

BAB III

ANALISIS DAN PERANCANGAN

3.1 Analisa

3.1.1 Identifikasi Masalah

Langkah paling awal yang harus dilakukan oleh peneliti, setelah ia memperoleh dan menentukan topik penelitiannya adalah mengidentifikasi permasalahan yang hendak dipelajari. Identifikasi ini dimaksud sebagai penegasan batas-batas permasalahan, sehingga cakupan penelitian tidak keluar dari tujuan. Identifikasi masalah pada penelitian ini adalah :

- 1) Pengklasifikasian sentimen komentar pada facebook daihatsu masih dilakukan secara manual
- 2) Facebook belum bisa mengetahui secara otomatis sentimen dukungan pengguna (pro/kontra) pada komentar di status.

3.1.2 Pemecahan Masalah

Berkaitan dengan identifikasi masalah di atas, maka didapat pemecahan masalahnya yaitu dengan sebuah sistem analisis sentimen untuk menganalisis sentimen terhadap komentar facebook Daihatsu. Pada permasalahan yang akan diteliti, akan dibagi menjadi 3 kategori klasifikasi yaitu positif, netral dan negatif.

Untuk menganalisis sentimen komentar digunakan *Naive Bayes* pada data komentar. Naive Bayes merupakan suatu teknik prediksi dengan menggunakan probabilitas sederhana berdasarkan pengalaman masa lalu dengan asumsi independensi (ketidaktergantungan) yang kuat pada setiap atribut. Jika pada

perhitungan tersebut salah satu kategori (positif, netral, negatif) menghasilkan Vmap tertinggi maka kategori tersebut menjadi suatu hasil klasifikasi

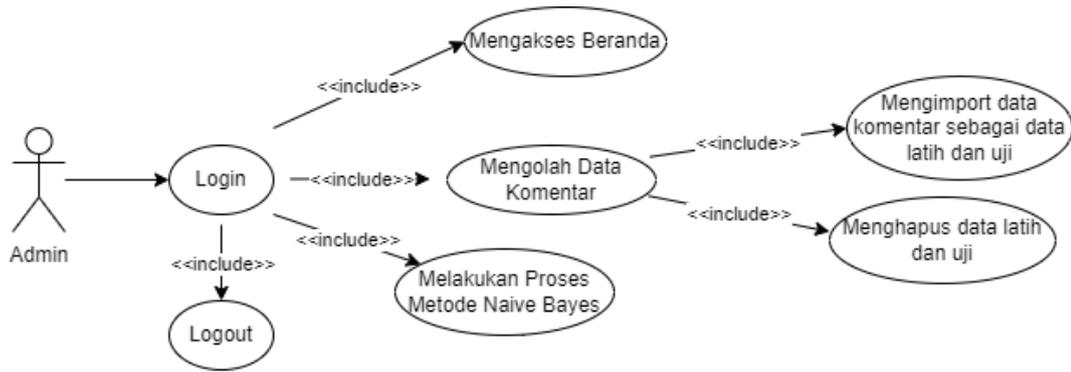
3.2. Perancangan

3.2.1 Perancangan Sistem

Dalam pembahasan ini akan dibahas mengenai perancangan sistem yang dikembangkan pada tugas akhir ini. Tujuan perancangan sistem ini adalah untuk menerapkan *analisis sentiment* komentar pada facebook Daihatsu menggunakan metode *Naïve Bayes*. Sistem yang dikembangkan diharapkan dapat menganalisis sentiment komentar secara otomatis. .

3.2.1.1 Use Case Diagram

Use Case Diagram adalah gambaran grafis dari beberapa atau semua *actor*, *use case*, dan interaksi diantaranya yang memperkenalkan suatu sistem. *Use case diagram* tidak menjelaskan secara detail tentang penggunaan use case, tetapi hanya memberi gambaran singkat hubungan antara use case, aktor, dan sistem. Di dalam use case ini akan diketahui fungsi-fungsi apa saja yang berada pada sistem yang dibuat. Pada gambar 3.1 berikut merupakan Use Case Diagram *Admin* dari aplikasi ini :



Gambar 3.1 Use Case Diagram Admin

Keterangan gambar :

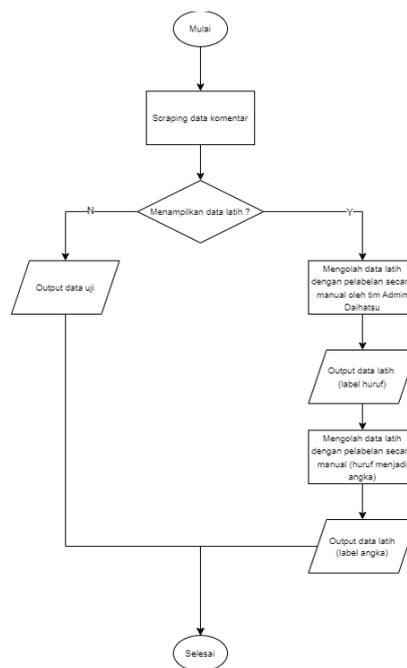
Use Case	Keterangan
Login	: Admin harus melakukan login terlebih dahulu untuk mengakses halaman admin
Beranda	: Admin dapat melihat tampilan awal sistem yang berisi jumlah data latih dan uji pada sistem.
Mengelola Data Komentar	: Admin dapat melihat, menambah dan menghapus data latih maupun data uji
Proses Metode	: Admin dapat melihat proses metode naive bayes dan hasil klasifikasi
Logout	: Admin dapat melakukan Logout

3.2.1.2 Flowchart

Flowchart merupakan salah satu diagram proses untuk menggambarkan, menampilkan langkah-langkah dalam bentuk simbol-simbol grafis, dan urutannya dihubungkan dengan panah. Diagram ini mewakili ilustrasi atau penggambaran penyelesaian masalah yaitu tentang penyelesaian metode *Item-Based Collaborative Filtering*. Berikut ini adalah proses pengklasifikasian komentar menggunakan metode *Naive Bayes* :

a. Pengumpulan Data Komentar

Pada gambar ini adalah proses pengumpulan data latih yang akan dimasukkan kedalam database dan pelabelan komentar yang dilakukan secara manual.



Gambar 3.2 Flowchart Pengumpulan Data Komentar

Data komentar facebook diambil menggunakan dengan proses scraping dengan menggunakan tool dari <https://exportcomments.com/>. Proses Scraping dilakukan dengan memasukkan url link yang ingin kita ambil datanya lalu melakukan export

file menjadi excel. Pada Tabel 3.1 adalah contoh hasil data yang telah melalui proses scraping:

Tabel 3.1 Hasil Scraping Data komentar

No	Komentar
1	Selalu membawa keunggulan dan ketangguhan!
2	Udah irit, luas, mantap dan bagus pokoknya!
3	Info harga dan angsuran?
4	Berapa promo ppnbmnya?
5	Jelek, boros dan mahal!
6	Norak dan nunggu lama!

Setelah dilakukan scraping data komentar, selanjutnya adalah dilakukan pelabelan data komentar untuk data latih secara manual oleh tim admin Daihatsu yang dikategorikan menjadi 3 kategori yaitu positif, negatif dan netral. Pada Tabel ini adalah contoh hasil data yang telah melalui proses pelabelan manual:

Tabel 3.2 Hasil Pelabelan Data Komentar Secara Manual (huruf)

No	Komentar	Label
1	Selalu membawa keunggulan dan ketangguhan!	Positif
2	Udah irit, luas, mantap dan bagus pokoknya!	Positif
3	Info harga dan angsuran?	Netral

4	Berapa promo ppnbmnya?	Netral
5	Jelek, boros dan mahal!	Negatif
6	Norak dan nunggu lama!	Negatif

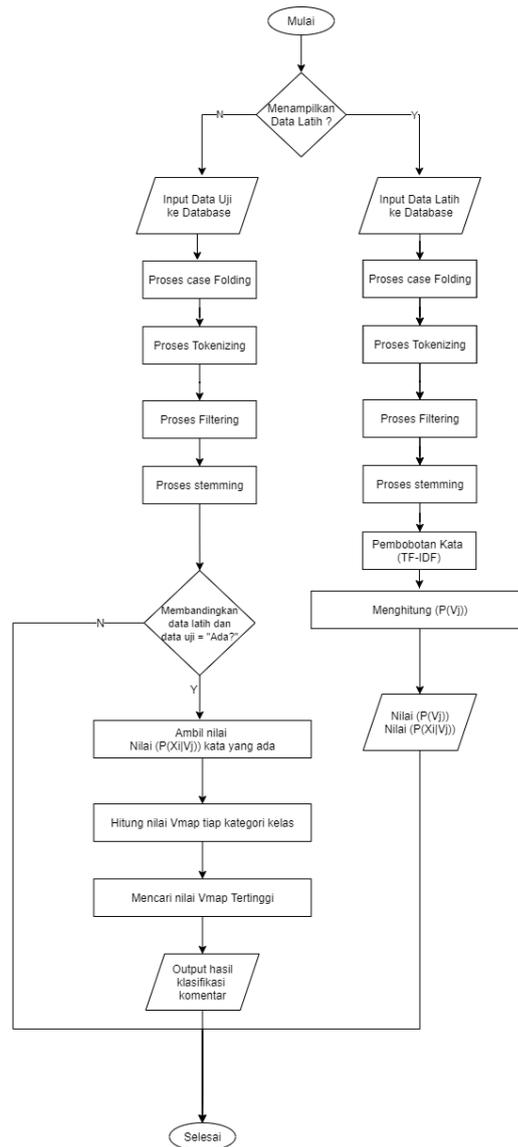
Setelah dilakukan pelabelan (huruf), selanjutnya dilakukan pelabelan menjadi angka oleh tim admin Daihatsu sebelum dilakukan import data ke sistem dengan kategori positif [1], negatif[-1] dan netral [0]. Pada Tabel ini adalah contoh hasil data yang telah melalui proses pelabelan manual:

Tabel 3.3 Hasil Pelabelan Data Komentar Secara Manual (angka)

No	Komentar	Label
1	Selalu membawa keunggulan dan ketangguhan!	1
2	Udah irit, luas, mantap dan bagus pokoknya!	1
3	Info harga dan angsuran?	0
4	Berapa promo ppnbmnya?	0
5	Jelek, boros dan mahal!	-1
6	Norak dan nunggu lama!	-1

b. Input Data dan Perancangan Algoritma Naïve Bayes

Berikut ini merupakan proses input data latih dan uji. Data uji akan melewati proses prediksi berdasarkan data latih. Flowchart proses tersebut bisa dilihat pada gambar 3.3 berikut ini:



Gambar 3.3 Proses Input Data dan Perancangan Algoritma Naïve Bayes

Berikut ini merupakan keterangan dari langkah-langkah proses dari Flowchart diatas :

1. Preprocessing

a) Case Folding

Hal ini dilakukan dengan mengubah kata menjadi lower case atau huruf kecil.

Tabel 3.4 Hasil Proses Preprocessing (Case Folding)

Sebelum	Sesudah
Keunggulan dan ketangguhan!	keunggulan dan ketangguhan!
Irit, luas, mantap dan bagus pokoknya!	irit, luas, mantap dan bagus pokoknya!
Info harga dan angsuran?	info harga dan angsuran?
Berapa promo ppnbmnya?	berapa promo ppnbmnya?
Jelek, boros dan mahal!	jelek, boros dan mahal!
Norak dan nunggu lama!	norak dan nunggu lama!

b) Tokenizing

Tahap Tokenizing merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Seperti karakter whitespace, seperti enter, tabulasi, spasi dianggap sebagai pemisah kata.

Tabel 3.5 Hasil Proses Preprocessing (Tokenizing)

Sebelum	Sesudah
keunggulan dan ketangguhan!	keunggulan dan ketangguhan !
irit, luas, mantap dan bagus pokoknya!	irit , luas , mantap dan bagus pokoknya

	!
info harga dan angsuran?	info harga dan angsuran ?
berapa promo ppnbmnya?	berapa promo ppnbmnya ?
jelek, boros dan mahal!	jelek , boros dan mahal !
norak dan nunggu lama!	norak dan nunggu lama !

c) Filtering

Tahap Filtering pada penelitian ini menggunakan algoritma stopwords adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan bahkan emoji.

Tabel 3.6 Hasil Proses Preprocessing (Filtering)

Sebelum	Sesudah
keunggulan dan ketangguhan !	keunggulan ketangguhan

irit, luas, mantap dan bagus pokoknya !	irit luas mantap bagus pokok
info harga dan angsuran?	info harga angsuran
berapa promo ppnbmnya?	berapa promo ppnbm
jelek, boros dan mahal!	jelek boros mahal
norak dan nunggu lama!	norak nunggu lama

d) Stemming

Stemming adalah proses mencari stem (kata dasar) dari kata hasil filtering. Ada dua aturan dalam melakukan stemming yaitu dengan pendekatan aturan dan pendekatan kamus.

Tabel 3.7 Hasil Proses Preprocessing (Stemming)

Sebelum	Sesudah
keunggulan ketanggaha n	unggul tangguh
irit	irit

luas mantap bagus pokok	luas mantap bagus pokok
info harga angsuran	info lengkap angsuran
berapa promo ppnbnm	berapa promo ppnbnm
jelek boros mahal	jelek boros mahal
norak nunggu lama	norak nunggu lama

2. TF-IDF

Setelah melakukan tahap pre-processing yaitu menggunakan tahapan TF-IDF yang digunakan untuk mengubah data huruf menjadi angka. Berikut adalah tahapan pembobotan kata melalui proses TF-IDF.

a) Term Frequency (TF)

Perhitungan ini dilakukan dengan cara menghitung Term Frequency setiap kata dalam data.

Tabel 3.8 Hasil Perhitungan Term Frequency

Kata	TF-IDF	TF
------	--------	----

	T1	T2	T3	T4	T5	T6	
unggul	1.0 0	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.00
tangguh	1.0 0	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.00
irit	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.00
luas	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.00
mantap	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.00
bagus	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.00
info	0.0 0	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	1.00
harga	0.0 0	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	1.00
angsuran	0.0 0	0.0 0	1.0 0	0.0 0	0.0 0	0.0 0	1.00
berapa	0.0 0	0.0 0	0.0 0	1.0 0	0.0 0	0.0 0	1.00
promo	0.0 0	0.0 0	0.0 0	1.0 0	0.0 0	0.0 0	1.00
ppnbnm	0.0 0	0.0 0	0.0 0	1.0 0	0.0 0	0.0 0	1.00
jelek	0.0 0	0.0 0	0.0 0	0.0 0	1.0 0	0.0 0	1.00
boros	0.0 0	0.0 0	0.0 0	0.0 0	1.0 0	0.0 0	1.00
mahal	0.0 0	0.0 0	0.0 0	0.0 0	1.0 0	0.0 0	1.00
norak	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 0	1.00
nunggu	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 0	1.00
lama	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 0	1.00

b) Invers Document Frequency (IDF)

Setelah di ketahui TF dari kemunculan setiap kata pada tabel 3.8 maka proses selanjutnya yaitu Invers Document Frequency (IDF).

Tabel 3.9 Hasil Perhitungan Invers Document Frequency

Kata	TF	IDF
unggul	1.00	0.77815
tangguh	1.00	1.07918
irit	1.00	1.07918
luas	1.00	1.07918
mantap	1.00	1.07918
bagus	1.00	1.07918
info	1.00	1.07918
harga	1.00	1.07918
angsuran	1.00	1.07918
berapa	1.00	1.07918
promo	1.00	1.07918
ppnbnm	1.00	1.07918
jelek	1.00	1.07918
boros	1.00	1.07918
mahal	1.00	1.07918
norak	1.00	1.07918
nunggu	1.00	1.07918
lama	1.00	1.07918

c) TF-IDF

Setelah melakukan proses pada tabel 3.7 dan 3.8 maka proses selanjutnya yaitu melakukan proses perhitungan TF-IDF.

Tabel 3.10 Hasil Perhitungan TF-IDF

Kata	TF-IDF					
	T1	T2	T3	T4	T5	T6
unggul	0.78	0.00	0.00	0.00	0.00	0.00
tangguh	1.08	0.00	0.00	0.00	0.00	0.00
irit	0.00	1.08	0.00	0.00	0.00	0.00
luas	0.00	1.08	0.00	0.00	0.00	0.00
mantap	0.00	1.08	0.00	0.00	0.00	0.00
bagus	0.00	1.08	0.00	0.00	0.00	0.00
info	0.00	0.00	1.08	0.00	0.00	0.00
harga	0.00	0.00	1.08	0.00	0.00	0.00
angsuran	0.00	0.00	1.08	0.00	0.00	0.00
berapa	0.00	0.00	0.00	1.08	0.00	0.00
promo	0.00	0.00	0.00	1.08	0.00	0.00
ppnbnm	0.00	0.00	0.00	1.08	0.00	0.00
jelek	0.00	0.00	0.00	0.00	1.08	0.00
boros	0.00	0.00	0.00	0.00	1.08	0.00
mahal	0.00	0.00	0.00	0.00	1.08	0.00
norak	0.00	0.00	0.00	0.00	0.00	1.08
nunggu	0.00	0.00	0.00	0.00	0.00	1.08

	0.0					
lama	0	0.00	0.00	0.00	0.00	1.08

3. Menghitung Peluang atau Metode Naïve Bayes

Setelah diketahui TF-IDF nya, selanjutnya adalah perhitungan metode *naïve bayes*. Sebelum sebuah data komentar diklasifikasikan maka system harus mempunyai data latih terlebih dahulu yang kemudian akan menjadi acuan untuk pengklasifikasian data komentar dengan metode *Naive Bayes*. Berikut adalah contoh proses perhitungan metode *Naive Bayes* :

a) Menghitung Nilai Prior

Berikut adalah contoh hasil perhitungan Nilai Prior.

Tabel 3.11 Tabel Nilai Prior

#	Jumlah Dokumen	P(c)
P(c) +	2	0.166666667
P(c)	2	0.166666667
P(c) -	2	0.166666667

b) Pemecahan Term Frequency

Setelah itu komentar akan di pecah perkata untuk dihitung kemunculan kata pada setiap kategori dengan *term frequency* (TF) positif, netral dan negatif

Tabel 3.12 Pemecahan Term Frequency

Kata	DOCN						TF	TF	TF
	T1	T2	T3	T4	T5	T6	(+)	()	(-)
unggul	0.7	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0
	8	0	0	0	0	0	8	0	0

tangguh	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0
irit	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0
luas	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0
mantap	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0
bagus	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0
info	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0
harga	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0
angsuran	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0
berapa	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0
promo	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0
ppnbm	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0
jelek	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	1.0 8
boros	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	1.0 8
mahal	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	0.0 0	1.0 8
norak	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	1.0 8
nunggu	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	1.0 8
lama	0.0 0	0.0 0	0.0 0	0.0 0	0.0 0	1.0 8	0.0 0	0.0 0	1.0 8

c) Peluang Kemunculan Kata Positif, Netral, dan Negatif

Peluang kemunculan kata positif, netral dan negatif yang akan dilambang kan dengan $P(w | +)$, $P(w |)$ dan $P(w | -)$. Berikut dibawah ini terdapat rumus dan tabel 3.13 terdapat tabel hasil perhitungan peluang.

Tabel 3.13 Peluang Kemunculan Kata Positif, Netral, dan Negatif

Kata	$P(W +)$	$P(W)$	$P(W -)$
unggul	0.074	0.041	0.041
tangguh	0.086	0.041	0.041
irit	0.086	0.041	0.041
luas	0.086	0.041	0.041
mantap	0.086	0.041	0.041
bagus	0.086	0.041	0.041
info	0.041	0.085	0.041
harga	0.041	0.085	0.041
angsuran	0.041	0.085	0.041
berapa	0.041	0.085	0.041
promo	0.041	0.085	0.041
ppnbnm	0.041	0.085	0.041
jelek	0.041	0.041	0.085
boros	0.041	0.041	0.085
mahal	0.041	0.041	0.085
norak	0.041	0.041	0.085
nunggu	0.041	0.041	0.085
lama	0.041	0.041	0.085

d) Testing

Kata dengan nilai $P(w | +)$ lebih besar maka akan masuk kedalam kategori kata positif sedangkan nilai $P(w |)$ lebih besar maka termasuk kedalam kategori netral dan nilai $P(w | -)$ lebih besar maka termasuk kedalam kategori negatif. Tahap yang terakhir adalah pengujian komentar apakah termasuk kedalam kategori positif, netral atau negatif. Contoh proses pengujian adalah sebagai berikut :

Tabel 3.14 Contoh Komentar untuk Data Uji

Komentar
Daihatsu SIGRA memang recommended banget buat keluarga Indonesia selain tangguh, irit perawatan dan juga nyaman loh.. Tapi lama datang unitnya..#FansoftheMonth #FansoftheYear

Setelah itu data dilakukan tahap *pre-processing* dan dibandingkan dengan data latih, jika terdapat kata yang sama maka kata tersebut yang akan seperti ini :

Tabel 3.15 Perhitungan Testing

Kata	$P(W +)$	$P(W)$	$P(W -)$
tangguh	0.086	0.041	0.041
irit	0.086	0.041	0.041
lama	0.041	0.041	0.085

e) Menghitung ARGMAX

Menghitung $argmax +$, $argmax$ netral dan $argmax -$ merupakan proses mengklasifikasikan komentar tersebut. Argmax dihitung dengan mengalikan semua peluang kata dengan hasil nilai prior. Karena nilai untuk $argmax +$ lebih besar

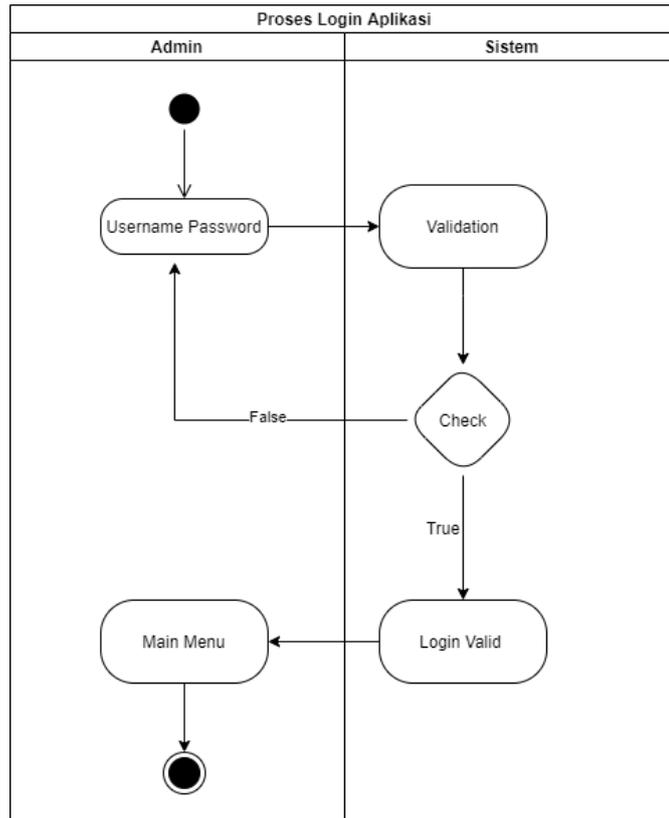
dibandingkan nilai argmax netral dan argmax - maka komentar facebook diklasifikasikan kedalam kategori **Positif**

Tabel 3.16 Hasil Perhitungan Argmax

ARGMAX(+)	ARGMAX()	ARGMAX(-)
0.000051	0.000011	0.000024

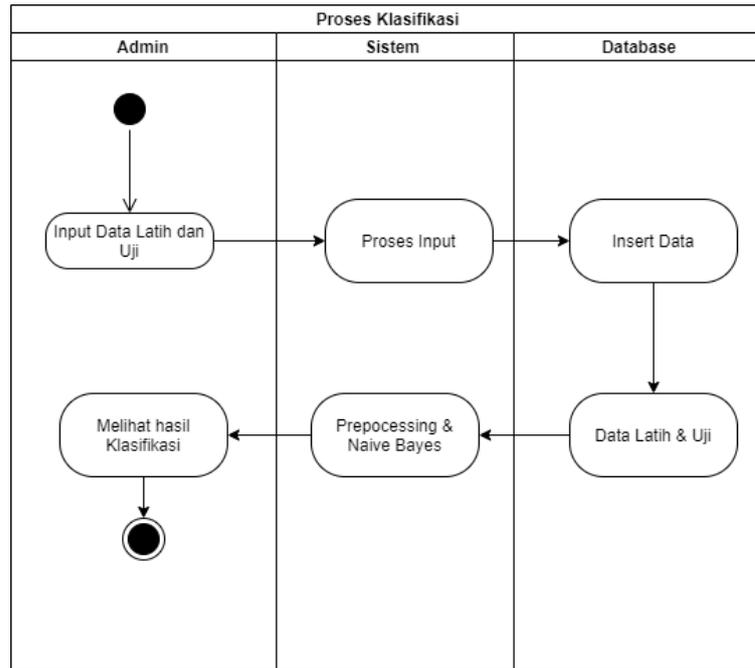
3.2.1.3 Activity Diagram

Activity Diagram merupakan bentuk visual alur kerja yang berisi aktivitas dan tindakan, yang dapat berisi pilihan, atau pengulangan. Dalam *Unified Modeling Language (UML)*, diagram aktivitas dibuat untuk menjelaskan aktivitas komputer maupun alur aktivitas dalam organisasi. Selain itu diagram aktivitas juga menggambarkan alur kontrol secara garis besar. Berikut gambar 3.4 merupakan Activity Diagram dari login aplikasi



Gambar 3.4 Activity Diagram Login aplikasi

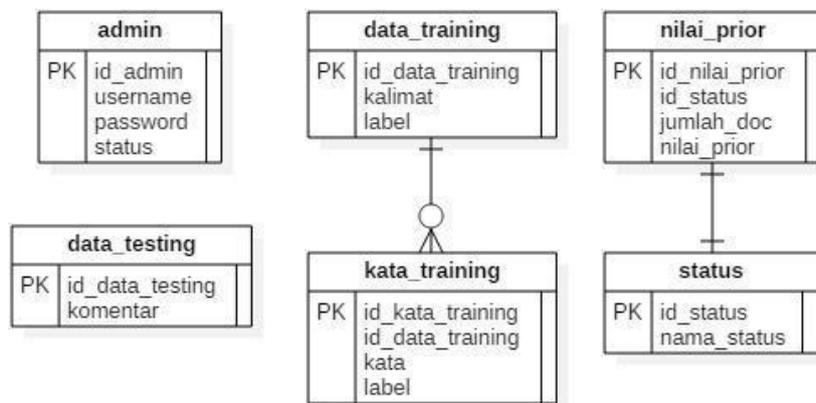
Berikut gambar 3.5 dibawah ini merupakan Activity Diagram dari alur klasifikasi pada sistem :



Gambar 3.5 Activity Diagram Alur Klasifikasi Pada Sistem

3.2.2 Perancangan Data

Perancangan data adalah suatu komponen yang digunakan untuk mendefinisikan struktur data, yang terdiri dari Perancangan *Entity Relationship Diagram (ERD)*. *Entity Relationship Diagram* merupakan model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. pada gambar 3.6 merupakan *Entity Relationship Diagram* dari sistem ini :



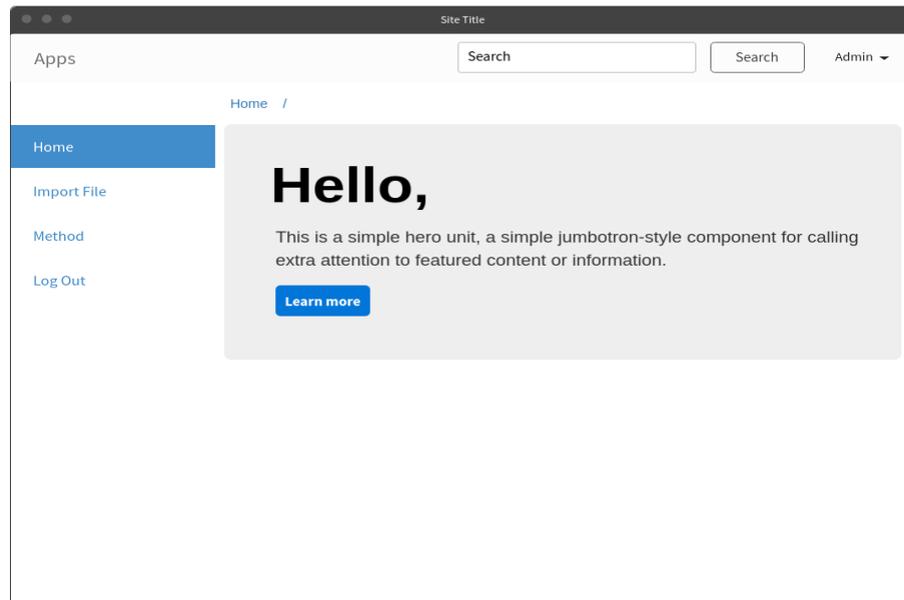
Gambar 3.6 Entity Relationship Diagram dari sistem

3.2.3 Perancangan User Interface / Mock-Up Aplikasi

Perancangan antarmuka (*user interface*) dijelaskan dalam bentuk gambar wireframe yang meliputi tampilan home, import file, dan perhitungan metode

a) Tampilan Home

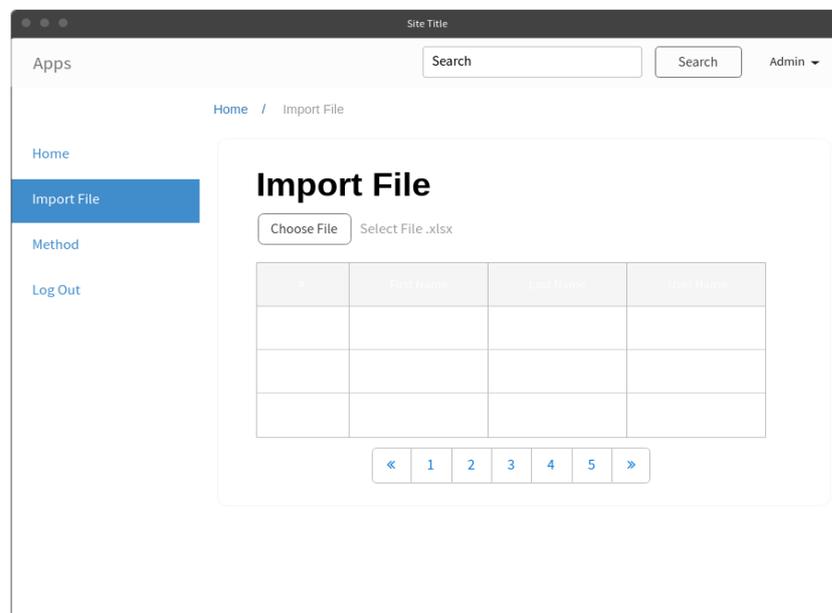
Tampilan home atau dashboard menampilkan petunjuk penggunaan aplikasi agar pengguna dapat memahami penggunaannya



Gambar 3.7 Tampilan Home

b) Tampilan Import File

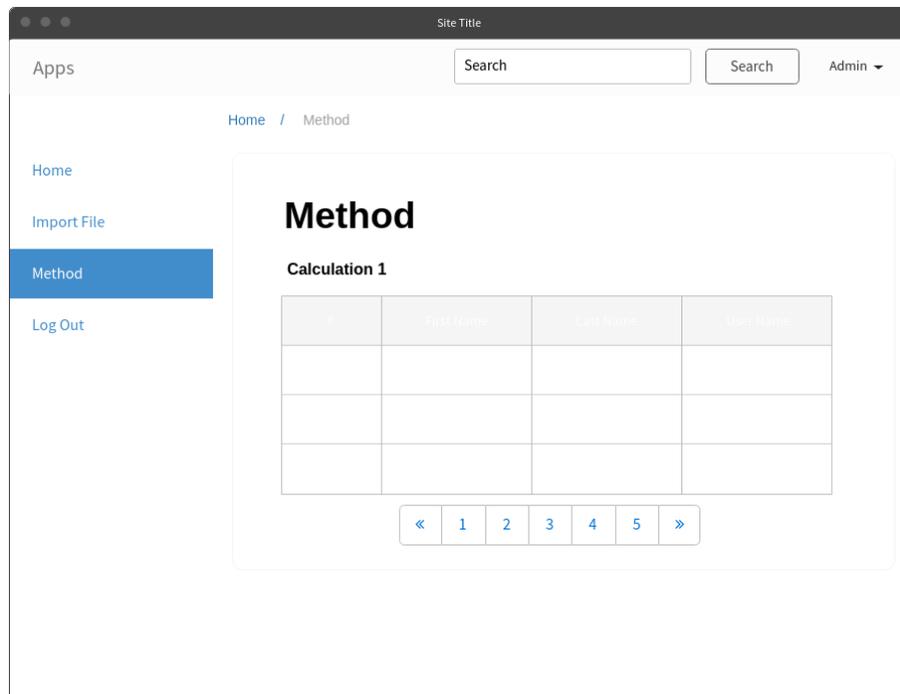
Tampilan ini merupakan tampilan data testing dan training yang telah diimport sebelumnya



Gambar 3.8 Tampilan Import File

c) Tampilan Perhitungan Metode

Tampilan tersebut menampilkan hasil proses perhitungan mulai dari training, testing, kemudian hasil yang didapatkan



Gambar 3.9 Tampilan Perhitungan Metode

3.3 Rancangan pengujian

Setelah selesai dilakukan implementasi, maka selanjutnya akan dilakukan pengujian terhadap sistem yang telah dibuat. Tahap pengujian sangat perlu dilakukan sebagai ukuran bahwa sistem dapat dijalankan sesuai dengan tujuan. Pada pengujian ini ada beberapa hal yang dijadikan acuan, yaitu :

a. Pengujian Sistem

Pengujian sistem dilakukan dengan cara melakukan pengujian *Blackbox* yaitu pengujian yang ditujukan untuk memeriksa apakah fitur berjalan sesuai dengan yang ditentukan.

b. Pengujian Algoritma

Pada proses ini, semua data uji akan diuji agar didapatkan akurasi dengan menggunakan *Confusion Matrix*. Data yang diuji adalah data yang telah diketahui sentimennya, sehingga setelah dilakukan klasifikasi maka akan dilakukan perhitungan akurasi atau ketetapan sistem dalam melakukan klasifikasi.