritas	1/5	1/31	1/31	1/31	1/31	1/31	1/31	1/31
Proker	1/5	2/29	1/29	1/29	1/29	1/29	1/29	1/29
Perilaku	1/5	1/30	2/30	2/30	3/30	2/30	2/30	1/30
Kinerja	1/5	1/32	1/32	1/32	1/32	1/32	1/32	3/32
Kapabilitas	1/5	1/29	1/29	1/29	1/29	1/29	1/29	1/29

Topik	$p(ci)$	$p(wkj.ci)$						
		Hasil	Wacana	Sukses	Pandai	Bodoh	Ragu	Lulus
Integritas	1/5	1/31	1/31	1/31	1/31	1/31	1/31	1/31
Proker	1/5	1/29	1/29	1/29	1/29	1/29	1/29	1/29
Perilaku	1/5	1/30	1/30	1/30	1/30	1/30	1/30	1/30
Kinerja	1/5	2/32	2/32	2/32	2/32	1/32	1/32	1/32
Kapabilitas	1/5	1/29	1/29	1/29	1/29	2/29	2/29	2/29

Topik	$p(ci)$	$p(wkj.ci)$	
		Sekolah	Sejahtera
Integritas	1/5	1/31	1/31
Proker	1/5	1/29	1/29
Perilaku	1/5	1/30	1/30
Kinerja	1/5	1/32	1/32
Kapabilitas	1/5	2/29	1/29

$$\begin{aligned}
p(\text{"integritas"}|T6) &= p(\text{"integritas"}) \times p(\text{"sukses"}|\text{"integritas"}) + \\
& p(\text{"bangun"}|\text{"integritas"}) + p(\text{"sejahtera"}|\text{"integritas"}) + \\
& p(\text{"hasil"}|\text{"integritas"}) \\
&= 1/5 \times (1/31 + 1/31 + 1/31 + 1/31) = 0.0000000216562
\end{aligned}$$

$$\begin{aligned}
p(\text{"proker"}|T6) &= p(\text{"proker"}) \times p(\text{"sukses"}|\text{"proker"}) + \\
& p(\text{"bangun"}|\text{"proker"}) + p(\text{"sejahtera"}|\text{"proker"}) + p(\text{"hasil"}|\text{"proker"}) \\
&= 1/5 \times (1/29 + 3/29 + 1/29 + 1/29) = 0.0000000848319
\end{aligned}$$

$$\begin{aligned}
p(\text{"perilaku"}|T6) &= p(\text{"Perilaku"}) \times p(\text{"sukses"}|\text{"Perilaku"}) + \\
& p(\text{"bangun"}|\text{"Perilaku"}) + p(\text{"sejahtera"}|\text{"Perilaku"}) + p(\text{"hasil"}|\text{"Perilaku"}) \\
&= 1/5 \times (1/30 + 1/30 + 1/30 + 1/30) = 0.0000000246914
\end{aligned}$$

$$\begin{aligned}
p(\text{"kinerja"}|T6) &= p(\text{"kinerja"}) \times p(\text{"sukses"}|\text{"kinerja"}) + \\
& p(\text{"bangun"}|\text{"kinerja"}) + p(\text{"sejahtera"}|\text{"kinerja"}) + p(\text{"hasil"}|\text{"kinerja"}) \\
&= 1/5 \times (2/32 + 1/32 + 1/32 + 3/32) = 0.000000114441
\end{aligned}$$

$$\begin{aligned}
p(\text{"kapabilitas"}|T6) &= p(\text{"kapabilitas"}) \times p(\text{"sukses"}|\text{"kapabilitas"}) + \\
& p(\text{"bangun"}|\text{"kapabilitas"}) + p(\text{"sejahtera"}|\text{"kapabilitas"}) + \\
& p(\text{"hasil"}|\text{"kapabilitas"}) \\
&= 1/5 \times (1/29 + 1/29 + 1/29 + 1/29) = 0.0000000282773
\end{aligned}$$

Dengan melakukan perhitungan probabilistik teks terhadap topik dengan menggunakan metode *naïve bayes* maka dapat diketahui bahwa teks *twitter* T6 termasuk dalam kategori topik Kinerja karena nilai bobot probabilistik teks *twitter* T6 terhadap topik Kinerja lebih besar dari pada nilai bobot teks T6 terhadap topik lainnya.

#### 3.3.1.7.4 Probabilitas Sentimen

Probabilitas sentimen merupakan proses penentuan teks *twitter* yang dihitung untuk didapatkan nilai tertinggi dari beberapa kategori dalam sentimen. Keenam 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 :janji(4), wujud(2), masyarakat(2), kerja(2), wujud(3), hasil(2), wacana(2), sukses(1)

Negatif :bangun(2), rata(1), bukti(1), gagal(1), tinggal(1)

Netral :bahasa(1), bicara(2), pantas(2), asal(1), tingkah(1), pandai(2), bodoh(1), ragu(1), lulus(1), sekolah(1)

? :sukses(2), bangun(1), sejahtera(1), hasil(2)

Seperti halnya untuk menghitung *tweet* terhadap topik. Perhitungan probabilitas *tweet* terhadap sentimen juga menggunakan rumus 2.1 dan rumus 2.2. dimana untuk menentukan probabilitas sentimen memerlukan nilai  $p(ci)$  dan  $p(wkj.ci)$ . Nilai  $p(ci)$  untuk probabilitas awal sentimen didapatkan dari jumlah *tweet* yang tergabung dalam topik  $ci$  dibagi dengan semua topik. Dari tabel 3.1 yang merupakan *tweet* awal, didapatkan bahwa ada lima (5) jumlah keseluruhan *tweet* data latih yang terdiri dari dua untuk sentimen positif dan negatif serta satu untuk sentimen netral.

$$p(\text{positif}) = 2/5$$

$$p(\text{netral}) = 1/5$$

$$p(\text{negatif}) = 2/5$$

Nilai  $p(wkj.ci)$  pada pembobotan sentimen dilakukan dengan rumus 2.2 dimana fitur kemunculan term ditambahkan satu lalu dibagi dengan hasil penjumlahan antara semua jumlah term pada topik  $ci$  ditambah dengan keseluruhan term. Diketahui bahwa pada sentimen positif terdapat 17 kali fitur kemunculan kata pada sentimen positif, 7 pada sentimen netral, 12 pada sentimen negatif dan jumlah keseluruhan term ada 23.

$$p(\text{janji, positif}) = \frac{6+1}{17+23} = \frac{5}{40}$$

$$p(\text{janji, netral}) = \frac{0+1}{7+23} = \frac{1}{30}$$

$$p(\text{janji, negatif}) = \frac{0+1}{12+23} = \frac{5}{35}$$

**Tabel 3. 9** Bobot probabilitas naïve bayes untuk sentimen

Sentimen	$p(ci)$	$p(wkj.ci)$						
		Janji	Wujud	Masyarakat	Bangun	Rata	Bukti	Gagal
Positif	2/5	5/40	6/40	3/40	1/40	1/40	1/40	1/40
Netral	1/5	1/30	1/30	1/30	1/35	1/30	1/30	1/30
Negatif	2/5	1/35	1/35	1/35	3/35	2/35	2/35	2/35

Sentimen	$p(ci)$	$p(wkj.ci)$						
		Tinggal	Bahasa	Bicara	Pantas	Asal	Tingkah	Kerja
Positif	2/5	1/40	1/40	1/40	1/40	1/40	1/40	3/40
Netral	1/5	1/30	2/30	3/30	3/30	2/30	2/30	1/30
Negatif	2/5	2/35	1/35	1/35	1/35	1/35	1/35	1/35

Sentimen	$p(ci)$	$p(wkj.ci)$						
		Hasil	Wacana	Sukses	Pandai	Bodoh	Ragu	Lulus
Positif	2/5	3/40	2/40	2/40	1/40	1/40	1/40	1/40
Netral	1/5	1/30	1/30	1/30	1/30	1/30	1/30	1/30
Negatif	2/5	1/35	1/35	1/35	3/35	2/35	2/35	2/35

Sentimen	$p(ci)$	$p(wkj.ci)$	
		Sekolah	Sejahtera
Positif	2/5	1/40	1/40
Netral	1/5	1/30	1/30
Negatif	2/5	2/35	1/35

$$\begin{aligned}
 p(\text{"positif"}|\text{"T6"}) &= p(\text{"positif"}) \times p(\text{"sukses"}|\text{"positif"}) \times p(\text{"bangun"}|\text{"positif"}) \\
 &\times p(\text{"sejahtera"}|\text{"positif"}) \times p(\text{"hasil"}|\text{"positif"}) \\
 &= 2/5 \times (2/40 + 1/40 + 1/40 + 3/40) = 0.00000125
 \end{aligned}$$

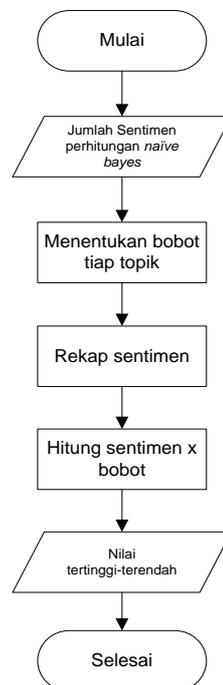
$$\begin{aligned}
 p(\text{"netral"}|\text{"T6"}) &= p(\text{"netral"}) \times p(\text{"sukses"}|\text{"netral"}) \times p(\text{"bangun"}|\text{"netral"}) \\
 &\times p(\text{"sejahtera"}|\text{"netral"}) \times p(\text{"hasil"}|\text{"netral"}) \\
 &= 1/5 \times (1/30 + 1/30 + 1/30 + 1/30) = 0.000000246
 \end{aligned}$$

$$\begin{aligned}
 p(\text{"negatif"}|\text{"T6"}) &= p(\text{"negatif"}) \times p(\text{"sukses"}|\text{"negatif"}) \times \\
 &p(\text{"bangun"}|\text{"negatif"}) \times p(\text{"sejahtera"}|\text{"negatif"}) \times p(\text{"hasil"}|\text{"negatif"}) \\
 &= 2/5 \times (1/35 + 3/35 + 1/35 + 1/35) = 0.0000000399
 \end{aligned}$$

Dengan melakukan perhitungan probabilistik teks terhadap sentimen dengan menggunakan metode *naïve bayes* maka dapat diketahui bahwa teks *twitter* T6 termasuk dalam kategori sentimen positif karena nilai bobot probabilistik teks *twitter* T6 terhadap topik positif lebih besar dari pada nilai bobot teks T6 terhadap topik lainnya.

### 3.3.1.7.5 Perhitungan *Simple Additive Weighting*

Pada penelitian ini penggunaan metode *simple additive weighting* digunakan untuk menghitung nilai sentimen yang ada pada topik setiap tokoh politik sehingga didapatkan urutan-urutan nama tokoh politik yang mendapatkan nilai sentimen terbaik sampai dengan sentimen terburuk. Alur proses perhitungan dengan metode *simple additive weighting* seperti pada gambar 3.8



**Gambar 3. 8** alur proses perhitungan metode *simple additive weighting*

Awal dari proses perhitungan metode *simple additive weighting* adalah menentukan nilai bobot pada masing-masing topik. Penentuan nilai bobot topik

dilakukan untuk mendapatkan bobot prioritas pada topik. Bobot prioritas dihitung dengan rumus (2.5)

Contoh kasus apabila diketahui jumlah tweet yang tergabung dalam topik integritas terdapat 20 *tweet*, topik proker terdapat 22 *tweet*, topik perilaku terdapat 48 *tweet*, topik kinerja terdapat 40 *tweet*, topik kapabilitas ada 20 *tweet* dan jumlah keseluruhan *tweet* yakni 130.

$$\text{Bobot prioritas topik (Integritas)} = 20 / 130 = 0.15$$

$$\text{Bobot prioritas topik (Proker)} = 22 / 130 = 0.17$$

$$\text{Bobot prioritas topik (Perilaku)} = 48 / 130 = 0.22$$

$$\text{Bobot prioritas topik (Kinerja)} = 40 / 130 = 0.31$$

$$\text{Bobot prioritas topik (Kapabilitas)} = 20 / 130 = 0.15$$

Detail hasil perhitungan bobot prioritas untuk perhitungan diatas seperti pada tabel 3.10

**Tabel 3. 10** Bobot prioritas tiap topik

Topik	Jumlah	Bobot
Integritas	20	0.15
Proker	22	0.17
Perilaku	48	0.22
Kinerja	40	0.31
Kapabilitas	20	0.15
JUMLAH	130	1.00

Nilai sentimen merupakan nilai yang didapatkan dari rekap jumlah sentimen dari tokoh politik. Sentimen yang didapatkan dikelompokan berdasarkan topik dan sentimen-nya (positif, netral dan negatif). Dimana dari tabel 3.1 terdapat lima (5) jumlah *tweet* yang teridentifikasi sebagai topik proker untuk tokoh politik Joko Widodo, dari kelima *tweet* tersebut tiga (3) diantaranya termasuk sentimen positif dan terdapat dua (2) yang termasuk ke sentimen negatif. Dari jumlah *tweet* joko widodo untuk topik proker dan memiliki sentimen maka akan di hitung untuk menentukan tokoh politik dengan sentimen terbaik. Dari tabel 3.1 terbentuk data pada tabel 3.11 untuk perhitungan metode *simple additive weighting*.

**Tabel 3. 11** Rekap jumlah sentimen tokoh politik

Topik	Joko Widodo			Prabowo		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Integritas	6	0	2	1	0	2
Proker	3	0	2	2	0	1
Perilaku	3	0	1	2	0	3
Kinerja	8	0	2	6	0	3
Kapabilitas	2	0	1	3	0	1

Topik	Zulkifi Hasan			Gatot Nurmantiyo		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Integritas	0	0	1	1	0	1
Proker	2	0	2	3	0	0
Perilaku	5	0	2	4	0	0
Kinerja	3	0	4	3	0	1
Kapabilitas	2	0	1	3	0	0

Topik	AHY			Susi Pudjiastuti		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Integritas	1	0	1	3	0	1
Proker	2	0	1	2	0	1
Perilaku	1	0	1	2	0	3
Kinerja	4	0	2	3	0	0
Kapabilitas	2	0	0	2	0	0

Topik	Zainul Majid (TGB)		
	Positif	Netral	Negatif
Integritas	1	0	1
Proker	2	0	1
Perilaku	1	0	0
Kinerja	1	0	0
Kapabilitas	1	0	2

Setelah didapatkan rekap sentimen pada masing-masing tokoh politik, maka tahap selanjutnya adalah menghitung nilai sentimen tiap topik pada masing-masing tokoh politik dengan melakukan kalkulasi pada tiap nilai sentimen.

Kalkulasi dilakukan dengan cara mengurangi jumlah sentimen positif dengan nilai negatif. Nilai netral tidak akan diikutkan dalam perhitungan karena netral tidak memiliki kecenderungan nilai sentimen tersebut baik atau buruk.

$$Joko\ widodo(integritas) = 6 - 2 = 4$$

$$Joko\ widodo(proker) = 3 - 2 = 1$$

$$Joko\ widodo(perilaku) = 3 - 0 = 3$$

$$Joko\ widodo(kinerja) = 8 - 2 = 0$$

$$Joko\ widodo(kapabilitas) = 2 - 1 = 0$$

**Tabel 3. 12** rekapan nilai sentimen tokoh politik

Topik	Joko Widodo	Prabowo	Zulkifli Hasan	Gatot	AHY	Susi P	TGB
Integritas	4	-1	-1	1	1	2	0
Proker	1	0	0	1	1	1	1
Perilaku	2	3	3	4	0	-1	1
Kinerja	6	-1	-1	2	2	3	1
Kapabilitas	1	1	1	3	2	2	-1

Nilai sentimen seperti pada tabel 3.12 akan digunakan dalam perhitungan metode *simple additive weighting* untuk mendapatkan tokoh politik dengan sentimen terbaik. Sebelum proses dilanjutkan, dengan adanya ada yang minus maka perlu dilakukan normalisasi data terlebih dahulu. Normalisasi dilakukan menggunakan normalisasi linier jangkauan [0, 1] dengan menggunakan rumus 2.4. Perhitungan normalisasi data dengan normalisasi linier [0, 1] seperti berikut :

$$x(joko\ widodo, integritas) = (4 - (-1)) / (4 - (-1)) = 1$$

$$x(joko\ widodo, proker) = (1 - 0) / (1 - 1) = 1$$

$$x(joko\ widodo, Perilaku) = (2 - (-1)) / (4 - (-1)) = 0.6$$

$$x(joko\ widodo, kinerja) = (6 - (-1)) / (6 - (-1)) = 1$$

$$x(joko\ widodo, kapabilitas) = (1 - (-1)) / (3 - (-1)) = 0.5$$

$$x(prabowo, integritas) = ((-1) - (-1)) / (4 - (-1)) = 0$$

$$x(prabowo, proker) = (1 - 0) / (1 - 1) = 1$$

$$x(prabowo, Perilaku) = ((-1) - (-1)) / (4 - (-1)) = 0$$

$$x(\text{prabowo, kinerja}) = (1 - (-1)) / (6 - (-1)) = 0.5714286$$

$$x(\text{prabowo, kapabilitas}) = (1 - (-1)) / (3 - (-1)) = 0.75$$

$$x(\text{zulkifli, integritas}) = ((-1) - (-1)) / (4 - (-1)) = 0$$

$$x(\text{zulkifli, proker}) = (0 - 0) / (1 - 1) = 0$$

$$x(\text{zulkifli, Perilaku}) = (3 - (-1)) / (4 - (-1)) = 0.8$$

$$x(\text{zulkifli, kinerja}) = ((-1) - (-1)) / (6 - (-1)) = 0$$

$$x(\text{zulkifli, kapabilitas}) = (1 - (-1)) / (3 - (-1)) = 0.5$$

$$x(\text{gatot, integritas}) = (1 - (-1)) / (4 - (-1)) = 4$$

$$x(\text{gatot, proker}) = (1 - 0) / (1 - 1) = 1$$

$$x(\text{gatot, Perilaku}) = (4 - (-1)) / (4 - (-1)) = 1$$

$$x(\text{gatot, kinerja}) = (2 - (-1)) / (6 - (-1)) = 0.42857$$

$$x(\text{gatot, kapabilitas}) = (3 - (-1)) / (3 - (-1)) = 1$$

$$x(\text{ahy, integritas}) = (1 - (-1)) / (4 - (-1)) = 0$$

$$x(\text{ahy, proker}) = (1 - 0) / (1 - 1) = 0$$

$$x(\text{ahy, Perilaku}) = (0 - (-1)) / (4 - (-1)) = 0.8$$

$$x(\text{ahy, kinerja}) = (2 - (-1)) / (6 - (-1)) = 0$$

$$x(\text{ahy, kapabilitas}) = (2 - (-1)) / (3 - (-1)) = 0.5$$

$$x(\text{susi, integritas}) = (2 - (-1)) / (4 - (-1)) = 0.6$$

$$x(\text{susi, proker}) = (1 - 0) / (1 - 1) = 1$$

$$x(\text{susi, Perilaku}) = ((-1) - (-1)) / (4 - (-1)) = 0$$

$$x(\text{susi, kinerja}) = (3 - (-1)) / (6 - (-1)) = 0.57143$$

$$x(\text{susi, kapabilitas}) = (2 - (-1)) / (3 - (-1)) = 0.75$$

$$x(\text{zainul, integritas}) = (0 - (-1)) / (4 - (-1)) = 0.2$$

$$x(\text{zainul, proker}) = (1 - 0) / (1 - 1) = 1$$

$$x(\text{zainul, Perilaku}) = (1 - (-1)) / (4 - (-1)) = 0.4$$

$$x(\text{zainul, kinerja}) = (1 - (-1)) / (6 - (-1)) = 0.28571$$

$$z(zainul, kapabilitas) = ((-1) - (-1)) / (3 - (-1)) = 0$$

Dari proses normalisasi maka didapatkan nilai sentimen pada masing - masing tokoh politik seperti pada tabel 3.23

**Tabel 3. 13** Hasil normalisasi liner [0,1]

Topik	Joko Widodo	Prabowo	Zulkifli Hasan	Gatot	AHY	Susi P	TGB
Integritas	1	0	0	0.4	0.4	0.6	0.2
Proker	1	1	0	1	1	1	1
Perilaku	0.6	0	0.8	1	0.2	0	0.4
Kinerja	1	0.57	0	0.43	0.43	0.57	0.28
Kapabilitas	0.5	0.75	0.5	1	0.75	0.75	0

Setelah data nilai sentimen dilakukan normalisasi, maka tahap terakhir pada perhitungan metode *simple additive weighting* yakni menentukan tokoh politik yang memiliki nilai tertinggi. Penentuan nilai sentimen tertinggi akan menggunakan metode *simple additive weighting* dimana nilai sentimen yang telah dinormalisasi dikali dengan bobot awal pada masing - masing topik.

*Joko widodo*

$$= (1 \times 0.15) + (1 \times 0.17) + (0.6 \times 0.22) + (1 \times 0.31) + (0.5 \times 0.15) = 0.83692$$

*Prabowo*

$$= (0 \times 0.15) + (1 \times 0.17) + (0 \times 0.22) + (0.57 \times 0.31) + (0.75 \times 0.15) = 0.46044$$

*Zulkifli hasan*

$$= (0 \times 0.15) + (0 \times 0.17) + (0.8 \times 0.22) + (0 \times 0.31) + (0.5 \times 0.15) = 0.24923$$

*Gatot nurmantiyo*

$$= (0.4 \times 0.15) + (1 \times 0.17) + (1 \times 0.22) + (0.43 \times 0.31) + (1 \times 0.15) = 0.73187$$

*AHY*

$$= (0.4 \times 0.15) + (1 \times 0.17) + (1 \times 0.22) + (0.43 \times 0.31) + (1 \times 0.15) = 0.5211$$

*Susi pudjiastuti*

$$= (0.6 \times 0.15) + (1 \times 0.17) + (0 \times 0.22) + (0.57 \times 0.31) + (0.75 \times 0.15) = 0.55275$$

*TGB*

$$= (0.2 \times 0.15) + (1 \times 0.17) + (0.4 \times 0.22) + (0.28 \times 0.31) + (0 \times 0.15) = 0.37407$$

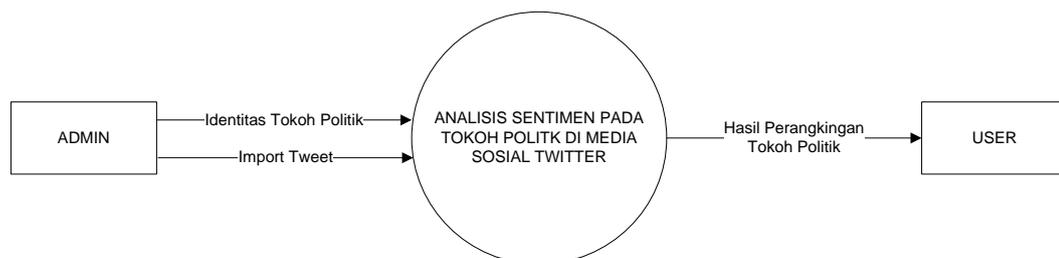
**Tabel 3. 14** Hasil akhir perangkingan tokoh politik

Tokoh	Nilai Akhir	Rangking
Joko Widodo	0.83692	1
Prabowo	0.46044	5
Zulkifli Hasan	0.24923	7
Gatot Nurmantiyo	0.73187	2
AHY	0.5211	4
Susi Pudjiastuti	0.55275	3
Zainul Majid (TGB)	0.37407	6

Dari proses perhitungan nilai sentimen dengan bobot topik diatas, didapatkan nilai akhir sentimen pada masing – masing tokoh politik dengan nilai tertinggi yakni 0.83692 (Joko Widodo), 0.73187 (Gatot Nurmantiyo), 0.55275 (Susi Pudjiastuti) dan nilai terendah yakni 0.24923 (Zulkifli Hasam). Maka kesimpulan didapatkan bahwa Joko Widodo memiliki nilai elektabilitas tertinggi, disusul oleh Gatot Nurmantiyo dan Susi Pudjiastuti.

### 3.3.2 Diagram Konteks

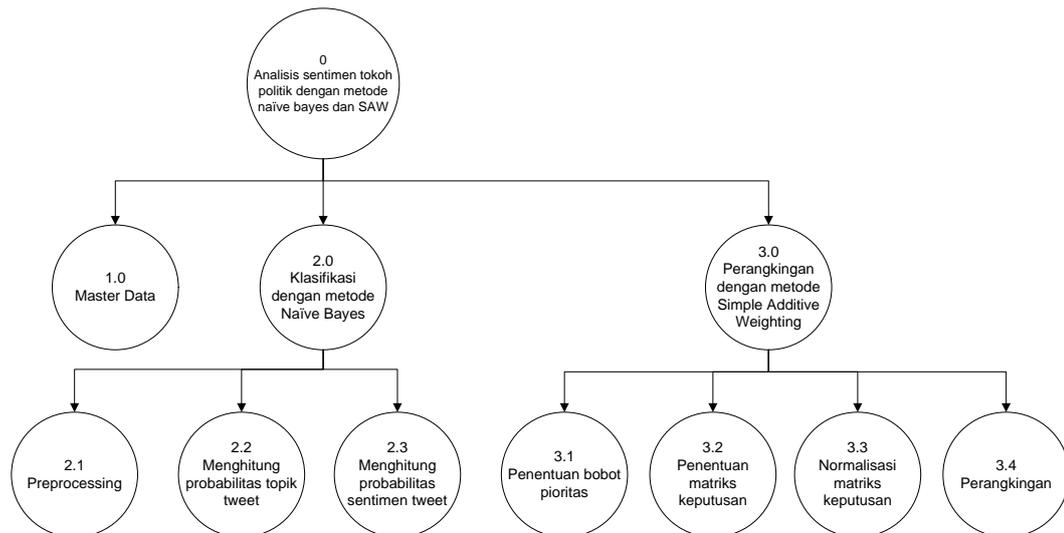
Diagram konteks merupakan diagram yang menunjukkan sebuah proses tunggal dalam sistem yang berhubungan dengan bagian yang terkait. Rangkaian diagram konteks yang digunakan pada penelitian ini seperti pada gambar 3.9

**Gambar 3. 9** Diagram Konteks Analisis Sentimen pada Tokoh Politik

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

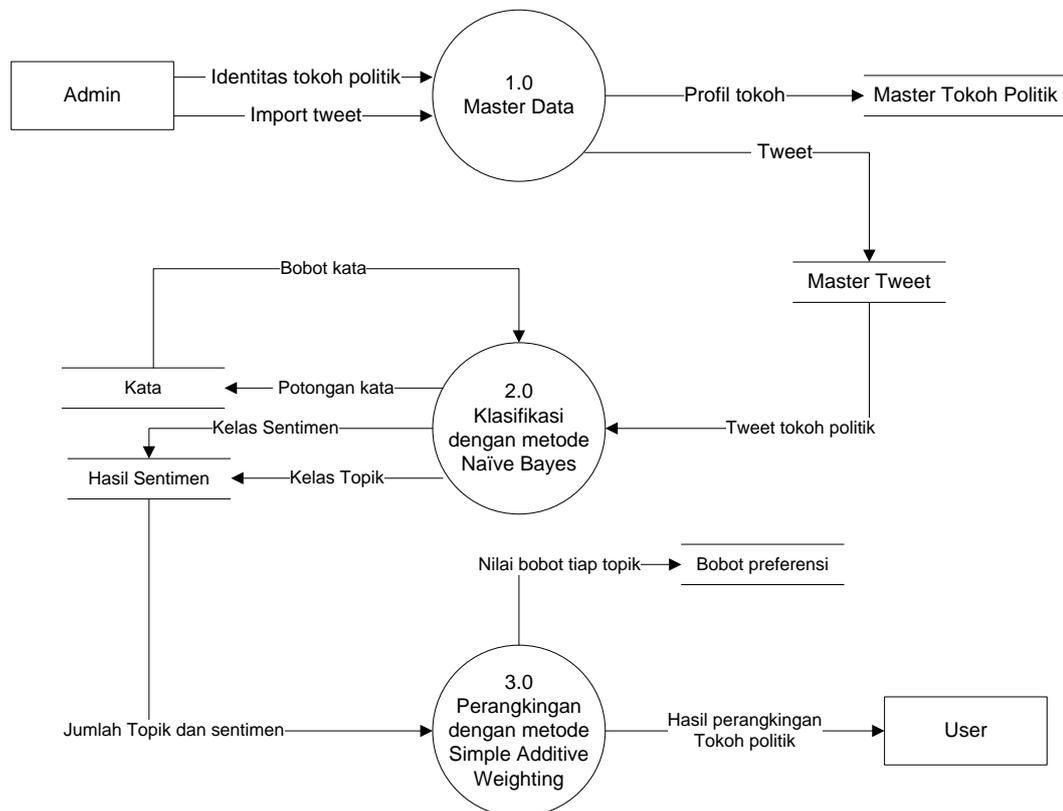
Berikut penjelasan gambar 3.10 berdasarkan kerangka diagram berjenjang diatas terlihat bahwa sistem yang dibuat terdiri dari tiga level

1. Top level : Analisis sentimen tokoh politik dengan metode *Naïve Bayes* dan *SAW*.
2. Level 0 : merupakan hasil *break down* dari proses keseluruhan dari analisis sentimen pada tokoh politik menjadi beberapa sub proses yaitu :
  - a. Master data.
  - b. Klasifikasi dengan metode *Naïve Bayes*.
  - c. Perangkingan dengan metode *Simple Additive Weighting*.
3. Level 1 : merupakan sub proses dari beberapa proses pada level 0 dalam analisis sentimen pada tokoh politik 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 *simple additive weighting* :
  - a. Penentuan bobot prioritas.
  - 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 tokoh politik

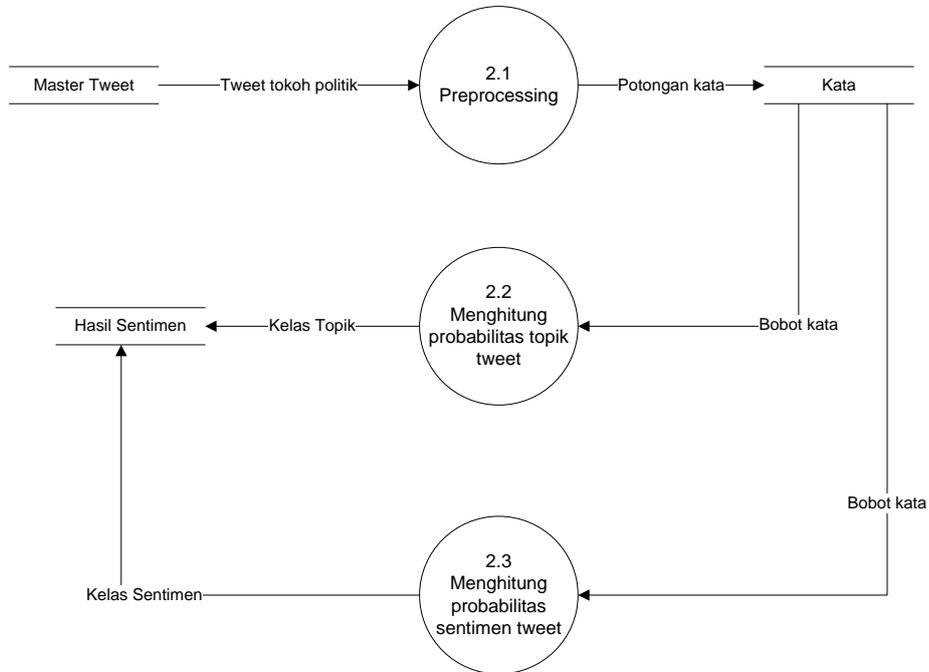
Berdasarkan pada gambar 3.11, dapat dijelaskan bahwa DFD level 0 menjelaskan beberapa proses yang terjadi. Beberapa proses yang ada pada DFD level 0 antara lain :

- a. Master data
- b. Klasifikasi dengan metode Naïve Bayes

c. Perangkingan dengan metode simple additive weighting

### 3.3.4.2 Data Flow Diagram level 1

#### 1) DFD Level 1 Proses 2

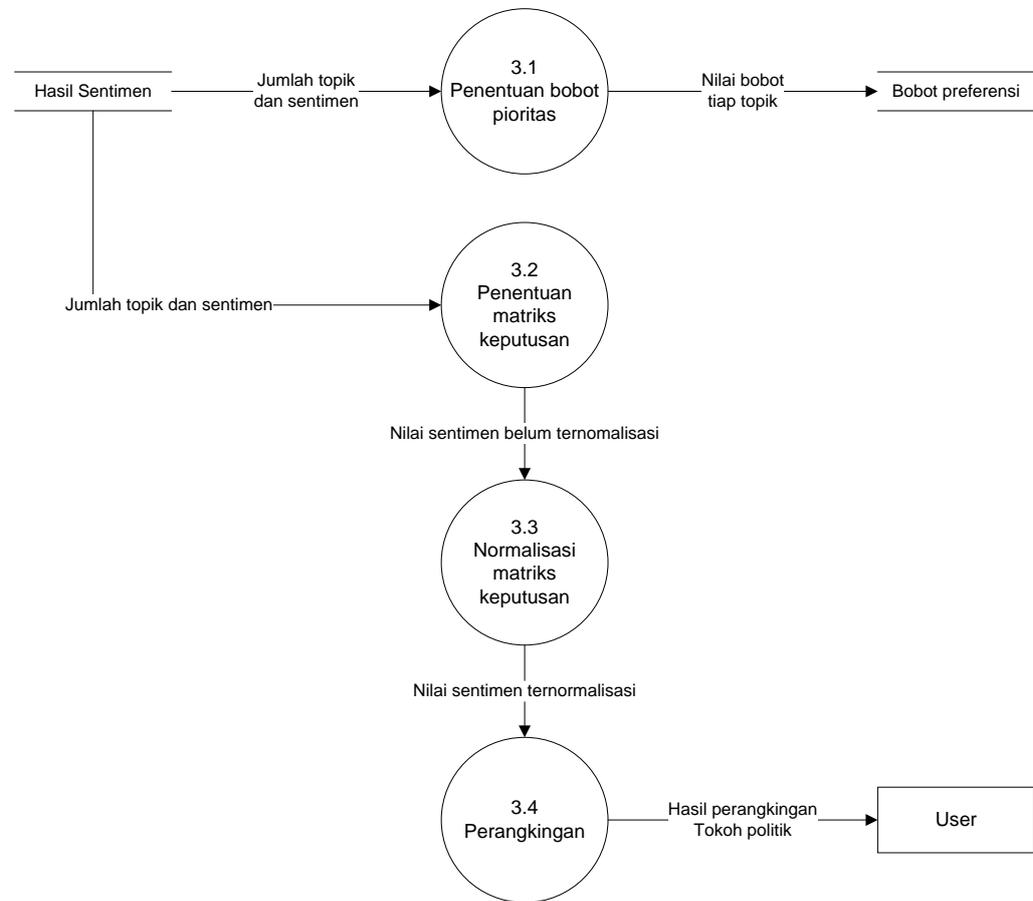


**Gambar 3. 12** Data Flow Diagram level 1 untuk proses kedua analisis sentimen pada tokoh politik

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 tokoh politik

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

- a. Penentuan bobot prioritas.
- 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 tokoh politik 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 sentiment. Atribut yang ada pada tabel master *tweet* adalah *id\_tweet*, *isi\_tweet* dan *id\_tokoh\_politik*. Skema tabel master *tweet* seperti pada tabel 3.15

**Tabel 3. 15** 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_tokoh_politik</i>	Int	15	<i>Foreign key</i>

#### 2. Tabel Master Tokoh Politik

Tabel master tokoh politik digunakan sebagai master yang berisikan para tokoh-tokoh yang akan diketahui hasil analisis sentimen-nya. Atribut yang ada pada tabel master tokoh politik adalah *id\_tokoh\_politik* dan *nama\_tokoh*. Skema tabel master tokoh politik seperti pada tabel 3.16

**Tabel 3. 16** tabel master tokoh politik

Nama kolom	Tipe	Ukuran	Keterangan
<i>Id_tokoh_politik</i>	Int	15	<i>Primary key</i>
<i>Nama_tokoh</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 3.17

**Tabel 3. 17** 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 3.18

**Tabel 3. 18** 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 3.19

**Tabel 3. 19** 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

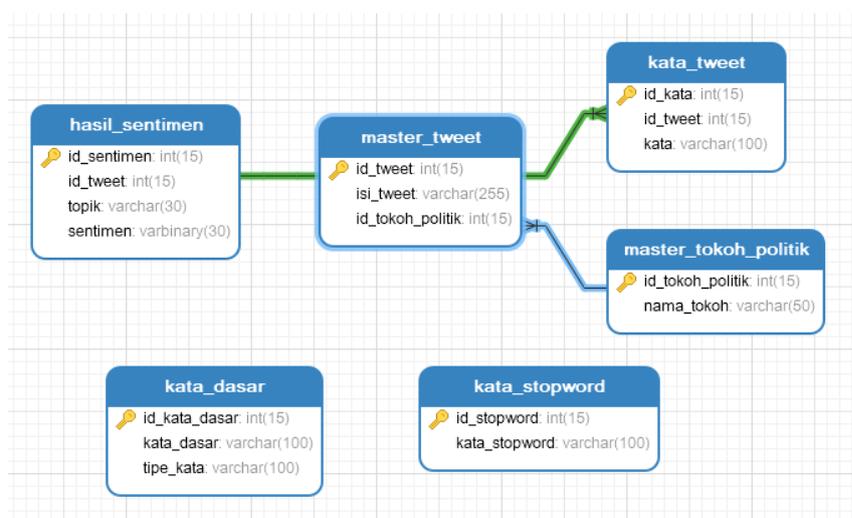
Tabel sentimen digunakan sebagai penyimpanan hasil pemrosesan analisis sentimen mulai dari pembobotan *tweet* terhadap topik maupun *tweet* terhadap sentimen. Atribut yang ada pada tabel sentimen adalah *id\_sentimen*, *id\_tweet*, topik dan sentimen. Skema tabel sentimen seperti pada gambar 3.20

**Tabel 3. 20** 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 tokoh politik seperti pada gambar 3.14



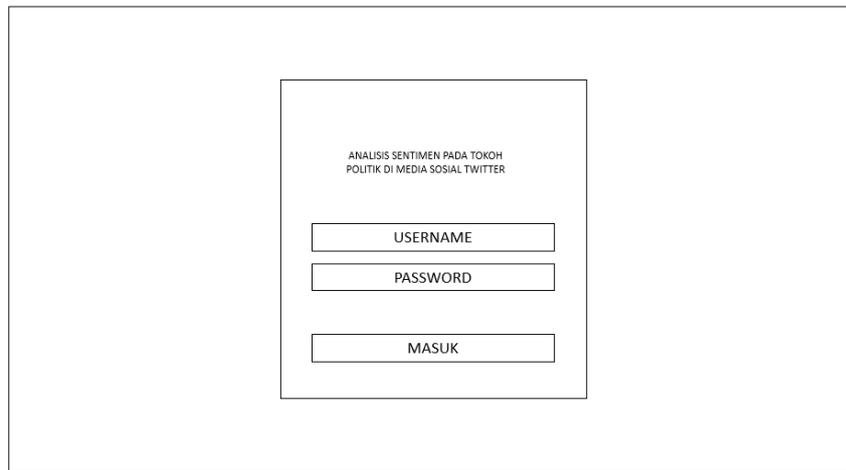
**Gambar 3. 14** Entity relationship diagram dalam analisis sentimen tokoh politik

### 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 tokoh politik.

#### 3.5.1 Halaman *Login*

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

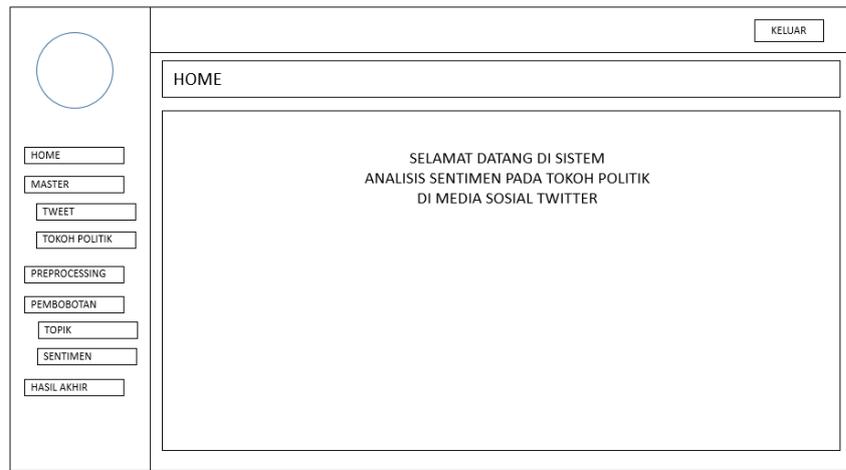


The image shows a wireframe of a login page. At the top center, the text reads "ANALISIS SENTIMEN PADA TOKOH POLITIK DI MEDIA SOSIAL TWITTER". Below this text are three rectangular input fields stacked vertically. The first field is labeled "USERNAME", the second is labeled "PASSWORD", and the third is labeled "MASUK". The entire login form is centered within a larger rectangular frame.

**Gambar 3.15** Tampilan rancangan halaman *login*

#### 3.5.2 Halaman *Home*

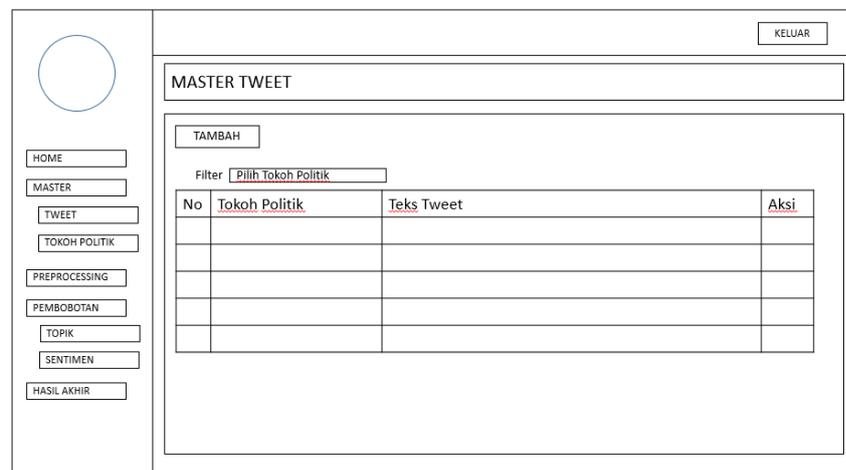
Halaman *home* merupakan halaman pertama yang muncul ketika pengguna berhasil *login*. Halaman ini berisikan informasi mengenai nama sistem dan kegunaannya. Tampilan rancangan halaman home seperti pada gambar 3.16



**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, tokoh politik dan aksi untuk menghapus ataupun mengubah teks. Tampilan rancangan halaman master tweet seperti pada gambar 3.17



**Gambar 3. 17** Tampilan rancangan halaman master tweet

### 3.5.4 Halaman *Master Tokoh Politik*

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

No	Tokoh Politik	Aksi

**Gambar 3. 18** Tampilan rancangan halaman master tokoh politik

### 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 teks tweet dan tokoh politik 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 Tokoh Politik

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

**Gambar 3. 20** Tampilan rancangan halaman tambah tokoh politik

### 3.5.7 Halaman *Preprocessing*

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

No	Tweet	Aksi

No	Term

**Gambar 3. 21** Tampilan rancangan halaman preprocessing

### 3.5.8 Halaman Pembobotan Topik

Halaman pembobotan topik akan menampilkan hasil proses perhitungan pembobotan tweet to topic. 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

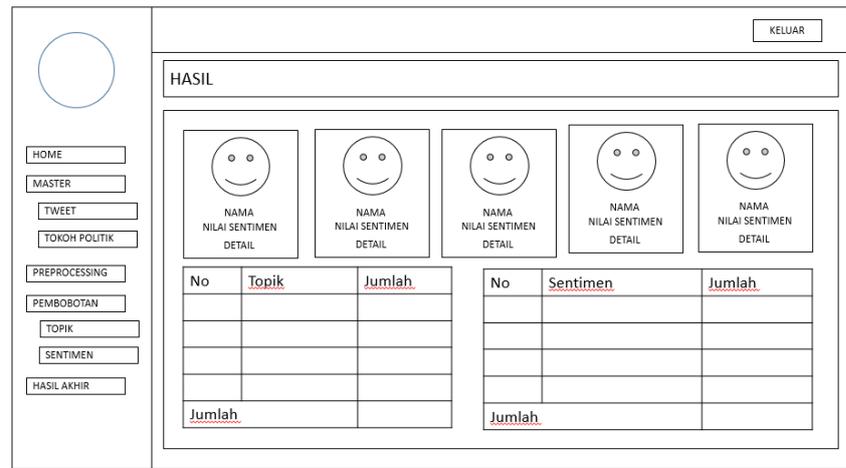
### 3.5.9 Halaman Pembobotan Sentimen

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 tokoh-nya serta menampilkan detail perhitungan dengan metode *simple additive weighting*. 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*. 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 500 data uji akan menentukan topik dan sentimen dari ke 500 data uji tersebut. Penentuan topik dan sentimen menggunakan rumus (2.1, 2.2, 2.3) yakni *naïve bayes classifier* serta menggunakan lebih dari 1000 kata yang termasuk dalam korpus dalam membantu penentuan topik dan sentimen, setiap data uji akan dilakukan perhitungan dengan menghitung masing-masing bobot pada topik 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 (2.4). Jika dalam pengembangannya terdapat penambahan master data tweet baru maka akan dilakukan kembali penentuan topik dan sentimen dengan rumus (2.1, 2.2, 2.3) dan dilakukan kembali penghitungan akurasi dengan rumus (2.4)

### 3.7 Spesifikasi Pembuatan Sistem

Pembuatan sistem untuk proses analisis sentimen pada tokoh politik 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