

BAB 3

METODE PENELITIAN

1.1 Kerangka Pemikiran

Kerangka berpikir yaitu suatu gambaran yang menjelaskan mengenai alur logika berjalannya penelitian secara garis besar. Berikut tahap-tahap yang dilakukan dalam kerangka berpikir penelitian, antara lain:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari portal twitter. Data yang diperoleh merupakan kumpulan berita yang didapatkan dengan menggunakan *web scraping*.

2. Pelabelan

Pelabelan adalah tahapan dimana berita diberi label yang nantinya akan digunakan pada proses *training* di tahap klasifikasi. Terdapat berita yang disediakan yaitu berita yang berisi informasi mengenai artis. Pelabelan dilakukan dengan menggunakan program aplikasi *web* dan dilakukan secara manual oleh annotator. Kelemahan dari pelabelan secara manual yang dilakukan oleh annotator adalah hasil yang subjektif. Hasil dari tahapan ini adalah kumpulan berita yang memiliki label positive dan negative.

3. Preprocessing

Proses yang dilakukan dalam tahapan ini adalah sebagai berikut:

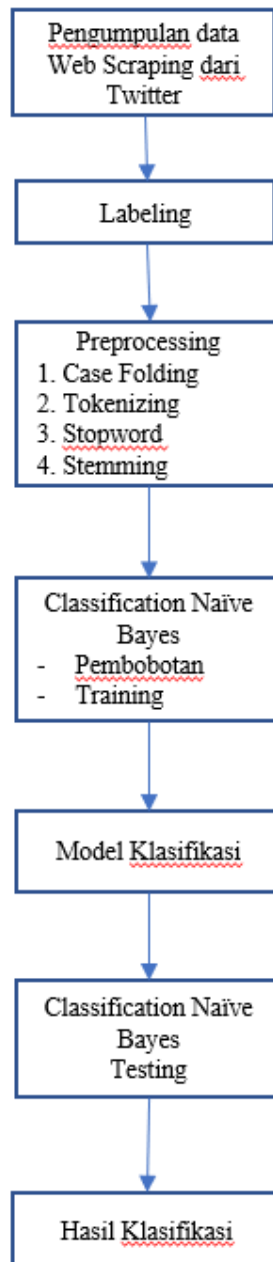
- a. *Case Folding*, yaitu untuk menyeragamkan bentuk huruf.
- b. *Tokenizing*, yaitu pemenggalan suku kata
- c. *Stopword*, yaitu menghilangkan kata yang tidak deskriptif.
- d. *Stemming*, yaitu mengubah suku kata menjadi kata dasar.

4. Klasifikasi

Proses klasifikasi dibedakan menjadi dua proses yaitu:

- a. *Training*, proses ini digunakan untuk melatih algoritma klasifikasi yang digunakan yaitu algoritma *Naive Bayes* agar mampu melakukan prosesnya sesuai dengan yang diharapkan. Pada tahap ini pertama-tama akan dilakukan proses pembobotan terhadap kumpulan berita hasil pelabelan menggunakan perhitungan TF-IDF dengan hanya menghitung TF (*term frequency*)-nya saja. Selanjutnya akan dihasilkan model klasifikasi yang nantinya digunakan pada tahap *testing*.

- b. *Testing*, proses ini dilakukan untuk melakukan pengklasifikasian terhadap dataset dengan memanfaatkan model klasifikasi yang dihasilkan pada proses *training*. Hasil pada tahap ini adalah kumpulan berita yang telah diklasifikasikan ke dalam kategori berita. Untuk jelasnya dapat dilihat pada Gambar 3.1.

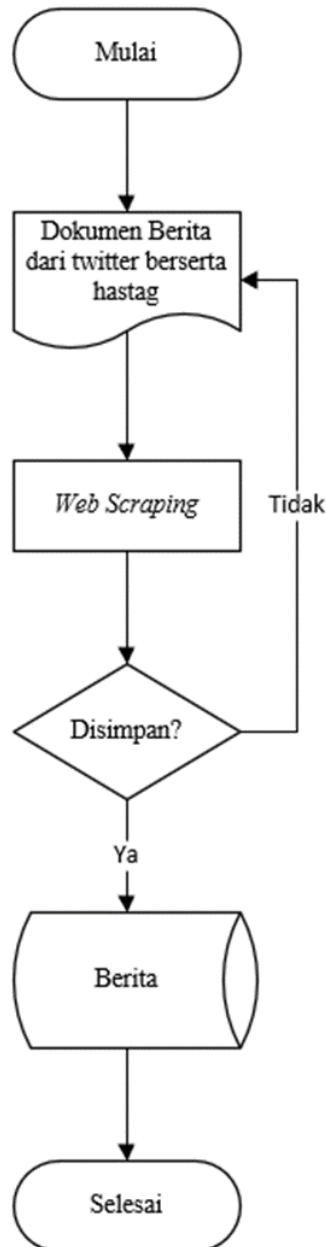


Gambar 3. 1 Kerangka Berpikir

1.2 Metode Penelitian

1.2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari twitter dengan memanfaatkan *web scraping*. Untuk lebih jelasnya dapat dilihat pada Gambar 3.2 *flowchart* pengumpulan data berikut :



Gambar 3. 2 Flowchart Pengumpulan Data

Berikut contoh hasil Data Scraping:

Tabel 3. 1 Contoh Data Scraping

Contoh Hasil Data Scraping Twitter
@chikautami12 di block syahrini entah karena salah apa
@benibernard belajar dari syahrini ga usah banyak ga enakan sama temenlah
@maudhyaa please jangan hujat syahrini semua ada sebab akibatnya
@RajaPurwa Ini bukti bahwa pak @Deddy_Mizwar_ asal bicara,tanpa bukti, sangat disayangkan sebagai publik figur
@thefailedjedii Mereka lupa kalau publik figur itu tindak tanduknya akan dicontoh masyarakat (awam)
@asrinoov Akhirnya video nikahan syahrini udah nongol + preskonnya. Bisa tidur nyenyak gw :”
@RichelFachrein Jadi? Kalau bukan publik figur sah-sah aja kalau sharing? Ini salah wawancarain apa gimana deh? Goblog amat.
@jessvbrn Syahrini ama luna yang punya urusan, netizen""nya pada pusing, salah satunya aku

Data yang diambil setiap kali melakukan *web scraping* berjumlah maksimal 100.

```
function get_tweet($keyword)
{
    $jumlah_tweet = 100;

    //membuat koneksi ke Twitter API
    $connection = $this->twitteroauth->create($this->config->
    item('twitter_consumer_token'), $this->config->item(
    'twitter_consumer_secret'), $this->config->item(
    'twitter_access_token'), $this->config->item(
    'twitter_access_secret'));

    //mendapatkan tweet
    $data['tweets'] = $connection->get("search/tweets", ["q"
    => $keyword,"count" => $jumlah_tweet,"lang"=>"id"]);
    return $data['tweets'];
}
```

Gambar 3. 3 Snippet dari code untuk mengambil tweet

Data yang diambil sebanyak 1291 tweet, lalu dibagi menjadi 2 bagian untuk mencari hasil evaluasi dengan perhitungan *confusion matrix* : 80% untuk *data training* dan 20% *data testing*. Data yang diambil untuk dijadikan model berjumlah 1033 tweet.



Gambar 3. 4 Query untuk menghitung jumlah tweet untuk *Data Training*

1.2.2 Labeling

Pelabelan adalah tahapan dimana berita diberi label yang nantinya akan digunakan pada proses *training* di tahap klasifikasi. Terdapat berita yang disediakan yaitu berita yang berisi informasi mengenai artis. Pelabelan dilakukan dengan menggunakan program aplikasi *web* dan dilakukan secara manual oleh annotator. Sebagai contoh, annotator akan memberikan label secara manual ke masing-masing berita atau *tweet* apakah berita tersebut mengandung sentimen positif, negatif, atau netral berdasarkan pemikiran annotator tersebut. Kelemahan dari pelabelan secara manual yang dilakukan oleh annotator adalah hasil yang subjektif. Hasil dari tahapan ini adalah kumpulan berita yang memiliki label *positive*, *negative*, dan *neutral*. Berikut gambar dari potongan code untuk melakukan perubahan kategori pada tweet.

```

public function doNegative($id)
{
    $dataArray['tweet_tag'] = '0';

    $update = $this->db->update("tbl_tweet" , $dataArray, array('tweet_id' => $id ));
    if($update){
        $this->m_umum->generatePesan("Berhasil mengubah data","berhasil");
        redirect('admin/tweet');
    }else{
        $this->m_umum->generatePesan("Gagal mengubah data","gagal");
        redirect('admin/tweet');
    }
}

public function doPositive($id)
{
    $dataArray['tweet_tag'] = '1';

    $update = $this->db->update("tbl_tweet" , $dataArray, array('tweet_id' => $id ));
    if($update){
        $this->m_umum->generatePesan("Berhasil mengubah data","berhasil");
        redirect('admin/tweet');
    }else{
        $this->m_umum->generatePesan("Gagal mengubah data","gagal");
        redirect('admin/tweet');
    }
}

public function doNeutral($id)
{
    $dataArray['tweet_tag'] = '2';

    $update = $this->db->update("tbl_tweet" , $dataArray, array('tweet_id' => $id ));
    if($update){
        $this->m_umum->generatePesan("Berhasil mengubah data","berhasil");
        redirect('admin/tweet');
    }else{
        $this->m_umum->generatePesan("Gagal mengubah data","gagal");
        redirect('admin/tweet');
    }
}
}

```

Gambar 3.5 Snippet dari code untuk mengganti tag pada tweet

- Untuk tweet *negative* diberi tag 0
- Untuk tweet *positive* diberi tag 1
- Untuk tweet *neutral* diberi tag 2

Dari 1033 data yang telah diambil dari web scraping sebelumnya, dilakukan pelabelan secara manual dan menghasilkan data positive, negative, dan neutral sebagai berikut :

- Tweet *positive* sebanyak 271 data (diberi tag 1)

```

SELECT COUNT(*) FROM `tbl_tweet` WHERE tweet_tag='1'
.....
+ Options
COUNT(*)
271

```

Gambar 3. 6 Query untuk menghitung jumlah tweet positif

- Tweet *negative* sebanyak 369 data (diberi tag 0)

```
SELECT COUNT(*) FROM `tbl_tweet` WHERE tweet_tag='0'
```

+ Options
COUNT(*)
369

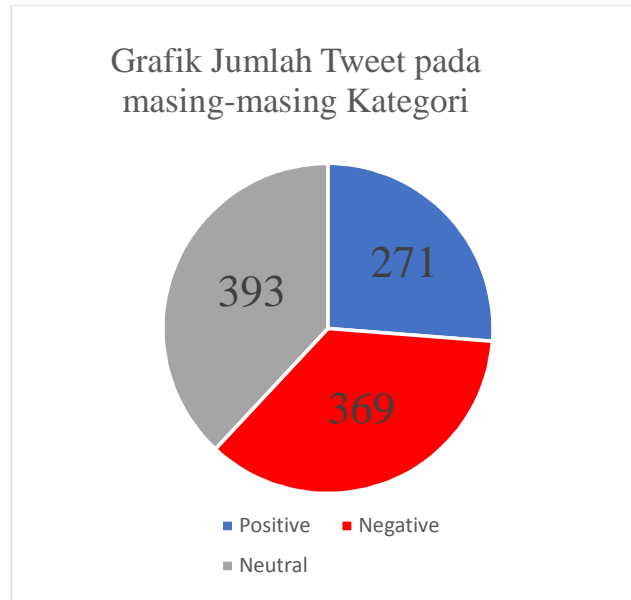
Gambar 3. 7 Query untuk menghitung jumlah tweet negatif

- Tweet *neutral* sebanyak 393 data (diberi tag 2)

```
SELECT COUNT(*) FROM `tbl_tweet` WHERE tweet_tag='2'
```

+ Options
COUNT(*)
393

Gambar 3. 8 Query untuk menghitung jumlah tweet netral



Gambar 3. 9 Grafik jumlah tweet pada masing-masing kategori

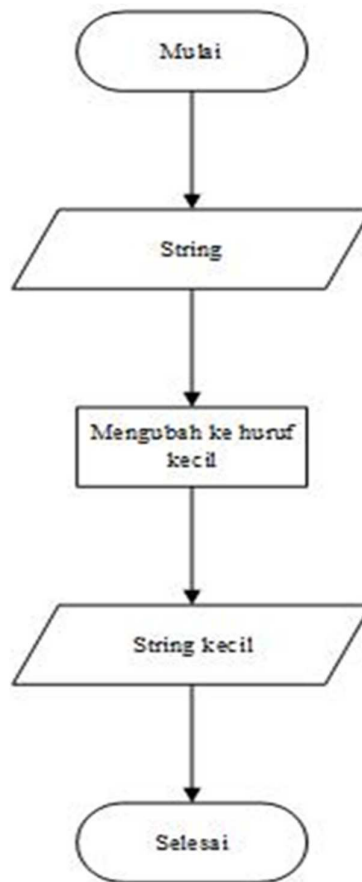
Tahap *preprocessing* adalah tahapan dimana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen. Proses *preprocessing* meliputi

- 1) *Case Folding* (menyeragamkan bentuk huruf)
- 2) *Tokenizing* (pemenggalan suku kata)
- 3) *Stopword* (menghilangkan kata yang tidak deskriptif)
- 4) *Stemming* (mengubah suku kata menjadi kata dasar)

Yang nanti akan diberikan label, pembobotan dan proses klasifikasi menggunakan *Naive Bayes Classifier*.

1.2.3 *Case Folding*

Tidak semua dokumen teks konsisten dalam menggunakan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan untuk mengkonversikan keseluruhan teks dalam dokumen menjadi suatu bentuk standar (huruf kecil atau *lowercase*). Sebagai contoh, user yang ingin mendapatkan informasi tentang “BERITA” dan mengetik “BeRiTa”, “BERITA”, atau “berita” tetap diberikan hasil retrieval yang sama yakni “berita”. *Case folding* adalah mengubah semua huruf dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf akan dihilangkan. Lebih jelasnya dapat dilihat pada Gambar 3.10 Flowchart *case folding*.



Gambar 3. 10 Flowchart *Case Folding*

```

function caseFolding($tweet)
{
    $new = clone $tweet;
    $new->tweet_isi = strtolower($new->tweet_isi);
    $new->tweet_isi = preg_replace('/[0-9]/','',$new->tweet_isi);
    return $new;
}
  
```

Gambar 3. 11 Snippet dari code untuk melakukan *Case Folding*

1.2.4 Tokenizing

Tahap *tokenizing* proses mengubah dokumen menjadi kumpulan *term* dengan cara menghapus semua karakter tanda baca yang terdapat pada *token*. Hingga pada akhirnya yang diperoleh hanya kumpulan kata-kata dari suatu teks/dokumen. Contoh dari tahap *tokenizing* dapat dilihat pada Tabel 3.2. Contoh *Tokenizing* sebagai berikut :

Tabel 3. 2 Contoh *Tokenizing*

Teks Input	Teks Output
Beri bantuan khas kepada kami atau tiada lagi hiburan untuk anda'	beri bantuan khas kepada kami atau tiada lagi hiburan untuk anda

Tokenizing secara garis besar memecah sekumpulan karakter dalam suatu teks kedalam satuan kata. Sebagai contoh, karakter *whitespace*, seperti enter, tabulasi, spasi dianggap sebagai pemisah kata. Namun untuk karakter tunggal (‘), titik (.), semikolon (;), titik dua (:), atau lainnya, dapat memiliki peran yang cukup banyak sebagai pemisah kata.

```
function tokenizing($tweet)
{
    $new = clone $tweet;
    $return = str_replace("\n", ' ', $new->tweet_isi);
    $return = str_replace(' ', '-', $return);
    $return = preg_replace('/[^\A-Za-z0-9\-\_]/', '', $return);
    $return = str_replace('-', ' ', $return);
    $return = str_replace(" ", "|", $return);
    $return = explode('|', $return);

    foreach ($return as $keys => $values) {
        if (ctype_space($values)) {
            unset($return[$keys]); //menghilangkan whitespace
        }
    }
    foreach ($return as $keys => $values) {
        if (empty($values)) {
            unset($return[$keys]); //menghilangkan values yang kosong
        }
    }
    $return = implode('|', $return);
    $new->tweet_isi = $return;
    return $new;
}
```

Gambar 3. 12 Snippet dari code untuk melakukan *Tokenizing*

1.2.5 *Stopword*

Pada tahap ini dilakukan pembuangan kata-kata yang kurang penting atau kata-kata yang sering muncul (*Stopword*), seperti kata penghubung dan kata keterangan yang bukan merupakan kata unik misalnya “sebuah”, “oleh”, “pada”, dan sebagainya (Parveen & Pandey, 2017). Contoh dari tahap *stopword* dapat dilihat pada Tabel 3.3. Contoh *Stopword* sebagai berikut :

Tabel 3. 3 Contoh Stopword

Hasil <i>Tokenizing</i>	Hasil <i>Filtering</i>
Beri bantuan khas kepada kami atau tiada lagi hiburan untuk anda	bantuan khas tiada hiburan

```
function stopWord($tweet)
{
    $kataStopWord = 'ada adalah adanya adapun agak agaknya agar akan akankah akhir akhiri
    apalagi apatah artinya asal asalkan atas atau ataukah ataupun awal awalnya bagi bagai
    balik banyak bapak baru bawah beberapa begini beginian beginikah beginilah begitu begi
    berada berakhir berakhirlah berakhirnya berapa berapakah berapalah berapapun berarti k
    berkehendak berkeinginan berkenaan berlainan berlalu berlangsung berlebihan bermacam k
    berturut berturut-turut bertutur berujar berupa besar betul betulkah biasa biasanya bi
    cara caranya cukup cukupkah cukuplah cuma dahulu dalam dan dapat dari daripada datang
    diantaranya diberi diberikan diberikannya dibuat dibuatnya didapat didatangkan diguna
    dikarenakan dikatakan dikatakannya dikerjakan diketahui diketahuinya dikira dilakukan
    dimulai dimulailah dimulainya dimungkinkan dini dipastikan diperbuat diperbuatnya dipe
    diri dirinya disampaikan disebut disebutkan disebutkannya disini disinilah ditambhakar
    ditunjukkannya ditunjuknya dituturkan dituturkannya diucapkan diucapkannya diungkapk
    harus haruslah harusnya hendak hendaklah hendaknya hingga ia ialah ibarat ibaratkan it
    jadi jadilah jadinya jangan janganlah jauh jawab jawaban jawabnya jelas jelas
    kalian kami kamilah kamu kamulah kan kapan kapankah kapanpun karena karenanya kasus ke
    kelihatan kelihatannya kelima keluar kembali kemudian kemungkinan kemungkinannya kena
    kinilah kira kira-kira kiranya kita kitalah kok kurang lagi lagian lah lain lainnya le
    mampu mampukah mana manakala manalagi masa masalah masalahnya masih masikhkah masing ma
    memberikan membuat memerlukan memihak meminta memintakan memisalkan memperbuat memper
    memungkinkan menaiki menambahkan menandakan menanti menanti-nanti menantikan menanya
    mengapa mengatakan mengatakannya mengenai mengerjakan mengetahui menggunakan menghend
    mengucapkannya mengungkapkan menjadi menjawab menjelaskan menuju menunjuk menunjuki me
    menyeluruh menyiapkan merasa mereka merekalah merupakan meski meskipun meyakini meyaki
    namun nanti nantinya nyaris nyatanya oleh olehnya pada padahal padanya pak paling panj
    persoalan pertama pertama-tama pertanyaan pertanyakan pihak pihaknya pukul pula pun pu
    sampai-sampai sampaikan sana sangat sangatlah satu saya sayalah se sebab sebabnya sebe
    sebegitu sebelum sebelumnya sebenarnya seberapa sebesar sebetulnya sebisanya sebuah se
    segala segalanya segera seharusnya sehingga seingat sejak sejauh sejenak sejumlah seke
    sekiranya sekitar sekitarnya sekurang-kurangnya sekurangnya sela selain selaku selalu
    semasa semasih semata semata-mata semaunya sementara semisal semisalnya sempat semua s
    sepantasnyalah seperlunya seperti sepertinya sepihak sering seringnya serta serupa ses
    setempat setengah seterusnya setiap setiba setibanya setidak-tidaknya setidaknya setir
    sudahlah supaya tadi tadinya tahu tahun tak tambah tambahannya tampak tampaknya tandas t
    tentulah tentunya tepat terakhir terasa terbanyak terdahulu terdapat terdiri terhadap
    terlihat termasuk ternyata tersampaikan tersebut tersebutlah tertentu tertuju terus te
    tutur tuturnya ucap ucapnya ujar ujarnya umum umumnya ungkap ungkapnya untuk usah usai

    $kataStopWord = explode(' ', $kataStopWord);
    $new = clone $tweet;
    $return = preg_replace('/\b('implode('|', $kataStopWord).')\b/', '', $new->tweet_isi);
    $return = explode('|', $return);
    foreach ($return as $keys => $values) {
        if (empty($values) || $values == '-') {
            unset($return[$keys]);
        }
    }
    $new->tweet_isi = implode('|', $return);
    return $new;
}
```

Gambar 3. 13 Snippet dari code untuk melakukan Stopword

Tabel 3. 4 List Stopword

Kata	Kata	Kata	Kata
ada	beginilah	biasanya	dikatakan
adalah	begitu	bila	dikatakannya
adanya	begitukah	bilakah	dikerjakan
adapun	begitulah	bisa	diketahui
agak	begitupun	bisakah	diketuainya
agaknya	bekerja	boleh	dikira
agar	belakang	bolehkah	dilakukan
akan	belakangan	bolehlah	dilalui
akankah	belum	buat	dilihat
akhir	belumah	bukan	dimaksud
akhiri	benar	bukankah	dimaksudkan
akhirnya	benarkah	bukanlah	dimaksudkannya
aku	benarlah	bukannya	dimaksudnya
akulah	berada	bulan	diminta
amat	berakhir	bung	dimintai
amatlah	berakhirlah	cara	dimisalkan
anda	berakhirnya	caranya	dimulai
andalah	berapa	cukup	dimulailah
antar	berapakah	cukupkah	dimulainya
antara	berapalah	cukuplah	dimungkinkan
antaranya	berapapun	cuma	dini
apa	berarti	dahulu	dipastikan
apaan	berawal	dalam	diperbuat
apabila	berbagai	dan	diperbuatnya
apakah	berdatangan	dapat	dipergunakan
apalagi	beri	dari	diperkirakan
apatah	berikan	daripada	diperlihatkan
artinya	berikut	datang	diperlukan
asal	berikutnya	dekat	diperlukannya
asalkan	berjumlah	demi	dipersoalkan
atas	berkali-kali	demikian	dipertanyakan
atau	berkata	demikianlah	dipunyai
ataukah	berkehendak	dengan	diri
ataupun	berkeinginan	depan	dirinya
awal	berkenaan	di	disampaikan
awalnya	berlainan	dia	disebut
bagai	berlalu	diakhiri	disebutkan
bagaikan	berlangsung	diakhirinya	disebutkannya
bagaimana	berlebihan	dialah	disini
bagaimanakah	bermacam	diantara	disinilah
bagaimanapun	bermacam-macam	diantaranya	ditambahkan
bagi	bermaksud	diberi	ditandaskan
bagian	bermula	diberikan	ditanya
bahkan	bersama	diberikannya	ditanyai
bahwa	bersama-sama	dibuat	ditanyakan
bahwasanya	bersiap	dibuatnya	ditegaskan
baik	bersiap-siap	didapat	ditujukan
bakal	bertanya	didatangkan	ditunjuk
bakalan	bertanya-tanya	digunakan	ditunjuki
balik	berturut	diibaratkan	ditunjukkan
banyak	berturut-turut	diibaratkannya	ditunjukkannya
bapak	bertutur	diingat	ditunjuknya
baru	berujar	diingatkan	dituturkan
bawah	berupa	diinginkan	dituturkannya
beberapa	besar	dijawab	diucapkan

begini beginian beginikah	betul betulkah biasa	dijelaskan dijelaskannya dikarenakan	diucapkannya diungkapkan dong
---------------------------------	----------------------------	--	-------------------------------------

Kata	Kata	Kata	Kata
dua	justru	lain	menandaskan
dulu	kala	lainnya	menanti
empat	kalau	lalu	menanti-nanti
enggak	kalaupun	lama	menantikan
enggaknya	kalian	lamanya	menanya
entah	kami	lanjut	menanyai
entahlah	kamilah	lanjutnya	menanyakan
guna	kamu	lebih	mendapat
gunakan	kamulah	lewat	mendapatkan
hal	kan	lima	mendatang
hampir	kapan	luar	mendatangi
hanya	kapankah	macam	mendatangkan
hanyalah	kapanpun	maka	menegaskan
hari	karena	makanya	mengakhiri
harus	karenanya	makin	mengapa
haruslah	kasus	malah	mengatakan
harusnya	kata	malahan	mengatakannya
hendak	katakan	mampu	mengenai
hendaklah	katakanlah	mampukah	mengerjakan
hendaknya	katanya	mana	mengetahui
hingga	ke	manakala	menggunakan
ia	keadaan	manalagi	menghendaki
ialah	kebetulan	masa	mengibaratkan
ibarat	kecil	masalah	mengibaratkannya
ibaratkan	kedua	masalahnya	mengingat
ibaratnya	keduanya	masih	mengingatkan
ibu	keinginan	masihkah	menginginkan
ikut	kelamaan	masing	mengira
ingat	kelihatan	masing-masing	mengucapkan
ingat-ingat	kelihatannya	mau	mengucapkannya
ingin	kelima	maupun	mengungkapkan
inginkah	keluar	melainkan	menjadi
inginkan	kembali	melakukan	menjawab
ini	kemudian	melalui	menjelaskan
inikah	kemungkinan	melihat	menuju
inilah	kemungkinannya	melihatnya	menunjuk
itu	kenapa	memang	menunjuki
itukah	kepada	memastikan	menunjukkan
itulah	kepadanya	memberi	menunjuknya
jadi	kesampaian	memberikan	menurut
jadilah	keseluruhan	membuat	menuturkan
jadinya	keseluruhannya	memerlukan	menyampaikan
jangan	keterlalu	memihak	menyangkut
janganakan	ketika	meminta	menyatakan
janganlah	khususnya	memintakan	menyebutkan
jauh	kini	memisalkan	menyeluruh
jawab	kinilah	memperbuat	menyiapkan
jawaban	kira	mempergunakan	merasa
jawabnya	kira-kira	memperkirakan	mereka
jasas	kiranya	memperlihatkan	merekalah
jelaskan	kita	mempersiapkan	merupakan
jelastah	kitalah	mempersoalkan	meski
jelastnya	kok	mempertanyakan	meskipun
jika	kurang	mempunyai	meyakini
jikalau		memulai	meyakinkan

juga jumlah jumlahnya	lagi lagian lah	memungkinkan menaiki menambahkan	minta mirip misal
-----------------------------	-----------------------	--	-------------------------

Kata	Kata	Kata	Kata
sambil	sekarang	sesuatu	terdapat
sampai	sekarang	sesuatunya	terdiri
sampai-sampai	sekecil	sesudah	terhadap
sampaikan	seketika	sesudahnya	terhadapnya
sana	sekiranya	setelah	teringat
sangat	sekitar	setempat	teringat-ingat
sangatlah	sekitarnya	setengah	terjadi
satu	sekurangnya	seterusnya	terjadilah
saya	kurangnya	setiap	terjadinya
sayalah	sekurangnya	setiba	terkira
se	sela	setibanya	terlalu
sebab	selain	setidak-tidaknya	terlebih
sebabnya	selaku	setidaknya	terlihat
sebagai	selalu	setinggi	termasuk
sebagaimana	selama	seusai	ternyata
sebagainya	selama-lamanya	sewaktu	tersampaikan
sebagian	selamanya	siap	tersebut
sebaik	selanjutnya	siapa	tersebutlah
sebaik-baiknya	seluruh	siapakah	tertentu
sebaiknya	seluruhnya	siapapun	tertuju
sebaliknya	semacam	sini	terus
sebanyak	semakin	sinilah	terutama
sebegini	semampu	soal	tetap
sebegini	semampunya	soalnya	tetapi
sebelum	semasa	suatu	tiap
sebelumnya	semasih	sudah	tiba
sebenarnya	semata	sudahkah	tiba-tiba
seberapa	semata-mata	sudahlah	tidak
sebesar	semaunya	supaya	tidakkah
sebetulnya	sementara	tadi	tidaklah
sebisanya	semisal	tadinya	tiga
sebuah	semisalnya	tahu	tinggi
sebut	sempat	tahun	toh
sebutlah	semua	tak	tunjuk
sebutnya	semuanya	tambah	turut
secara	semula	tambahnya	tutur
secukupnya	sendiri	tampak	tuturnya
sedang	sendirian	tampaknya	ucap
sedangkan	sendirinya	tandas	ucapnya
sedemikian	seolah	tandasnya	ujar
sedikit	seolah-olah	tanpa	ujarnya
sedikitnya	seorang	tanya	umum
sejaknya	sepanjang	tanyakan	umumnya
segala	sepantasnya	tanyanya	ungkap
segalanya	sepantasnyalah	tapi	ungkapnya
segera	seperlunya	tegas	untuk
seharusnya	seperti	tegasnya	usah
sehingga	sepertinya	telah	usai
seingat	sepihak	tempat	waduh
sejak	sering	tengah	wah
sejauh	seringnya	tentang	wahai
sejenak	serta	tentu	waktu
sejumlah	serupa	tentulah	waktunya
sekadar	sesaat	tentunya	walaupun
sekadarnya	sesama	tepat	

sekali sekali-kali sekalian sekaligus sekalipun	sesampai sesegera sesekali seseorang	terakhir terasa terbanyak terdahulu	wong yaitu yakini yakni yang
---	---	--	--

1.2.6 Stemming

Tahap *Stemming* adalah proses menghapus imbuhan, awalan, akhiran yang bertujuan untuk mengubah kata-kata sesuai dengan kata dasarnya. Contoh dari tahap stemming dapat dilihat pada Tabel 3.5. Contoh Stemming sebagai berikut :

Tabel 3. 5 Contoh Stemming

Hasil <i>Filtering</i>	Hasil <i>Stemming</i>
bantuan khas tiada hiburan	bantu khas tiada hibur

```
function stemming($kata){ //fungsi pencarian akar kata

    $cekKata = $this->cariKamusData($kata);
    if($cekKata == true){ // Cek Kamus
        return $kata; // Jika Ada maka kata tersebut adalah kata dasar
    }else{ //jika tidak ada dalam kamus maka dilakukan stemming
        //untuk menghapus suffix seperti -ku, -mu, -kah, dsb
        $kata = $this->Del_Inflection_Suffixes($kata);
        if($this->cariKamusData($kata)){
            return $kata;
        }
        // Hapus Derivation Suffixes ("-i", "-an" atau "-kan")
        $kata = $this->Del_Derivation_Suffixes($kata);
        if($this->cariKamusData($kata)){
            return $kata;
        }
        // Hapus Derivation Prefix ("di-", "ke-", "se-", "te-", "be-", "me-", atau "pe-")
        $kata = $this->Del_Derivation_Prefix($kata);
        if($this->cariKamusData($kata)){
            return $kata;
        }
    }
    return $kata;
}
```

Gambar 3. 14 Snippet dari code untuk melakukan Stemming

1.2.7 Pembobotan Kata

Dalam klasifikasi berita, pembobotan kata digunakan untuk mendapatkan suatu kategori. Salah satu metode pembobotan adalah *TF-IDF* (*Term Frequency –Inverse Document Frequency*).

Nilai bobot suatu kata (*term*) menyatakan kepentingan bobot tersebut dalam merepresentasikan judul. Pada pembobotan *TF-IDF*, bobot akan semakin besar jika

frekuensi kemunculan kata semakin tinggi, tetapi bobot akan berkurang jika kata tersebut semakin sering muncul pada berita lainnya.

Dari persamaan tersebut, diketahui:

$$idf = \log\left(\frac{N}{df}\right) \quad (3.1)$$

N = Tweet

Df = Banyaknya tweet dimana suatu kata (term) muncul

Sebagai contoh terdapat tiga tweet untuk dihitung bobot katanya :

Tabel 3. 6 Contoh Pembobotan Kata

Tweet	Isi Tweet
Tweet 1	bantu khas tiada hibur
Tweet 2	pemerintah bantu masyarakat makan
Tweet 3	hibur malam makan daging

Lalu dihitung bobot tiap katanya :

Tabel 3. 7 Contoh Perhitungan Pembobotan Kata

TF				df	N/df	IDF	TF * IDF		
Kata	D1	D2	D3				D1	D2	D3
bantu	1	1	0	2	1.5	0.176	0.176	0.176	0
khas	1	0	0	1	3	0.477	0.477	0	0
tiada	1	0	0	1	3	0.477	0.477	0	0
hibur	1	0	1	2	1.5	0.176	0.176	0	0.176
pemerintah	0	1	0	1	3	0.477	0	0.477	0
masyarakat	0	1	0	1	3	0.477	0	0.477	0
makan	0	1	1	2	1.5	0.176	0	0.176	0.176
malam	0	0	1	1	3	0.477	0	0	0.477
daging	0	0	1	1	3	0.477	0	0	0.477


```

foreach ($dataPerkata as $key => $value) {
    foreach ($tweet as $keys => $values) {
        $sadaCuy = substr_count($values['preposisi'], $value['kata']);
        if ($sadaCuy > 0) {
            $sadaCuy = 1;
        }
        $dataPerkata[$key]['doc'][$key].$values['tweet_id'] += $sadaCuy;
        $dataPerkata[$key]['df'] += $sadaCuy; //kata ini muncul berapa kali
    }
    $dataPerkata[$key]['idf'] = round(log(($dataPerkata[$key]['jumlah_doc']/$dataPerkata[$key]['df']),4),5);
}

```

Gambar 3. 15 Snippet dari code untuk melakukan perhitungan pembobotan kata

1.2.8 Naive Bayes Classifier

Tahap ini merupakan tahap penentuan keterhubungan antara kata-kata pada data. Tahap ini menggunakan sebuah algoritma *Naive Bayes Classifier*. *Naive Bayes Classifier* terdiri dari dua proses dalam proses klasifikasi datanya. Kedua proses itu adalah proses pembelajaran *Naive Bayes Classifier* dan Proses klasifikasi *Naive Bayes Classifier*.

- a. Proses Prior Probabilitas

Tabel 3. 8 Contoh Data Testing

Tweet	Isi Tweet	Kelompok
Tweet 1	selamat syahrini sudah dapat pacar baru	P
Tweet 2	masyarakat merasa kasihan kepada teman artis tersebut	P
Tweet 3	artis selingkuh dengan teman pacar	N
Tweet4	syahrini merebut pacar teman	N
Tweet5	syahrini nikung teman	N
Tweet6	youtuber indonesia paling tonton kalo raditya dika	P
Tweet7	semoga publik figur ikut jejak bang ari untung	P
Tweet8	publik figur sah aja sharing salah wawancara apa gimana deh goblog	N

Lalu dilakukan preprocessing dan menghasilkan tweet sebagai berikut:

Tabel 3. 9 Contoh Data Testing

Tweet	Isi Tweet	Kelompok
Tweet 1	selamat pacar	P
Tweet 2	masyarakat kasihan teman artis	P
Tweet 3	artis selingkuh teman pacar	N
Tweet4	rebut pacar teman	N
Tweet5	nikung teman	N
Tweet6	youtuber indonesia paling tonton	P
Tweet7	semoga publik figur jejak	P
Tweet8	publik figur sah sharing salah wawancara goblog	N

Tahap ini melakukan perhitungan pada kata yang terdapat data test dan rumusnya adalah sebagai berikut:

$$P(v_j) = \frac{|doc\ j|}{|D\ training|}$$

$$P(a_i|v_j) = \frac{|n_k + 1|}{|n + kata|}$$

Keterangan :

1. Docs_j : kumpulan dokumen yang memiliki kategori v_j.
2. |D training| : jumlah dokumen yang digunakan dalam pelatihan (kumpulan data latih).
3. n : jumlah total kata yang terdapat di dalam kata tekstual yang memiliki nilai fungsi target yang sesuai.
4. n_k : jumlah kemunculan kata a_i pada semua data tekstual yang memiliki nilai fungsi target yang sesuai.
5. | kata | : jumlah kata yang berbeda yang muncul dalam seluruh data tekstual yang digunakan.

1. Kategori Positif:

$$P(V_{positif}) = \frac{(V_{positif})}{|D\ Training|} = \frac{4}{8} = 0.5$$

2. Kategori Negatif:

$$P(V_{negatif}) = \frac{(V_{negatif})}{|D Training|} = \frac{4}{8} = 0.5$$

Setelah $P(V_j)$ dihitung, maka langkah berikutnya adalah mencari nilai $P(a_i|V_j)$ dari masing-masing term pada masing-masing kategori yang dihitung dengan persamaan 2.8.

Perhitungan $P(a_i|V_j)$ sebagai berikut:

1. Term “selamat”

$$P(a_{selamat}|V_{positif}) = \frac{n_k+1}{n+|kata|} = \frac{1+1}{14+30} = \frac{2}{44} = 0,0454545454$$

$$P(a_{selamat}|V_{negatif}) = \frac{n_k+1}{n+|kata|} = \frac{0+1}{16+30} = \frac{1}{46} = 0,021739130434$$

2. Term “pacar”

$$P(a_{pacar}|V_{positif}) = \frac{n_k+1}{n+|kata|} = \frac{1+1}{14+30} = \frac{2}{44} = 0,0454545454$$

$$P(a_{pacar}|V_{negatif}) = \frac{n_k+1}{n+|kata|} = \frac{2+1}{16+30} = \frac{3}{46} = 0,0652173913$$

3. Term “selingkuh”

$$P(a_{selingkuh}|V_{positif}) = \frac{n_k+1}{n+|kata|} = \frac{0+1}{14+30} = \frac{1}{44} = 0,02272727$$

$$P(a_{selingkuh}|V_{negatif}) = \frac{n_k+1}{n+|kata|} = \frac{1+1}{16+30} = \frac{2}{46} = 0,043478260$$

4. Term “rebut”

$$P(a_{rebut}|V_{positif}) = \frac{n_k+1}{n+|kata|} = \frac{0+1}{14+30} = \frac{1}{44} = 0,02272727$$

$$P(a_{rebut}|V_{negatif}) = \frac{n_k+1}{n+|kata|} = \frac{1+1}{16+30} = \frac{2}{46} = 0,043478260$$

5. Term “teman”

$$P(a_{teman}|V_{positif}) = \frac{n_k+1}{n+|kata|} = \frac{1+1}{14+30} = \frac{2}{44} = 0.0243902439$$

$$P(a_{teman}|V_{negatif}) = \frac{n_k+1}{n+|kata|} = \frac{2+1}{16+30} = \frac{3}{46} = 0,04545454$$

6. Term “artis”

$$P(a_{\text{artis}}|V_{\text{positif}}) = \frac{n_k+1}{n+|\text{kata}|} = \frac{1+1}{14+30} = \frac{2}{44} = 0.0243902439$$

$$P(a_{\text{artis}}|V_{\text{negatif}}) = \frac{n_k+1}{n+|\text{kata}|} = \frac{1+1}{16+30} = \frac{2}{46} = 0,043478260$$

Jika dibuat menjadi sebuah tabel, maka probabilitas setiap kata pada data testing seperti pada tabel berikut :

Tabel 3.10 Perhitungan Probabilitas Data Testing

Kategori	P(v _j)	P(a _i V _j)					
		selamat	pacar	selingkuh	rebut	teman	artis
P	0.5	0,04545454	0,0454545	0,02272727	0,02272727	0.0243902439	0.0243902439
N	0.5	0,02173913	0,0652173	0,043478260	0,043478260	0,04545454	0,043478260

b. Proses Klasifikasi *Naive Bayes Classifier*

Secara umum proses ini menentukan kelas dari data test dengan menggunakan rumus sebagai berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2, \dots, a_n | V_j) P(V_j)$$

Keterangan:

VMAP: semua kategori yang diujikan

V_j: Kategori tweet, dengan:

a₁: Kata pertama

a₂: Kata kedua

a_n: Kata seterusnya

P(a_i|V_j): probabilitas a_i pada kategori V_j

P(V_j): probabilitas dari V_j

Dari hasil perhitungan $P(V)$ dan $P(a|V)$, selanjutnya adalah menentukan kategori dari kasus testing dengan persamaan VMAP dicari peluang dari masing-masing kategori.

Tabel 3. 11 Contoh Kasus *Testing*

Tweet	<i>Fitur</i>	Kategori
Tweet9	Syahrini jahat karena selingkuh dengan pacar temannya	?
Tweet10	Syahrini sudah dapat pacar baru.	?

Proses testing ini dihitung probabilitasnya dan dicari probabilitas tertinggi menggunakan persamaan VMAP berikut:

$$- P(\textit{Tweet9}|V_{positif})$$

$$= P(a_{jahat}|V_{positif}) \times P(a_{selingkuh}|V_{positif}) \times P(a_{pacar}|V_{positif}) \times P(a_{teman}|V_{positif}) \times P(V_{positif})$$

$$= 1 \times 0,02272727 \times 0,0454545454 \times 0,0243902439 \times 0,5$$

$$= \underline{1,2598263 \times 10^{-5}}$$

$$- P(\textit{Tweet9}|V_{negatif})$$

$$= P(a_{jahat}|V_{negatif}) \times P(a_{selingkuh}|V_{negatif}) \times P(a_{pacar}|V_{negatif}) \times P(a_{teman}|V_{negatif}) \times P(V_{negatif})$$

$$= 1 \times 0,043478260 \times 0,0652173913 \times 0,04545454 \times 0,5$$

$$= \underline{6,4443963 \times 10^{-5}}$$

$$- P(\textit{Tweet10}|V_{positif})$$

$$= P(a_{pacar}|V_{positif}) \times P(V_{positif})$$

$$= 0,0454545454 \times 0,5$$

$$= \underline{227272727 \times 10^{-2}}$$

$$\begin{aligned}
& - P(\textit{Tweet10} | V_{\textit{negatif}}) \\
& = P(a_{\textit{pacar}} | V_{\textit{negatif}}) \times P(V_{\textit{negatif}}) \\
& = 0,0652173913 \times 0.5 \\
& = \underline{\underline{326086955 \times 10^{-2}}}
\end{aligned}$$

Setelah menghitung probabilitas dari setiap data tester, diperoleh hasil pada tabel 3.10 sebagai berikut:

Tabel 3. 12 Nilai Probabilitas *Data Testing*

Tweet	Probabilitas	
	Positif	Negatif
Tweet9	$1,2598263 \times 10^{-5}$	$6,4443963 \times 10^{-5}$
Tweet10	227272727×10^{-2}	326086955×10^{-2}

Pada tabel 3.10 dapat dianalisis hasil dari data testing pertama yaitu Tweet9 didapatkan nilai probabilitas negatif lebih besar dari nilai probabilitas positif sehingga dapat disimpulkan bahwa Tweet9 termasuk sentimen negatif. Pada Tweet10 didapatkan nilai probabilitas negatif lebih besar dari nilai probabilitas negatif sehingga Tweet10 termasuk kategori sentimen negatif. Berikut gambar potongan code untuk perhitungan Naïve Bayes Classifier :

```

foreach ($dataTr as $key => $value) {
    $datanyaExp = explode("|", $value['preposisi']);
    $cocok = [];
    $masuk = [];
    foreach ($datanyaExp as $keys => $values) {
        if (!empty($datatfIdf[$values])) {
            if (!in_array($values, $masuk)) { //jika kata tidak ada di array masuk
                $cocok[] = $datatfIdf[$values];
                $masuk[] = $values; //simpan kata setelah dicocokkan.
            }
        }
    }

    $jumlahDoc = [];
    foreach ($cocok as $keyz => $valuez) {
        foreach ($valuez['p'] as $keys => $values) {
            if (empty($jumlahDoc[$keys])) {
                $jumlahDoc[$keys] = $values;
            } else {
                $jumlahDoc[$keys] = round($jumlahDoc[$keys]*$values,5);
                if ($jumlahDoc[$keys]<0.0001) {
                    $jumlahDoc[$keys] = 0;
                }
            }
        }
    }

    if (count($jumlahDoc)>0) {
        $keyMax = array_keys($jumlahDoc, max($jumlahDoc)); //mencari nilai terbesar
        $keyMax = str_replace("D", "", $keyMax[0]);
        $dataTr[$key]['kategori'] = $dataTraining[$keyMax]['tweet_tag']; //klasifikasi hasil perhitungan nbc
    } else {
        $dataTr[$key]['kategori'] = "tidak memiliki kategori";
    }
}
return $dataTr;
}

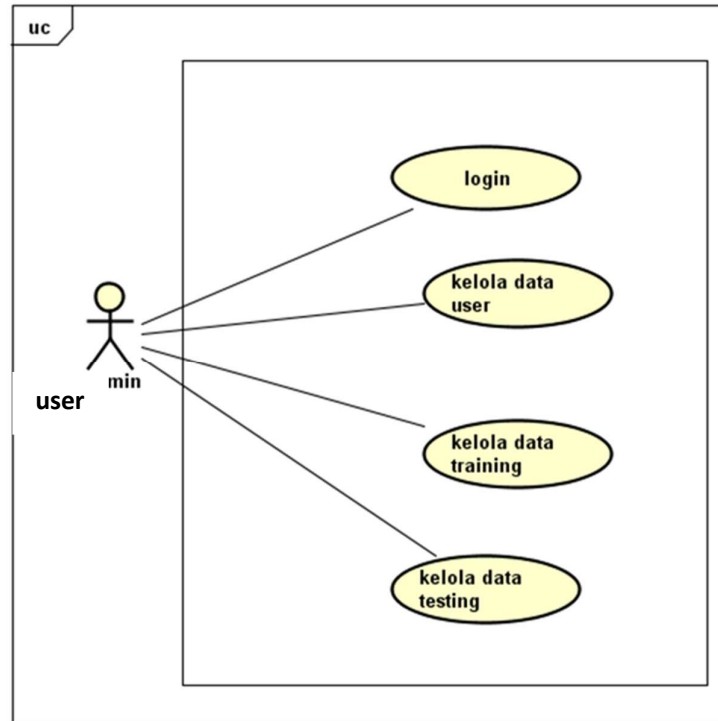
```

Gambar 3. 16 Snippet dari code untuk melakukan perhitungan *Naïve Bayes Classifier*

1.3 Perancangan Sistem

1.3.1 *Use Case Diagram*

Berikut ini adalah *use case diagram* pada sistem analisis sentimen berita artis di twitter:



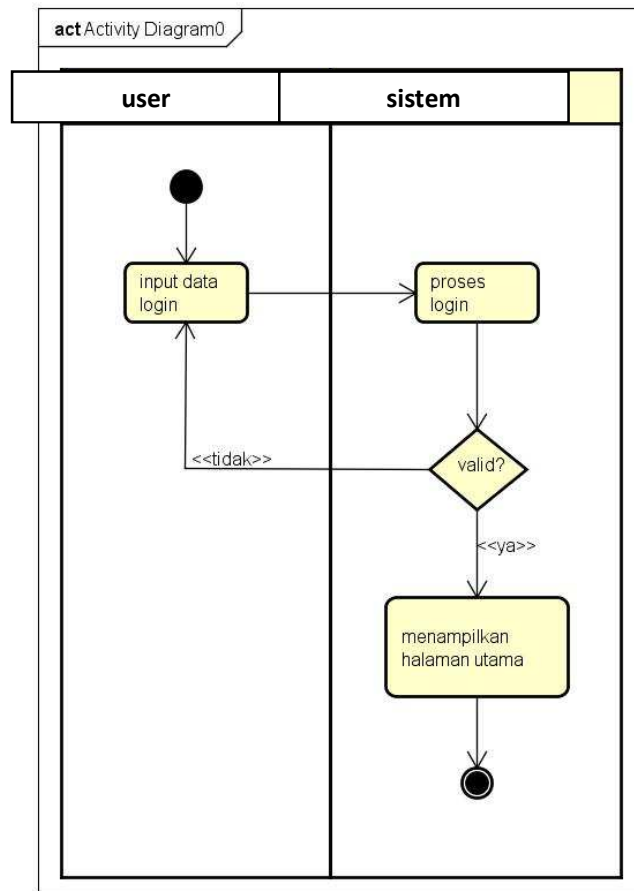
Gambar 3. 17 Use Case Diagram

1.3.2 Activity Diagram

Activity diagram merupakan alur aktivitas pengguna terhadap sistem. Dengan adanya *activity diagram* dapat mengetahui alur interaksi yang terjadi pada *use case diagram*:

1. Activity Diagram Login

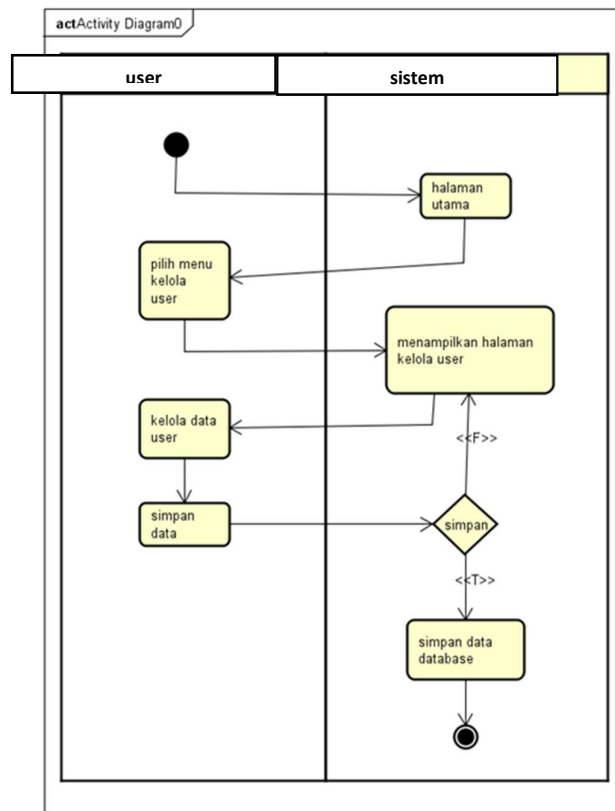
Diagram aktifitas ini menggambarkan aktifitas admin saat berada di halaman Login. Berikut ini alur aktifitas proses Login



Gambar 3. 18 Activity Diagram Login

2. Kelola Data *User*

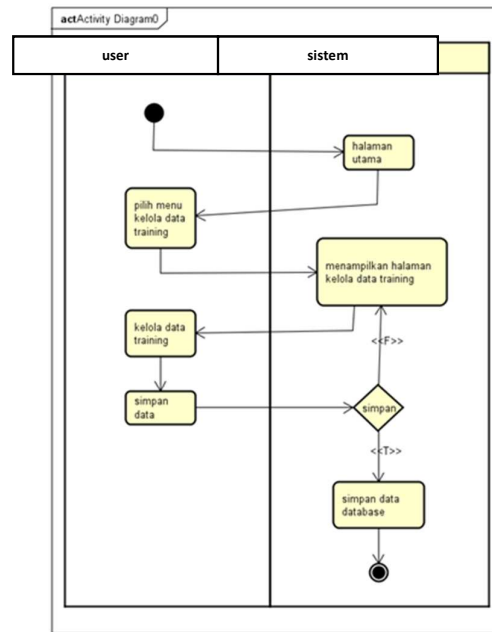
Merupakan aktivitas kelola data *user admin*. Sistem menyediakan fitur tambah, ubah dan hapus yang dapat dipilih oleh *admin* untuk kelola data *user*.



Gambar 3. 19 Activity Diagram Kelola Data User

3. Kelola Data *Training*

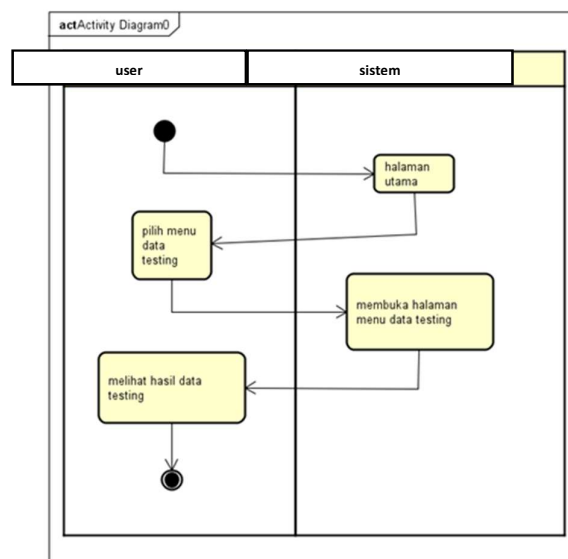
Merupakan aktivitas kelola data *training admin*. Sistem menyediakan fitur tambah, ubah dan hapus yang dapat dipilih oleh *admin* untuk kelola data *training*.



Gambar 3. 20 Activity Diagram Kelola Data Training

4. Lihat Data Testing

Merupakan aktivitas lihat data *testing admin*. Berikut adalah *activity diagram* lihat data *testing*



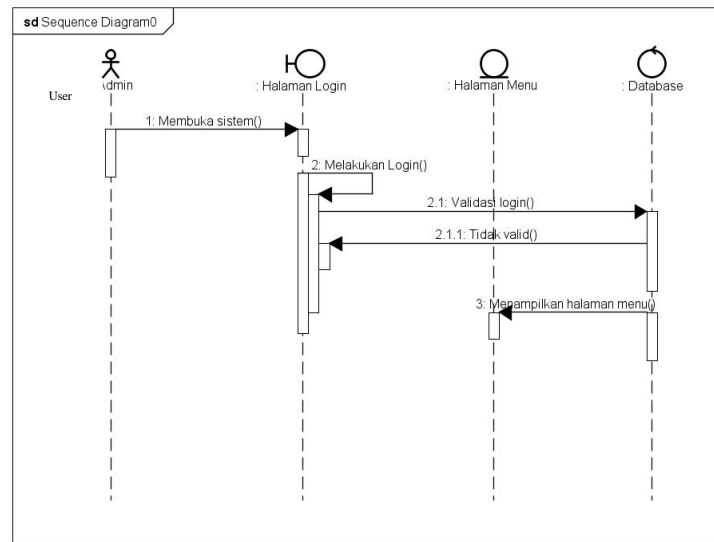
Gambar 3. 21 Activity Diagram Lihat Data Training

1.3.3 Sequence Diagram

Diagram sequence menggambarkan detail alur proses berdasarkan urutan waktu:

a. *Sequence Diagram Login*

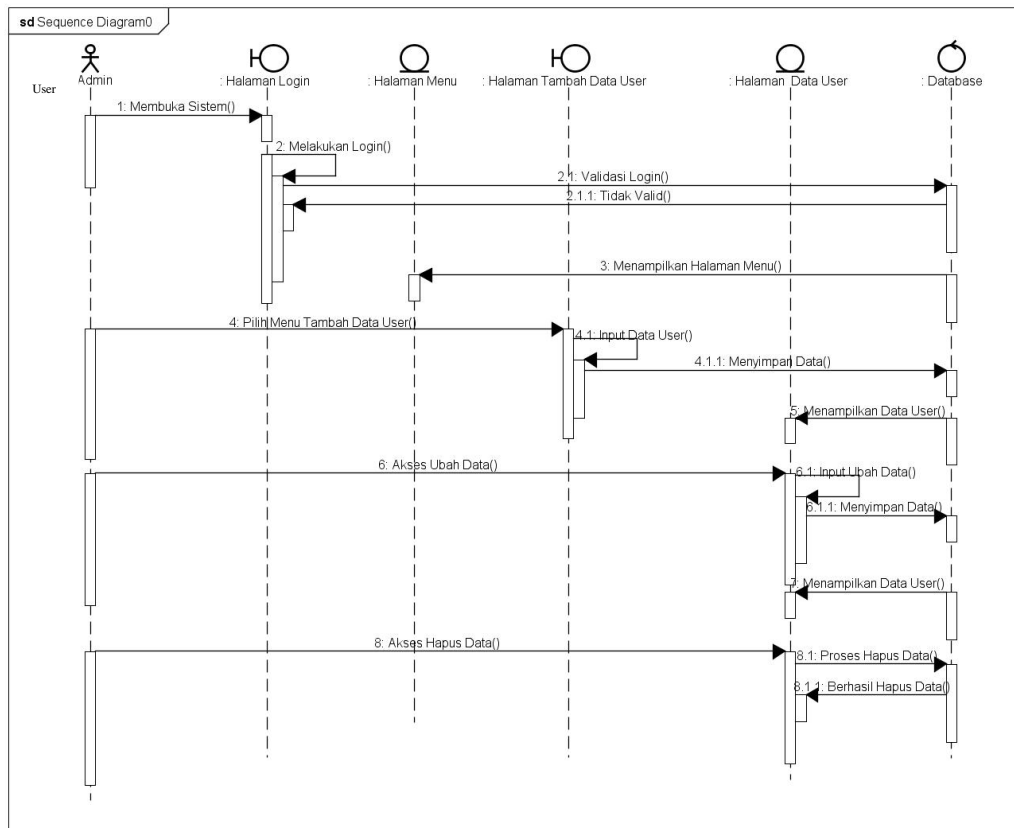
Pada diagram *sequence* ini dapat diketahui urutan proses *login*.



Gambar 3. 22 Sequence Diagram Login

b. *Sequence Diagram Kelola User*

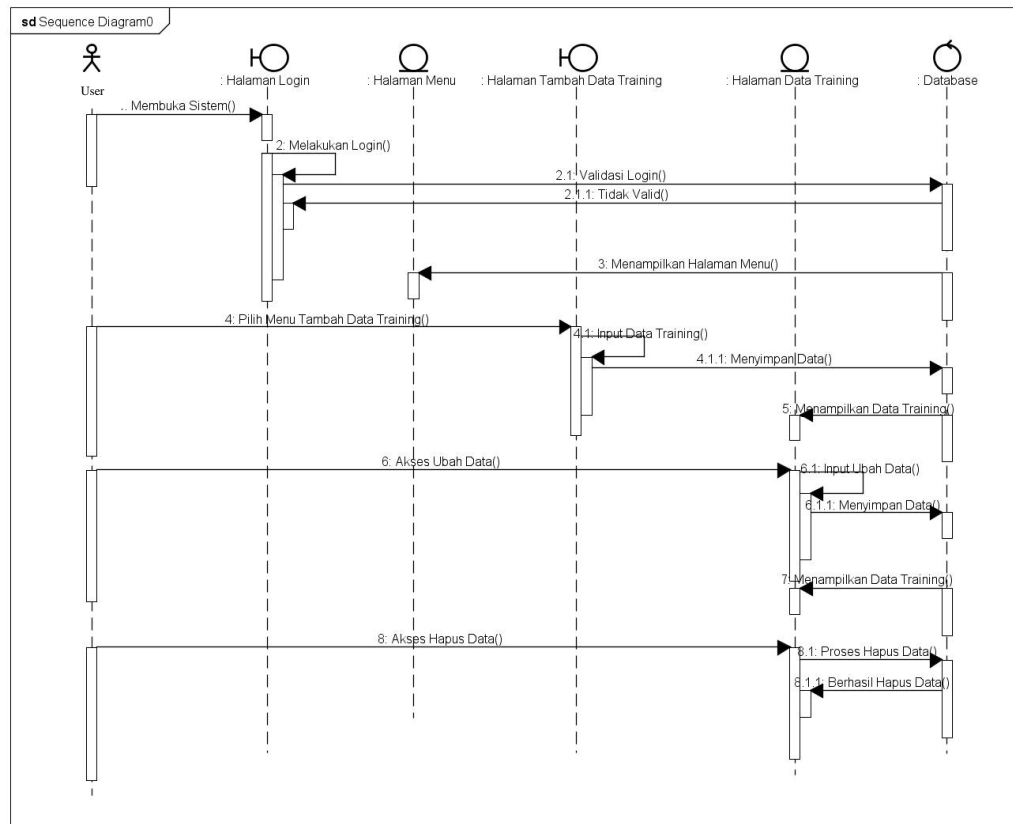
Pada diagram *sequence* ini dapat diketahui urutan proses Kelola user.



Gambar 3. 23 Sequence Diagram Kelola User

c. *Sequence Diagram* Data Training

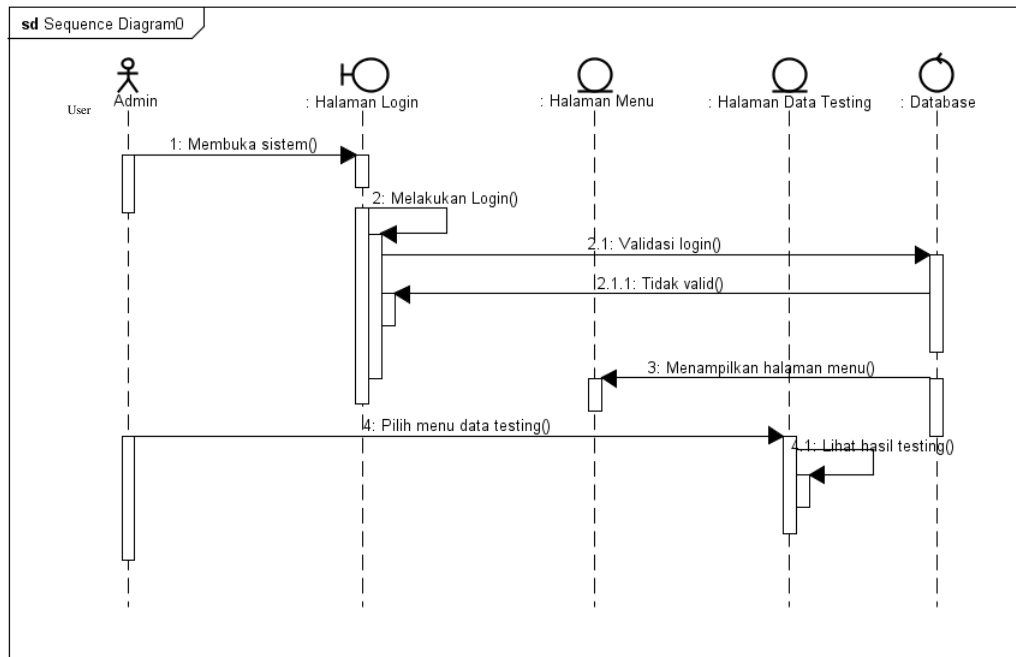
Pada diagram *sequence* ini dapat diketahui urutan proses data training.



Gambar 3. 24 Sequence Diagram Data Training

d. *Sequence Diagram Data Testing*

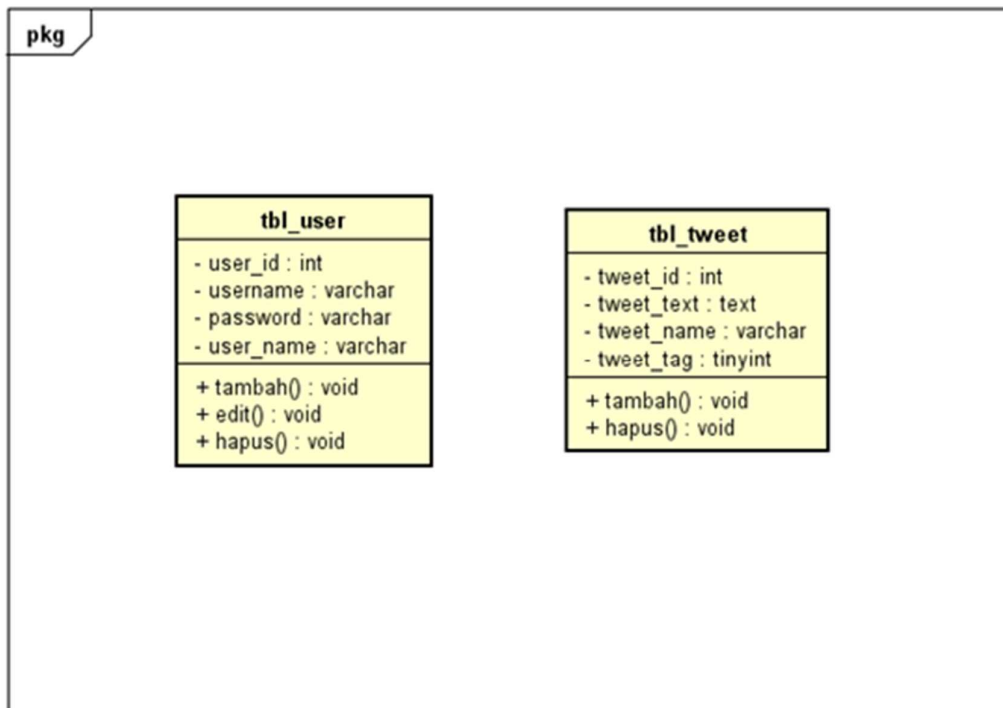
Pada diagram *sequence* ini dapat diketahui urutan proses data testing.



Gambar 3. 25 Sequence Diagram Data Testing

1.3.4 Class Diagram

Berikut ini adalah *class diagram* dari Analisis Sentimen Pada Twitter Terhadap Berita Artis Menggunakan Naïve Bayes.



Gambar 3. 26 Class Diagram