

## **BAB III**

### **ANALISIS DAN DESAIN SISTEM**

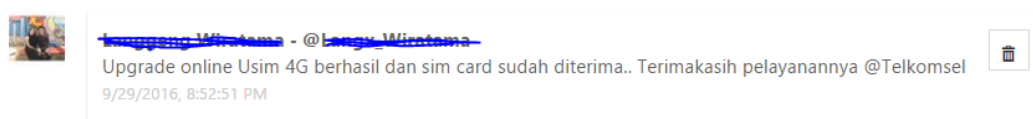
#### **III.1. Analisis Masalah**

Twitter adalah salah satu *microblogging* yang sangat populer di tengah masyarakat. Hal ini terlihat dari banyaknya jumlah pengguna Twitter yang mencapai 231,7 juta. Biasanya Twitter digunakan sebagai sarana untuk menyampaikan pendapat. Informasi yang terkandung dalam tweets ini sangat berharga sebagai alat penentu kebijakan. Salah satunya adalah untuk menilai suatu produk yang dikeluarkan oleh perusahaan atau yang sering disebut dengan analisis sentimen.

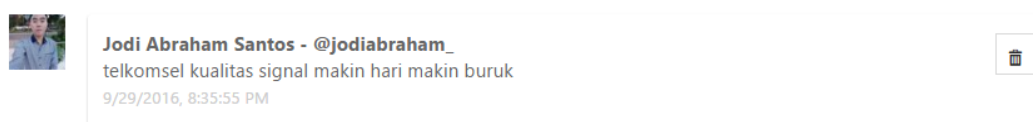
Perusahaan dapat memanfaatkan salah satu *microblogging* ini sebagai layanan feedback untuk pengguna produknya. Salah satu perusahaan yang menyediakan layanan feedback adalah Telkomsel. Feedback dari pengguna biasanya berupa opini-opini selama menggunakan produk dari perusahaan tersebut. Hal ini sangat berguna bagi perusahaan dalam meninjau kembali produknya. Opini dari pengguna itu nantinya akan dilakukan analisis sentimen apakah termasuk opini positif atau opini negatif. Namun, analisis sentimen ini mendapat tantangan berupa model bahasa tidak formal yang digunakan di Twitter. Maka dari itu, sebelum melakukan analisis sentimen, terlebih dahulu harus dilakukan preprocessing pada data tweets yang akan digunakan. Hal ini berguna untuk mengatasi model bahasa yang tidak formal yang sering digunakan pada Twitter.

### III.1.1. Evaluasi Sistem

Data yang digunakan pada penelitian ini diambil dari kumpulan tweets bahasa Indonesia yang diambil dari official akun Twitter Telkomsel. Data tweets ini diperoleh dengan membuat program crawling yang menggunakan Tweetinvi API. Dalam proses crawling, secara otomatis akan mengambil data tweets yang mengandung kata “Telkomsel”, “#Telkomsel”, dan “@Telkomsel”. Data tweets yang terkumpul nantinya akan melewati tahap preprocessing dan selanjutnya akan diklasifikasikan. Dalam sistem analisis sentimen ini, tweets akan diklasifikasikan ke dalam dua kelas (kategori), yaitu kelas sentimen positif dan kelas sentimen negatif



**Gambar III.1. Contoh Sentimen Positif**



**Gambar III.2. Contoh Sentimen Negatif**

### III.1.2. Strategi Sistem

Strategi sistem untuk melakukan pemrosesan teks merupakan proses menggali, mengolah, mengatur informasi dengan cara menganalisis hubungannya, aturan-aturan yang ada di data tekstual semi terstruktur atau tidak terstruktur. Untuk lebih efektif dalam proses pemrosesan dilakukan langkah transformasi data ke dalam suatu format yang memudahkan untuk kebutuhan pemakai.

Preprocessing merupakan salah satu langkah yang penting dalam analisis sentimen.

### III.2. Analisis Penerapan Metode Naive Bayes

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Ebranda A.W ; 2013 : 2-3).

Langkah Langkah Navie Bayes ini yaitu:

1. Analisis masalah.
2. Analisis sumber data.
3. Analisis preprocessing.
4. Analisis penerapan algoritma naïve bayes.
5. Analisis pembobotan tweet.

Persamaan Metode Naive Bayes Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots (1)$$

Di mana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

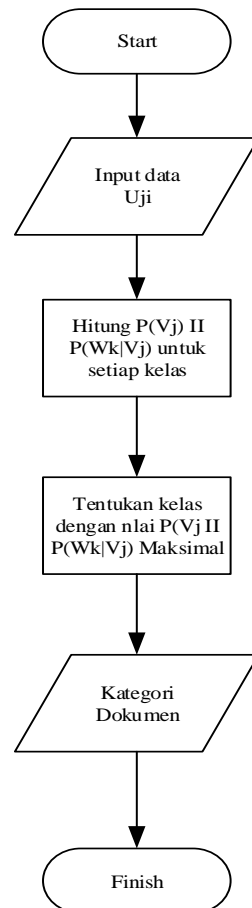
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis

H P(X) : Probabilitas X

## 2. Flowchart



**Gambar III.3. Flowchart Metode Naive Bayes**

## Dataset Tweet Dengan Label Sentimen Positif:

No.	Tweet
1	secara tak langsung bts telkomsel ikut bantu sejahterakan masy di daerah pelosok dan perbatasan
2	upgrade online usim 4g berhasil dan sim card sudah diterima.. terimakasih pelayanannya @telkomsel
3	jaringan 4g telkomsel di semarang stabil dan kencang
4	kualitas jaringan telkomsel sampai ke pelosok nusantara, terimakasih telkomsel
5	dibandingkan operator lain, telkomsel sangat memuaskan
6	salah satu jaringan 4g paling bagus stabil adalah telkomsel
7	puas dengan kecepatan lte nya telkomsel
8	pergi kemanapun gak ragu menggunakan telkomsel, jaringan terluas
9	singkat kata, puas dengan telkomsel
10	kualitas stabil jaringan terluas adalah telkomsel

## Distribusi Frekuensi Kemunculan Kata Pada Dataset Positif:

No.	Kata	Frekuensi	Probabilitas
1	secara	1/10	0.1
2	tak	1/10	0.1
3	langsung	1/10	0.1
4	bts	1/10	0.1
5	Telkomsel	9/10	0.9
6	Ikut	1/10	0.1
7	bantu	1/10	0.1
8	sejahterakan	1/10	0.1
9	masy	1/10	0.1
10	di	2/10	0.2
11	daerah	1/10	0.1
12	pelosok	2/10	0.2
13	dan	3/10	0.3
14	perbatasan	1/10	0.1
15	upgrade	1/10	0.1
16	online	1/10	0.1
17	usim	1/10	0.1
18	4g	3/10	0.3
19	berhasil	1/10	0.1
20	sim	1/10	0.1
21	card	1/10	0.1
22	sudah	1/10	0.1
23	diterima..	1/10	0.1
24	terimakasih	2/10	0.2
25	pelayanannya	1/10	0.1
26	@telkomsel	1/10	0.1
27	jaringan	5/10	0.5
28	semarang	1/10	0.1
29	stabil	3/10	0.3
30	kencang	1/10	0.1
31	kualitas	2/10	0.2
32	sampai	1/10	0.1
33	ke	1/10	0.1
34	nusantara,	1/10	0.1
35	dibandingkan	1/10	0.1
36	operator	1/10	0.1
37	lain,	1/10	0.1
38	sangat	1/10	0.1
39	memuaskan	1/10	0.1
40	salah	1/10	0.1

41	satu	1/10	0.1
42	paling	1/10	0.1
43	bagus	1/10	0.1
44	adalah	2/10	0.2
45	puas	2/10	0.2
46	dengan	2/10	0.2
47	kecepatan	1/10	0.1
48	lte	1/10	0.1
49	nya	1/10	0.1
50	pergi	1/10	0.1
51	kemanapun	1/10	0.1
52	gak	1/10	0.1
53	ragu	1/10	0.1
54	menggunakan	1/10	0.1
55	telkomsel,	1/10	0.1
56	terluas	2/10	0.2
57	singkat	1/10	0.1
58	kata,	1/10	0.1

Dataset Tweet Dengan Label Sentimen Negatif:

No.	Tweet
1	telkomsel kualitas signal makin hari makin buruk
2	jam segini ya sinyal telkomsel suka ngajak berantem
3	jaringan telkomsel lelet abis
4	telkomsel lelet super lelet
5	kalo udah ujan, telkomsel mulai lelet
6	lelet bgt internet. mqn game online jadi mati mulu berujung setres @telkomsel
7	telkomsel lagi gangguan gak sih? signal gw dari siang jelek
8	sinyal telkomsel sedang bermasalah buka facebook lite aja susah banget
9	lg gangguan kah telkomsel buka sosmed sm streaming lemot
10	@telkomsel min ini 4g tapi ko lemot banget ya bagus pindah operator

Distribusi Frekuensi Kemunculan Kata Pada Dataset Negatif:

<b>No.</b>	<b>Kata</b>	<b>Frekuensi</b>	<b>Probabilitas</b>
1	telkomsel	8/10	0.8
2	kualitas	1/10	0.1
3	signal	2/10	0.2
4	makin	2/10	0.2
5	hari	1/10	0.1
6	buruk	1/10	0.1
7	jam	1/10	0.1
8	segini	1/10	0.1
9	ya	2/10	0.2
10	sinyal	2/10	0.2
11	suka	1/10	0.1
12	ngajak	1/10	0.1
13	berantem	1/10	0.1
14	jaringan	1/10	0.1
15	lelet	5/10	0.5
16	abis	1/10	0.1
17	super	1/10	0.1
18	kalo	1/10	0.1
19	udah	1/10	0.1
20	ujan,	1/10	0.1
21	mulai	1/10	0.1
22	bgt	1/10	0.1
23	internet.	1/10	0.1

24	mqin	1/10	0.1
25	game	1/10	0.1
26	online	1/10	0.1
27	jadi	1/10	0.1
28	mati	1/10	0.1
29	mulu	1/10	0.1
30	berujung	1/10	0.1
31	setres	1/10	0.1
32	@telkomsel	2/10	0.2
33	lagi	1/10	0.1
34	gangguan	2/10	0.2
35	gak	1/10	0.1
36	sih?	1/10	0.1
37	gw	1/10	0.1
38	dari	1/10	0.1
39	siang	1/10	0.1
40	jelek	1/10	0.1
41	sedang	1/10	0.1
42	bermasalah	1/10	0.1
43	buka	2/10	0.2
44	facebook	1/10	0.1
45	lite	1/10	0.1
46	aja	1/10	0.1
47	susah	1/10	0.1

48	banget	2/10	0.2
49	lg	1/10	0.1
50	kah	1/10	0.1
51	sosmed	1/10	0.1
52	sm	1/10	0.1
53	streaming	1/10	0.1
54	lemot	2/10	0.2
55	min	1/10	0.1
56	ini	1/10	0.1
57	4g	1/10	0.1
58	tapi	1/10	0.1
59	ko	1/10	0.1
60	bagus	1/10	0.1
61	pindah	1/10	0.1
62	operator	1/10	0.1

Distribusi Frekuensi Kemunculan Kata Tanpa Memandang Sentimen:

No.	Kata	Frekuensi	Probabilitas
1	secara	1/20	0.05
2	tak	1/20	0.05
3	langsung	1/20	0.05
4	bts	1/20	0.05
5	telkomsel	17/20	0.85
6	ikut	1/20	0.05
7	bantu	1/20	0.05
8	sejahterakan	1/20	0.05
9	masy	1/20	0.05
10	di	2/20	0.1
11	daerah	1/20	0.05
12	pelosok	2/20	0.1
13	dan	3/20	0.15
14	perbatasan	1/20	0.05
15	upgrade	1/20	0.05
16	online	2/20	0.1
17	usim	1/20	0.05
18	4g	4/20	0.2
19	berhasil	1/20	0.05
20	sim	1/20	0.05
21	card	1/20	0.05
22	sudah	1/20	0.05
23	diterima..	1/20	0.05

24	terimakasih	2/20	0.1
25	pelayanannya	1/20	0.05
26	@telkomsel	3/20	0.15
27	jaringan	6/20	0.3
28	semarang	1/20	0.05
29	stabil	3/20	0.15
30	kencang	1/20	0.05
31	kualitas	3/20	0.15
32	sampai	1/20	0.05
33	ke	1/20	0.05
34	nusantara,	1/20	0.05
35	dibandingkan	1/20	0.05
36	operator	2/20	0.1
37	lain,	1/20	0.05
38	sangat	1/20	0.05
39	memuaskan	1/20	0.05
40	salah	1/20	0.05
41	satu	1/20	0.05
42	paling	1/20	0.05
43	bagus	2/20	0.1
44	adalah	2/20	0.1
45	puas	2/20	0.1
46	dengan	2/20	0.1
47	kecepatan	1/20	0.05

48	lte	1/20	0.05
49	nya	1/20	0.05
50	pergi	1/20	0.05
51	kemanapun	1/20	0.05
52	gak	2/20	0.1
53	ragu	1/20	0.05
54	menggunakan	1/20	0.05
55	telkomsel,	1/20	0.05
56	terluas	2/20	0.1
57	singkat	1/20	0.05
58	kata,	1/20	0.05
59	signal	2/20	0.1
60	makin	2/20	0.1
61	hari	1/20	0.05
62	buruk	1/20	0.05
63	jam	1/20	0.05
64	segini	1/20	0.05
65	ya	2/20	0.1
66	sinyal	2/20	0.1
67	suka	1/20	0.05
68	ngajak	1/20	0.05
69	berantem	1/20	0.05
70	lelet	5/20	0.25
71	abis	1/20	0.05

72	super	1/20	0.05
73	kalo	1/20	0.05
74	udah	1/20	0.05
75	ujan,	1/20	0.05
76	mulai	1/20	0.05
77	bgt	1/20	0.05
78	internet.	1/20	0.05
79	mqin	1/20	0.05
80	game	1/20	0.05
81	jadi	1/20	0.05
82	mati	1/20	0.05
83	mulu	1/20	0.05
84	berujung	1/20	0.05
85	setres	1/20	0.05
86	lagi	1/20	0.05
87	gangguan	2/20	0.1
88	sih?	1/20	0.05
89	gw	1/20	0.05
90	dari	1/20	0.05
91	siang	1/20	0.05
92	jelek	1/20	0.05
93	sedang	1/20	0.05
94	bermasalah	1/20	0.05
95	buka	2/20	0.1

96	facebook	1/20	0.05
97	lite	1/20	0.05
98	aja	1/20	0.05
99	susah	1/20	0.05
100	banget	2/20	0.1
101	lg	1/20	0.05
102	kah	1/20	0.05
103	sosmed	1/20	0.05
104	sm	1/20	0.05
105	streaming	1/20	0.05
106	lemot	2/20	0.1
107	min	1/20	0.05
108	ini	1/20	0.05
109	tapi	1/20	0.05
110	ko	1/20	0.05
111	pindah	1/20	0.05

Misalkan terdapat tweet dengan kalimat "**kualitas telkomsel bagus**", maka probabilitas sentimen tweet tersebut adalah:

$$P(\text{Positif}) = 10/20 = 0.5$$

$$P(\text{positif}|\text{kualitas, telkomsel, bagus}) = P(\text{positif}|\text{kualitas}) * P(\text{positif}|\text{telkomsel}) * P(\text{positif}|\text{bagus})$$

$$P(\text{positif}|\text{kualitas, telkomsel, bagus}) = 0.2 * 0.9 * 0.1 * 0.5 / 0.15 * 0.85 * 0.1$$

$$P(\text{positif}|\text{kualitas, telkomsel, bagus}) = 0.0090000000000000001 / 0.012750000000000001$$

$$P(\text{positif}|\text{kualitas, telkomsel, bagus}) = 0.7058823529411765$$

$$P(\text{positif}|\text{kualitas, telkomsel, bagus}) = \mathbf{0.7058823529411765}$$

$$P(\text{positif}|\text{kualitas, telkomsel, bagus}) = \mathbf{0.7058823529411765}$$

$$P(\text{negatif}) = 10/20 = 0.5$$

$$\begin{aligned}
P(\text{negatif}|\text{kualitas, telkomsel, bagus}) &= P(\text{negatif}|\text{kualitas}) * P(\text{negatif}|\text{telkomsel}) * \\
&P(\text{negatif}|\text{bagus}) * \text{negatif} / P(\text{kualitas}) * P(\text{telkomsel}) * P(\text{bagus}) \\
P(\text{negatif}|\text{kualitas, telkomsel, bagus}) &= 0.1 * 0.8 * 0.1 * 0.5 / 0.15 * 0.85 * 0.1 \\
P(\text{negatif}|\text{kualitas, telkomsel, bagus}) &= 0.004000000000000001 / \\
&0.012750000000000001 \\
P(\text{negatif}|\text{kualitas, telkomsel, bagus}) &= \mathbf{0.3137254901960785}
\end{aligned}$$

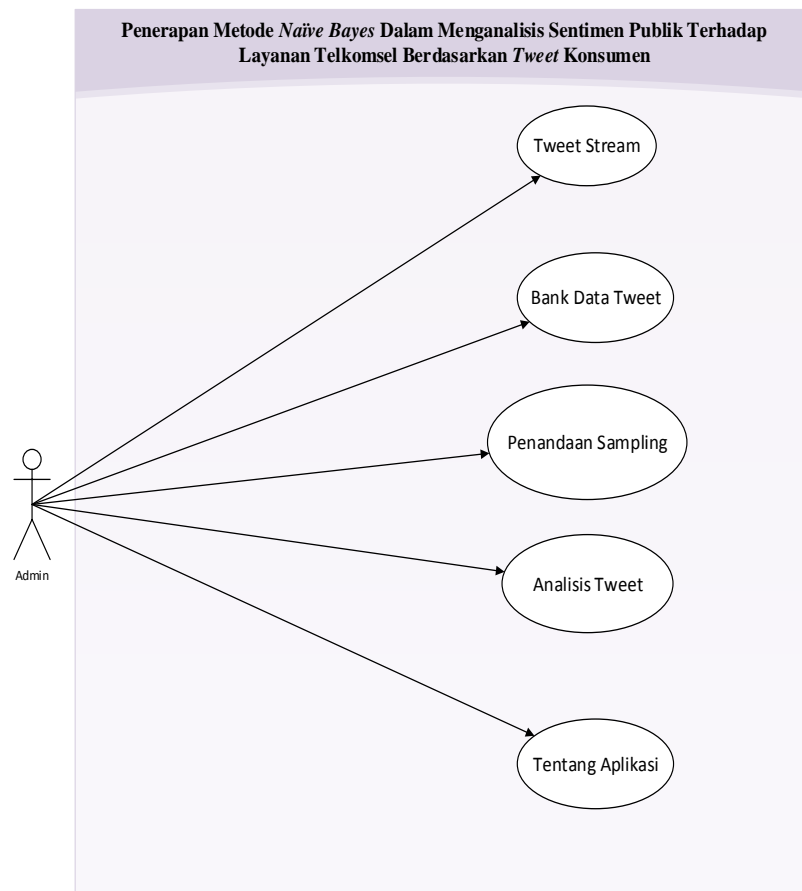
Karena probabilitas kalimat “kualitas telkomsel bagus” lebih besar pada sentimen positif, maka kalimat tersebut masuk ke sentimen positif.

### III.3. Desain Sistem Baru

Desain Sistem Baru menggunakan bahasa pemodelan UML yang terdiri dari *Usecase Diagram*, *Activity Diagram* dan *Sequence Diagram*.

#### III.3.1. *Usecase Diagram*

Secara garis besar, bisnis proses sistem yang akan dirancang digambarkan dengan *usecase diagram* yang terdapat pada Gambar III.4 :



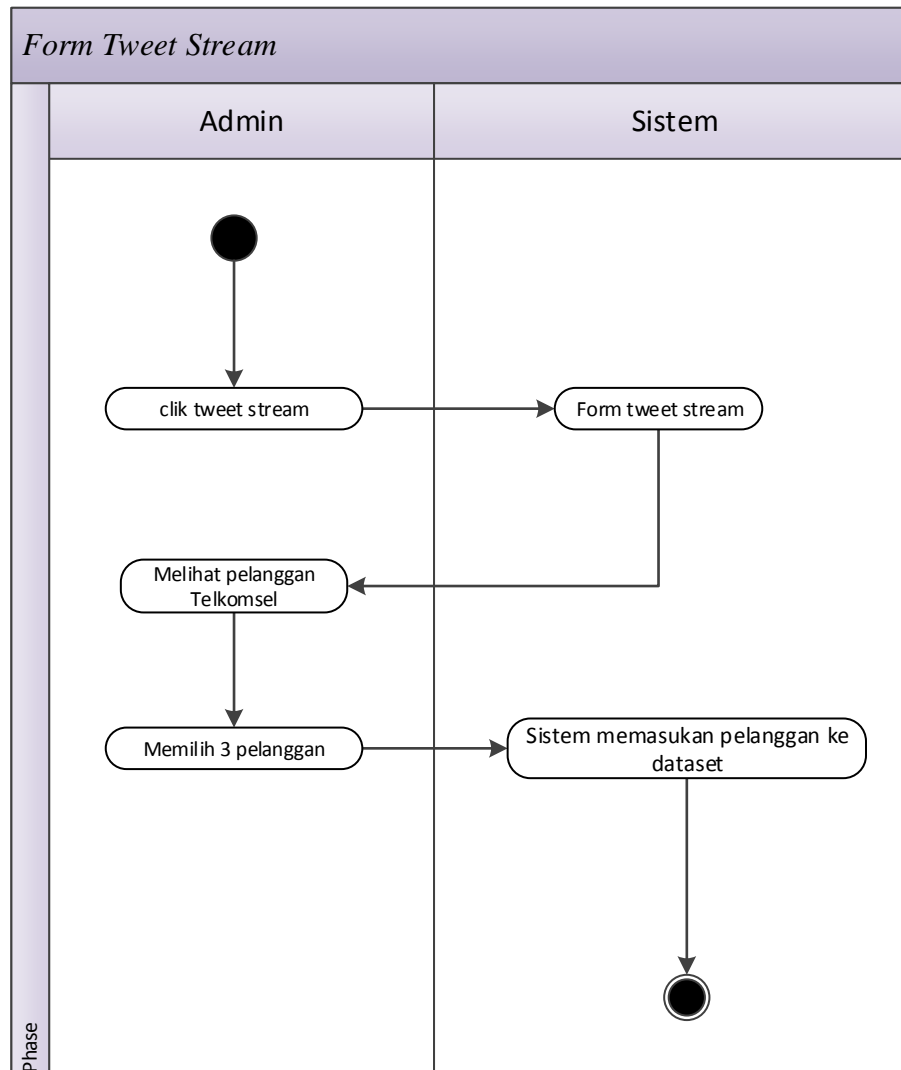
**Gambar III.4. Use Case Diagram**

### III.3.3. Activity Diagram

Bisnis proses yang telah digambarkan pada *usecase diagram* diatas dijabarkan dengan *activity diagram* :

#### 1. Activity Diagram Tweet Stream

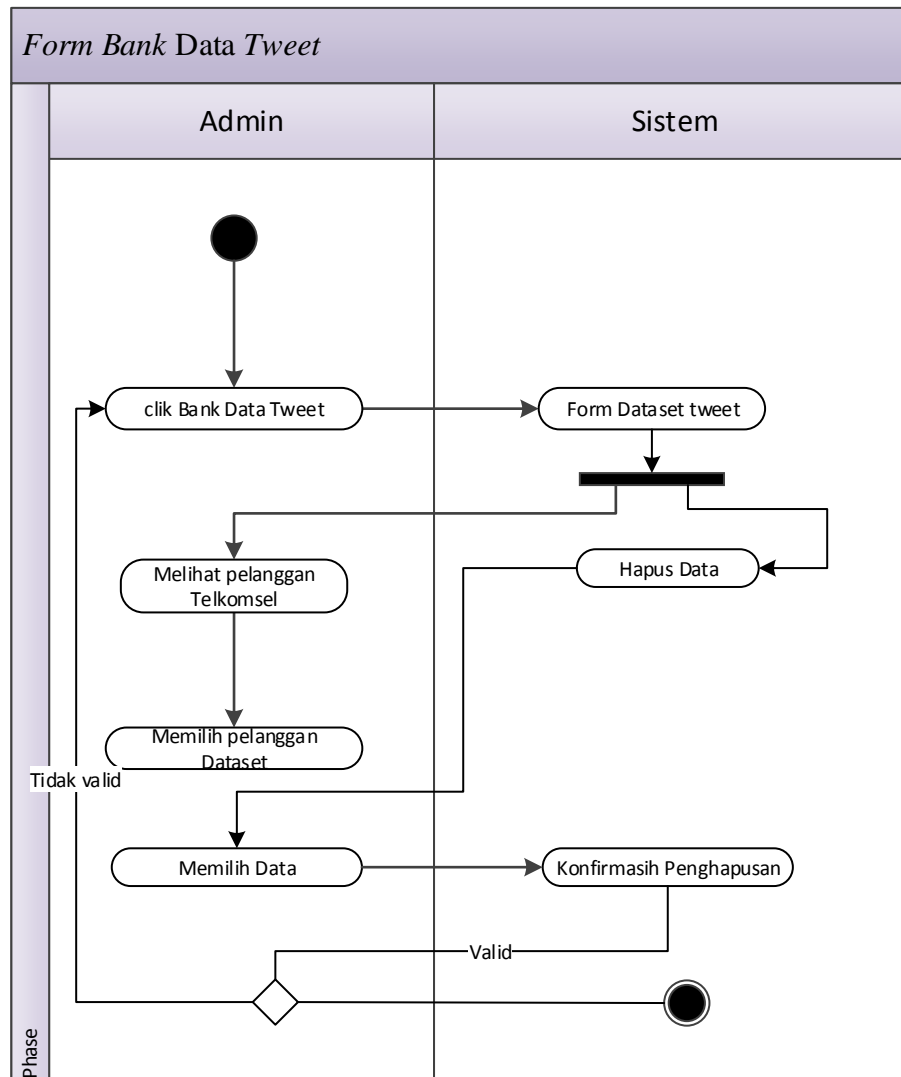
Aktifitas pada sistem *Tweet Stream* yang dilakukan oleh admin dapat diterangkan dengan langkah-langkah *state* yang ditunjukkan pada gambar III.5 berikut :



**Gambar III.5. Activity Diagram Tweet Stream**

## 2. Activity Diagram Bank Data Tweet

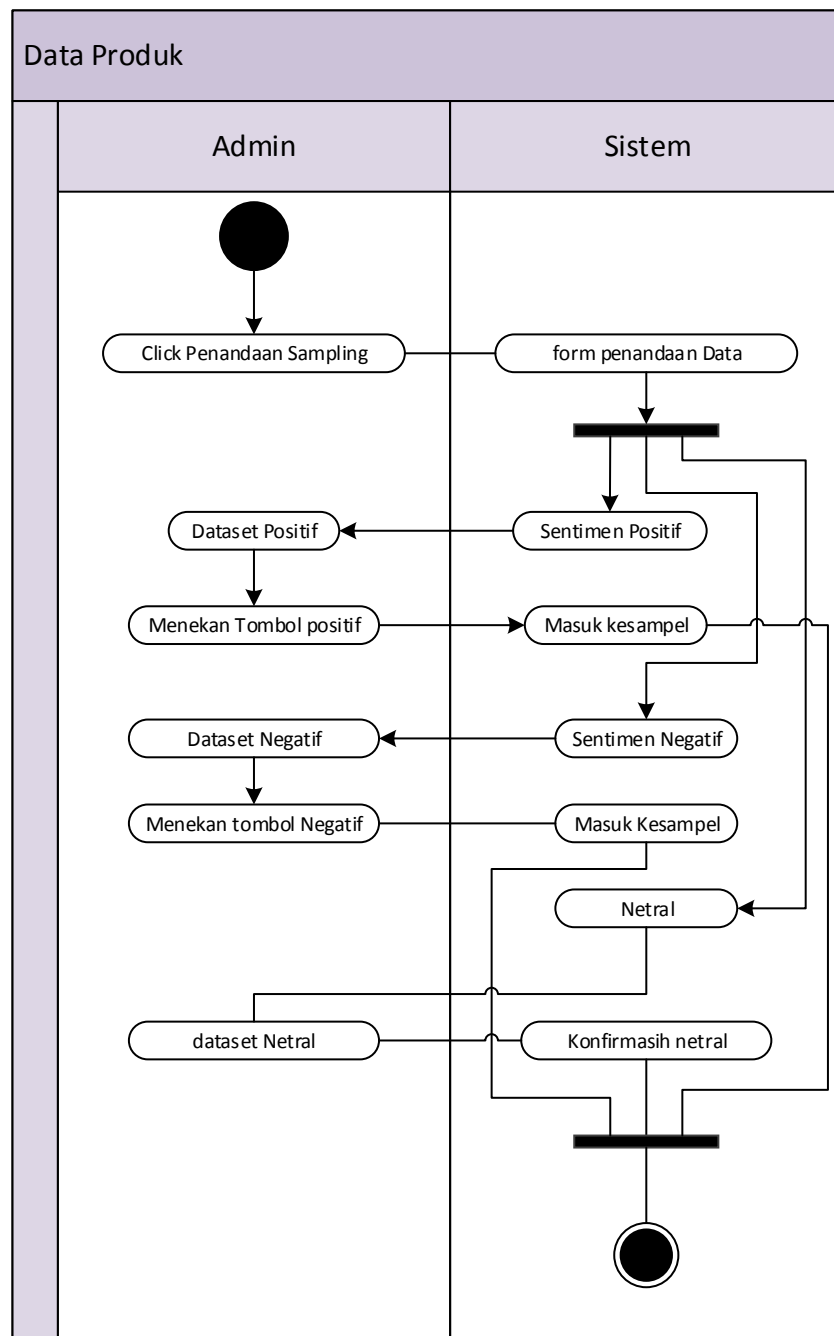
Aktivitas yang dilakukan olah dataset pelanggan pada form bank data tweet maka admin menyimpan data pelanggan seperti yang ditunjukkan pada gambar III.6:



**Gambar III.6. Activity Diagram Bank Data Tweet**

### 3. Activity Diagram Penandaan Tweet

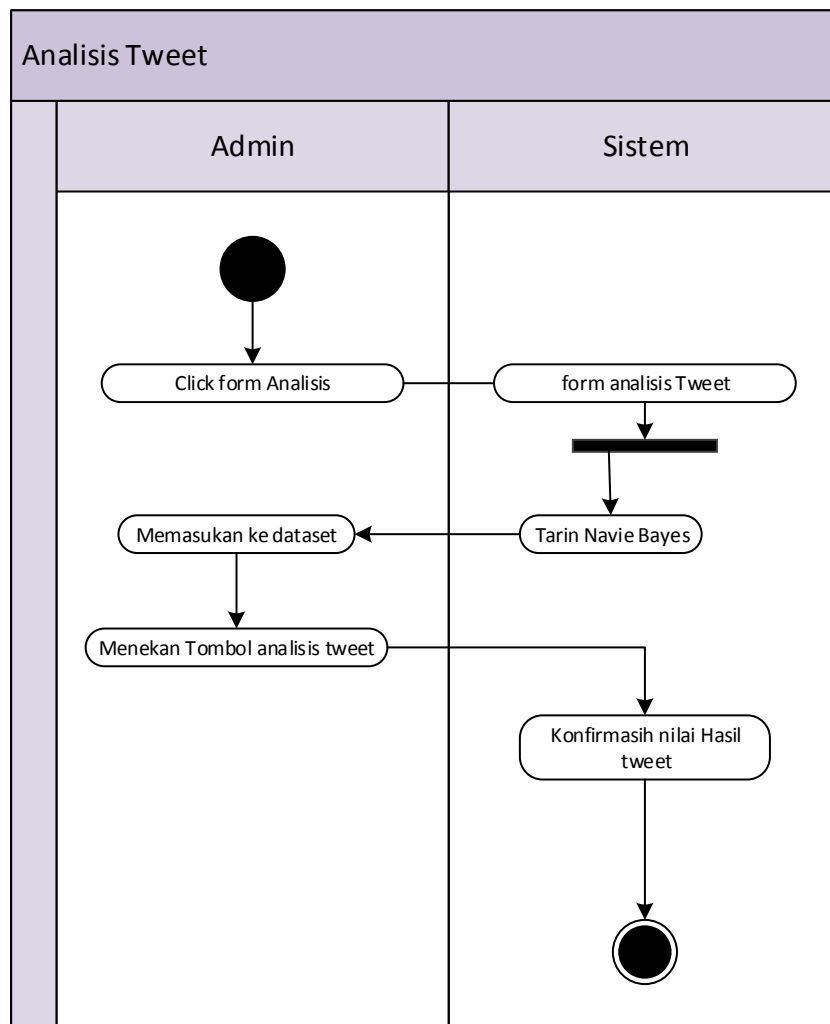
Aktivitas yang dilakukan olah dataset tweet pada form penandaan tweet maka admin menyimpan data sentimen negative, sentimen positif, netral sepeti yang ditunjukkan pada gambar III.7:



**Gambar III.7. Activity Diagram Penandaan Sampling**

#### 4. Activity Diagram Analisis Tweet

Aktivitas yang dilakukan dalam akan melakukan train naive bayes pada analisis tweet adalah admin mengklik bottom analisis tweet maka akan memunculkan hasil tweet seperti yang ditunjukkan pada gambar III.8 :



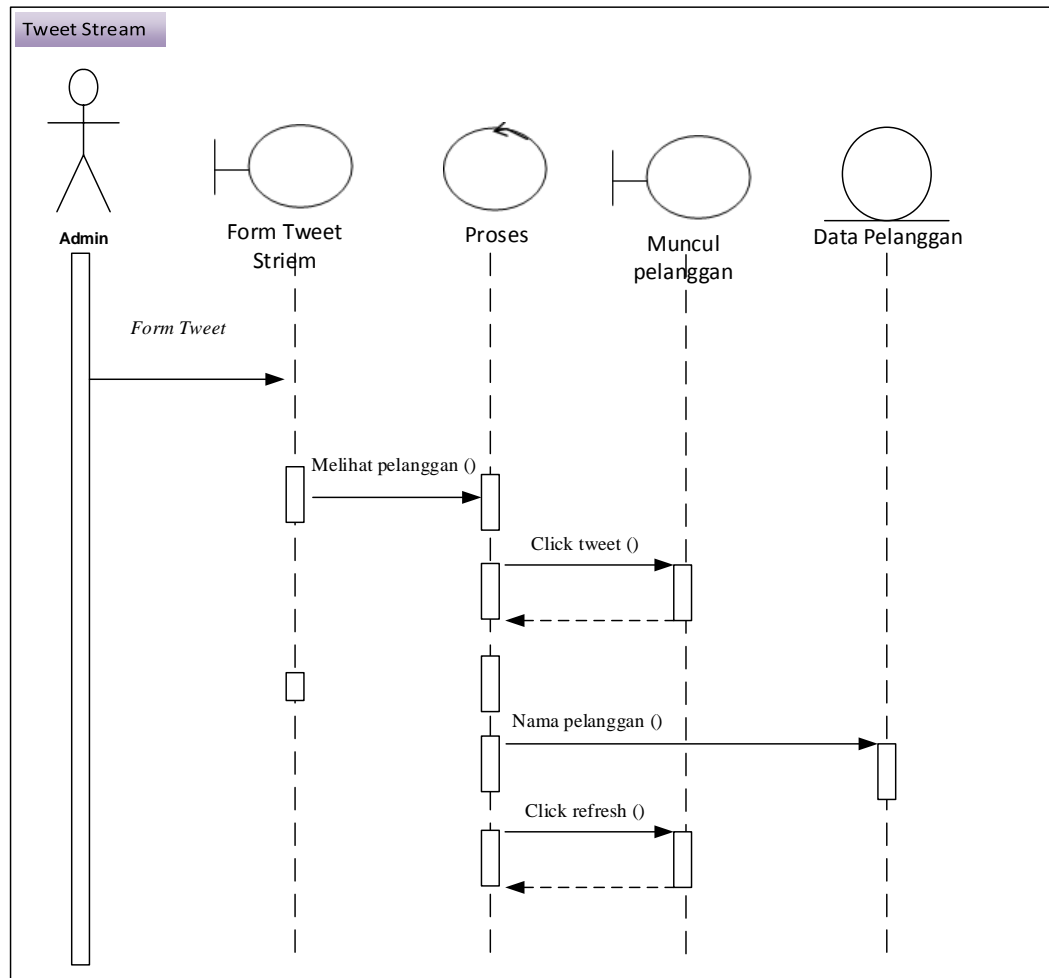
**Gambar III.8. Activity Diagram Analisis Tweet**

### III.3.4. Sequence Diagram

Rangkaian kegiatan pada setiap terjadi *event* sistem digambarkan pada *sequence* diagram berikut:

#### 1. Sequence Diagram Tweet Stream

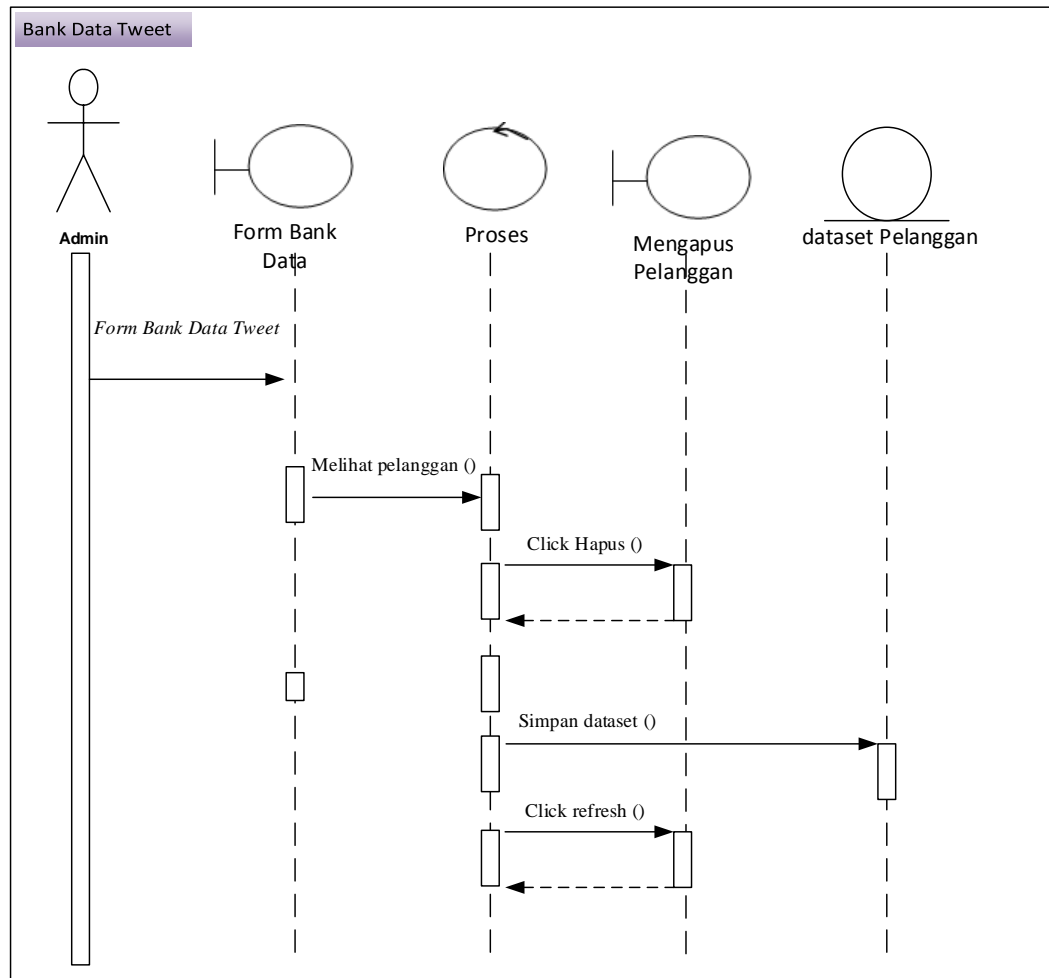
Serangkaian kinerja sistem *Tweet Stream* yang dilakukan oleh admin dapat diterangkan dengan langkah-langkah *state* yang ditunjukkan pada gambar III.9 berikut :



**Gambar III.9. Sequence Diagram Tweet Stream**

## 2. Sequence Diagram Bank Data Tweet

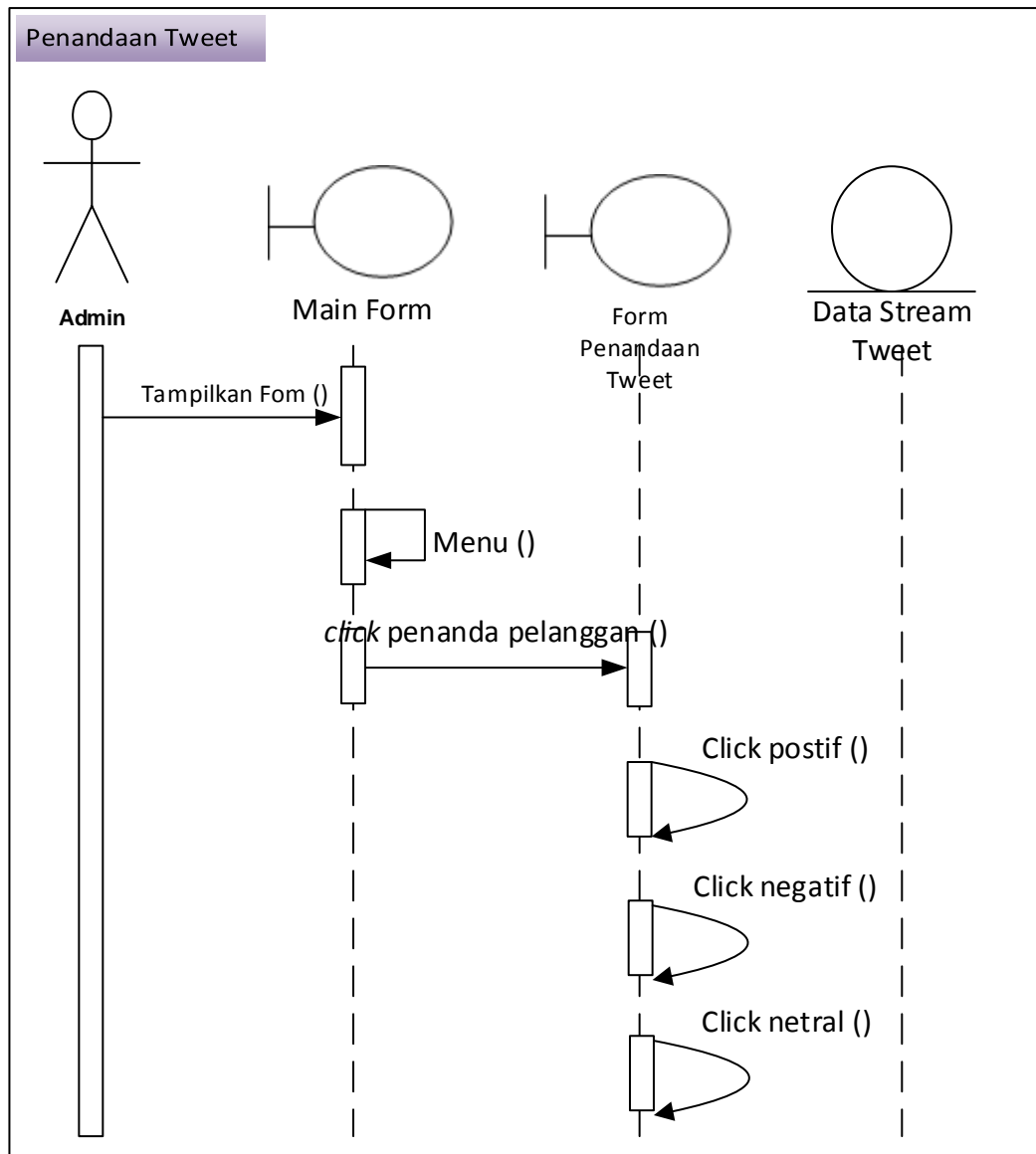
Serangkaian kinerja sistem yang dilakukan olah dataset pelanggan pada form bank data tweet maka admin menyimpan data pelanggan seperti yang ditunjukkan pada gambar III.10 berikut :



**Gambar III.10. Sequence Diagram Bank Data Tweet**

### 3. Sequence Diagram Penandaan Tweet

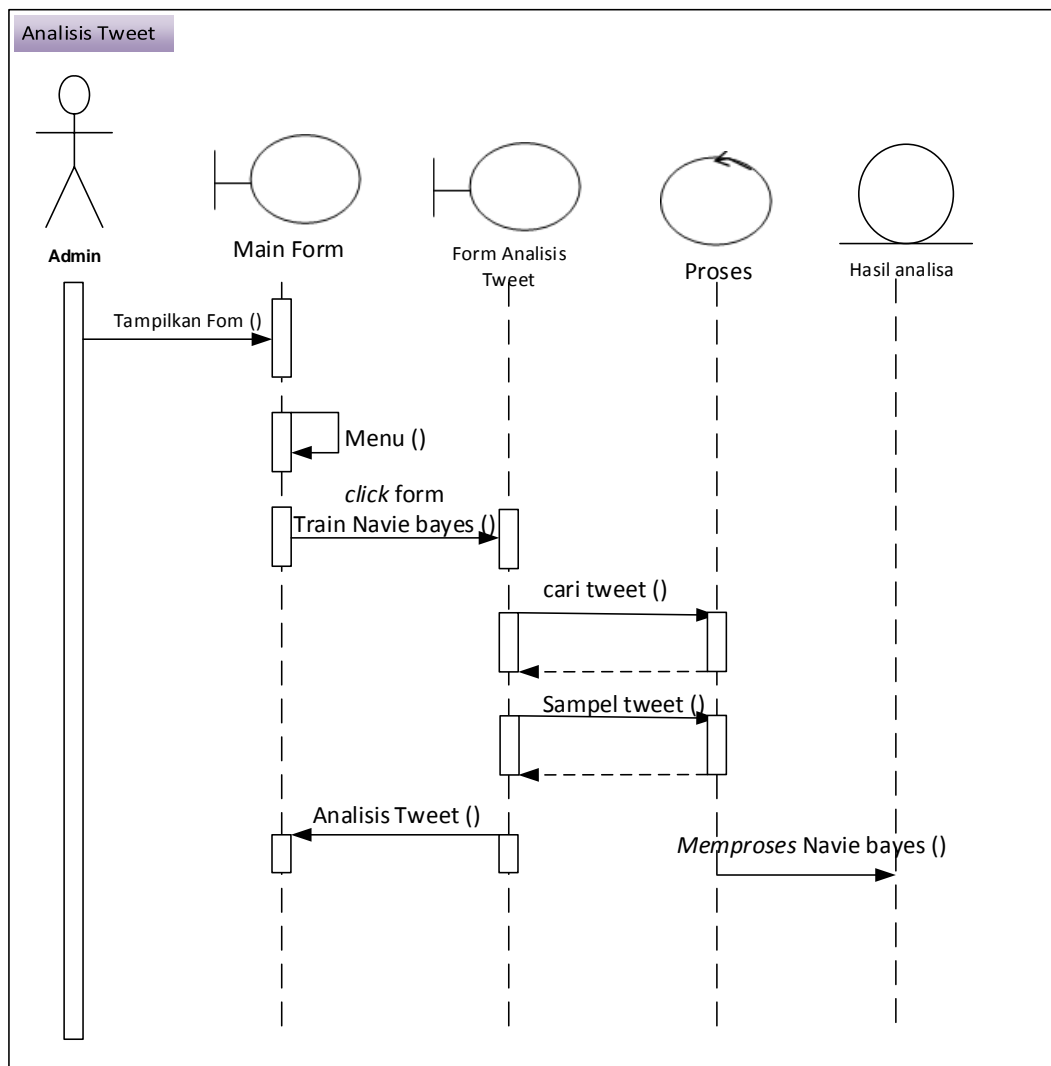
Serangkaian kinerja sistem dilakukan olah dataset tweet pada form penandaan tweet maka admin menyimpan data sentimen negative, sentimen positif, netral seperti yang ditunjukkan pada gambar III.11 berikut :



**Gambar III.11. Sequence Diagram Penandaan Tweet**

#### 4. Sequence Diagram Analisis Tweet

Serangkaian kinerja sistem yang dilakukan dalam akan melakukan train naive bayes pada analisis tweet adalah admin mengklik bottom analisis tweet maka akan memunculkan hasil tweet seperti yang ditunjukkan pada gambar III.12 berikut :



**Gambar III.12. Sequence Diagram Analisis Tweet**

### III.3.7. Desain User Interface

#### 1. Desain Form Tweet Stream

Desain yang disajikan oleh sistem yang dilakukan dalam mengelolah data tweet yang ditunjukkan pada gambar III.13 berikut :

**Gambar III.13. Desain *Form* Tweet Stream**

## 2. Desain *Form* Bank Data Tweet

Desain yang disajikan oleh sistem yang dilakukan pada form bank data tweet maka admin menyimpan data pelanggan seperti yang ditunjukkan pada gambar III.16 berikut :

**Gambar III.16. Desain *Form* Bank Data Tweet**

## 3. Desain *Form* Penandaan Sampling

Desain yang disajikan oleh dataset tweet pada form penandaan tweet maka admin menyimpan data sentimen negative, sentimen positif, netral seperti yang ditunjukkan pada gambar III.17 berikut :

Sentimen Publik Tweet Telkomsel

Tweet Stream Bank Data Tweet Penandaan Sampling Analisis Tweet Tentang

Penandaan Data

Dataset Tweet

Dataset Positif

Dataset Negatif

Dataset Netral

**Gambar III.17. Desain Form Penandaan Sampling**

#### 4. Desain Form Analisis Tweet

Desain yang disajikan oleh admin dalam akan melakukan train naive bayes pada analisis tweet adalah admin mengklik bottom analisis tweet maka akan memunculkan hasil tweet seperti yang ditunjukkan pada gambar III.18 berikut :

Sentimen Publik Tweet Telkomsel

Tweet Stream Bank Data Tweet Penandaan Sampling Analisis Tweet Tentang

Dataset Positif

Dataset Negatif

Dataset Netral

Analisis Tweet

Tanggal :

Nama :

Akun :

Tweet Text :

Hasil Analisis

Hasil Analisis Tweet

Total data :
Total Sampel :
Total Diperoses :
Total Positif :
Total Negatif :
Total Netral :

**Gambar III.18. Desain Form Analisis Tweet**