

BAB 3

METODE PENELITIAN

Penelitian ini adalah penelitian analisis sentimen positif, negatif, dan netral pada data Google Play Store. Data yang diambil adalah data yang berkaitan dengan Shopee, pengambilan data menggunakan API yang disediakan oleh Google Play Store dengan pengklasifikasian menggunakan metode Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). Setelah dilakukan pengambilan data maka akan dilakukan pengolahan data berupa *Preprocessing* untuk mendapatkan hasil yang diinginkan.

Penelitian ini dilatarbelakangi oleh masalah yang ada kemudian mengolah data yang didapatkan sehingga memperoleh data sentimen sesuai dengan yang diharapkan. Berikut ini adalah bahan, alat, dan juga jalan penelitian analisis sentimen Shopee guna menyelesaikan proses analisis sentimen menggunakan data Google Play Store

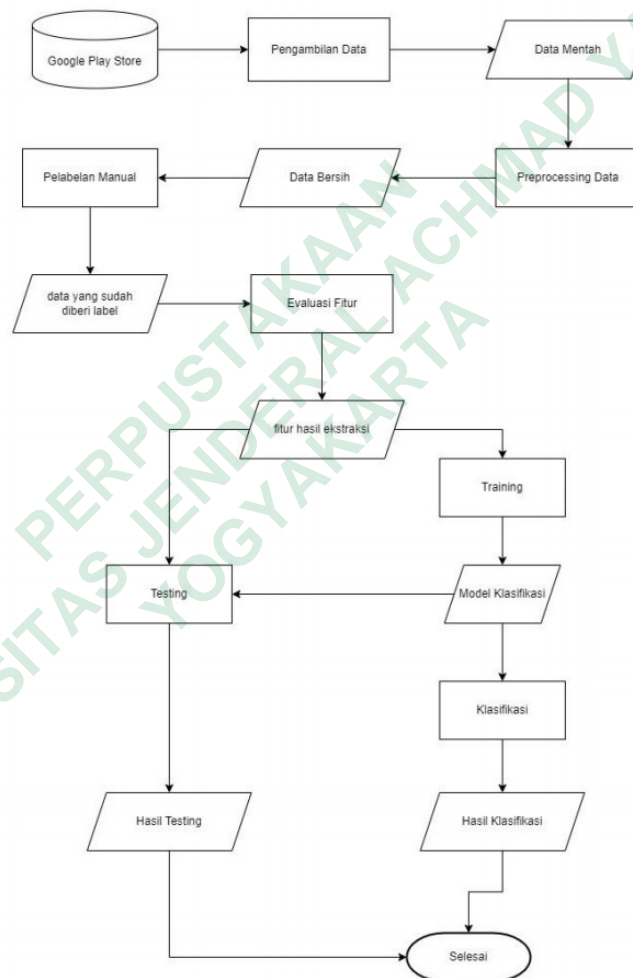
3.1 BAHAN DAN ALAT PENELITIAN

Bahan penelitian yang akan digunakan dalam penelitian ini adalah data *review* di Google Play Store yang berkaitan dengan Shopee. Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi cukup untuk menjalankan sistem operasi dan perangkat lunak pengembangan serta koneksitas Internet. Sistem Operasi dan program-program aplikasi yang dipergunakan dalam dalam pengembangan aplikasi ini adalah:

1. Sistem Operasi: Windows 10 64-bit.
2. Bahasa Pemograman Python 3.10.5.
3. Microsoft Office Excel 2019.
4. Anaconda 3 64-bit.
5. Sublime Text 4
6. Google Colab

3.2 JALAN PENELITIAN

Dalam pengambilan data di Google Play Store, penelitian ini menggunakan bahasa pemrograman Python, Anaconda 3 dan Google Colab. Data akan ditampilkan di Microsoft Office dan dimodelkan dengan bantuan *library* pada bahasa pemrograman Python. Jalan penelitian ini dapat diuraikan pada Gambar 3.1.



Gambar 3.1 Jalan Penelitian

3.2.1 PENGAMBILAN DATA

Pada tahapan ini dilakukan pengambilan dan pengolahan data *review* Shopee pada Google Play Store yang akan dilakukan menggunakan API yang sudah disediakan oleh Google Play Store, kemudian data yang telah

diambil akan ditampilkan ke dalam Microsoft Exel. Hasil dari pengambilan data dapat dilihat pada Tabel 3.1.

Tabel 3.1 Data Review

No	Data Review
1	1,2023-06-11 22:26:59,"TIAP TRANSAKSI LEMOT PARAH, MAU TOP UP LEMOT, MUTER2 BIKIN EMOSI, INI APLIKASI GIMANA SIH KECEWA PARAAAAAH"
2	2,2023-06-14 01:48:49,Aplikasi bobotnya berat banget terlalu menyerap energi hp cepet lelah lemot letoy malehoy
3	1,2023-06-11 13:57:34,"Aplikasinya mengecewakan, paylater dan spinjam saya tiba tiba di block tanpa tau salah saya di mana. Padahal saya selalu bayar sebelum jatuh tempo, sudah saya kirim email untuk pengaduan, tp tetap saja tidak ada solusi apapun."
4	1,2023-06-13 16:45:43,koneksi bagus tapi pas dibuka apknya loading terus ampe berjam jam tolong di benerin
5	5,2023-06-13 15:02:14,praktis banget deh pokok nya
6	5,2023-06-13 14:39:16,Cuma eventnya aja sih yg memuaskan
7	5,2023-06-13 14:35:59,Assalamualaikum salam alhamdulillah mantap aja deh oke makasih
8	5,2023-06-11 00:55:45,Semenjak ada aplikasi Shoppe belanja lebih mudah lebih hemat pokoknya
9	5,2023-06-12 14:22:25,"saya senang belanja apa saja di Shopee sudah tdk ribet, slalu dpat hadiah menarik menarik ."
10	5,2023-06-13 03:03:56,Belanja di shopee barang nya bAgus .. dan sangat puas tolong di pertahan kan harus sesuai photo barang yang di jual biar pelanggan tambah semangat belanja nya ðŸŽš

3.2.2 PREPROCESSING DATA

Tahapan ini akan dilakukan pemrosesan data dengan cara *cleansing* dan *normalization* yang digunakan untuk meningkatkan tingkat akurasi. Data hasil *preprocessing* dapat dilihat pada Tabel 3.3.

Berikut adalah tahapan-tahapan yang dilakukan pada proses *preprocessing*:

1. Case Folding

Tahapan *case folding* merupakan tahapan yang dilakukan untuk mengubah huruf setiap kata pada data menjadi huruf kecil.

2. *Number Removal*

Pada tahapan *number removal* akan dilakukan pembersihan data dari karakter angka pada setiap kalimatnya.

3. *Punctuation Removal*

Pada tahapan *punctuation removal* akan dilakukan pembersihan karakter pada data.

4. *Whitespaces Removal*

Pada tahapan *whitespace removal* dilakukan pembersihan spasi pada awal dan akhir kalimat data *review*.

5. *Tokenize*

Pada tahapan *tokenize* dilakukan pemisahan kata yang akan menjadikan kata-kata tersebut berdiri sendiri dalam suatu teks yang disebut token.

6. *Stopword Removal*

Pada tahapan ini dilakukan penghapusan kata yang memiliki informasi rendah yang terdapat pada kalimat data *review*. List *stopword removal* dapat dilihat pada Tabel 3.2.

Tabel 3.2 Stopword Removal

<p>"b", "d", "xef", "adajjjajajjjjakajajakakkakaaka", "abaca", "by", "bnyak", "yayaan", "yy", "acara", "yy", "zzzz", "x", "xa", "n", "xe", "xf", "xb", "xad", "xd", "xxxxxxx", "xba", "xc", "k", "xcche", "xd xd xaa xd xd xd xd", "m", "t", "xbb", "f", "xbf", "xbd", "xbc", "xab", "xae", "ya", "nya", "a", "c", "f", "e", "g", "h", "i", "j", "k", "l", "m", "n", "o", "p", "q", "r", "s", "t", "u", "v", "w", "x", "y", "z", "blablabla", "mas", "mbak", "isi", "uji", "tulisi", "pasti", "benah", "ajar", "pasti", "kan", "ya", "uji", "tolong", "bab", "bingung", "b", "d", "xef", "x", "xa", "n", "xe"</p>

7. *Normalisasi*

Pada tahapan normalisasi digunakan untuk merubah kata tidak baku atau kata dengan salah eja menjadi baku.

8. Stemming

Pada tahapan *stemming* merupakan tahapan untuk pemetaan dan penguraian bentuk kata menjadi kata dasar. Pada proses *stemming library* yang digunakan yaitu Sastrawi.

Tabel 3.3 *Preprocessing Data*

No	Data Review
1	transaksi lot parah top up lot muter emosi aplikasi gimana sih kecewa paraaaaah
2	aplikasi bobot berat banget serap energi handphone cepet lelah lot letoy malehoy
3	aplikasi kecewa paylater spinjam block salah bayar jatuh tempo kirim email adu solusi apa
4	koneksi bagus pas buka apknya loading ampe berjam jam tolong benerin
5	praktis banget deh pokok nya
6	eventnya sih muas
7	assalamualaikum salam alhamdulillah mantap deh oke makasih
8	semenjak aplikasi shoppe belanja mudah hemat pokok
9	senang belanja shopee tdk ribet slalu dpat hadiah tarik tarik
10	belanja shopee barang nya bagus puas tolong tahan sesuai photo barang jual biar langgan semangat belanja nya

3.2.3 PELABELAN MANUAL

Pada proses pelabelan manual dilakukan pengklasifikasian secara manual setiap kata dan kalimat pada dokumen guna adanya proses analisis lebih lanjut. Label yang akan digunakan pada penelitian ini untuk menentukan hasil yaitu positif, negatif dan netral. Data pelabelan manual dapat dilihat pada Table 3.4.

Tabel 3.4 Data Pelabelan Manual

No	Data Review	Label	Kelas
1	transaksi lot parah top up lot muter emosi aplikasi gimana sih kecewa paraaaaah	Negatif	-1
2	aplikasi bobot berat banget serap energi handphone cepet lelah lot letoy malehoy	Negatif	-1
3	aplikasi kecewa paylater spinjam block salah bayar jatuh tempo kirim email adu solusi apa	Negatif	-1
4	koneksi bagus pas buka apknya loading ampe berjam jam tolong benerin	Negatif	-1

5	praktis banget deh pokok nya	Netral	0
6	eventnya sih muas	Netral	0
7	assalamualaikum salam alhamdulillah mantap deh oke makasih	Netral	0
8	semenjak aplikasi shoppe belanja mudah hemat pokok	Positif	1
9	senang belanja shopee tdk ribet slalu dpat hadiah tarik tarik	Positif	1
10	belanja shopee barang nya bagus puas tolong tahan sesuai photo barang jual biar langgan semangat belanja nya	Positif	1

3.2.4 EKSTRAKSI FITUR

Ekstraksi fitur digunakan pada penelitian ini yaitu menggunakan *Term Frequency-Invers Document Frequency* (TF-IDF). Metode TF-IDF ini bekerja dengan cara menghitung bobot setiap kata yang umum digunakan.

3.2.5 TRAINING DATA

Tahapan ini dilakukan klasifikasi menggunakan metode NBC dan SVM yang digunakan untuk perbandingan, kemudian akan menggunakan TF-IDF *term frequency-invers document frequency* untuk perbandingan setiap kata pada dokumen yang dilakukan secara otomatis.

1. TF-IDF

Term Frequency Invers Document Frequency (TF-IDF) merupakan metode untuk mengkonversi input teks menjadi data numerik sehingga setiap kata atau fitur dapat dinilai. Hasil yang didapatkan dari perhitungan TF-IDF merupakan hasil dari perkalian antara TF dan IDF, dimana TF merupakan jumlah kemunculan kata pada setiap dokumen sedangkan IDF adalah menghitung bobot kata-kata di semua dokumen dalam korpus (Septian, 2019). Untuk perhitungan yang dilakukan yaitu menggunakan lima dokumen, dokumen yang digunakan untuk perhitungan manual dapat dilihat pada Tabel 3.5.

Tabel 3.5 Data TF-IDF

Dokumen (d)	<i>Review</i>
d1	update aplikasi buka live search duh
d2	aplikasi bagus gratis ongkir mantap

d3	pakai ahopee belanja mudah transaksi gampang makasih Shopee good luck
d4	kecewa akun Shopee login jaring masalah jaring bagus kenapaya
d5	aplikasi belanja favorit always pakai Shopee kasih nilai kesini biaya operasional Shopee

Tabel 3.5 merupakan contoh data yang akan dipakai dalam melakukan perhitungan TF-IDF , perhitungan TF-IDF dilakukan secara manual menggunakan 5 dokumen yaitu d1, d2, d3, d4 dan d5. Untuk contoh perhitungan *term frequency* (TF) terdapat pada Tabel 3.6.

Tabel 3.6 Perhitungan TF

term/kata	d1	d2	d3	d4	d5	df
akun				1		1
always					1	1
aplikasi	1	1			1	3
bagus		1	1	1		3
belanja					1	1
biaya					1	1
buka	1					1
duh	1					1
favorit					1	1
gampang			1			1
good			1			1
gratis		1				1
jaring				2		2
kasih					1	1
kecewa				1		1
kenapaya				1		1
kesini					1	1
live	1					1
login				1		1
luck			1			1
makasih			1			1
mantap		1				1
masalah				1		1
mudah			1			1
nilai					1	1

ongkir		1				1
operasional					1	1
pakai			1		1	2
search	1					1
Shopee			2	1	2	5
transaksi			1			1
update	1					1

Tabel 3.6 merupakan perhitungan *term frequency* (TF) yang akan digunakan untuk perhitungan perhitungan sebuah kata yang terdapat pada dokumen dengan N banyaknya dokumen menggunakan komponen *term* atau kata, *document frequency* (DF) dan *invers document frequency* (IDF). Perhitungan IDF untuk persamaan (1) terdapat pada Tabel 3.7.

Tabel 3.7 Perhitungan IDF

term/kata	df	idf
akun	1	0,69897
always	1	0,69897
aplikasi	3	0,221849
bagus	3	0,221849
belanja	1	0,69897
biaya	1	0,69897
buka	1	0,69897
duh	1	0,69897
favorit	1	0,69897
gampang	1	0,69897
good	1	0,69897
gratis	1	0,69897
jaring	2	0,39794
kasih	1	0,69897
kecewa	1	0,69897
kenapaya	1	0,69897
kesini	1	0,69897
live	1	0,69897
login	1	0,69897
luck	1	0,69897
makasih	1	0,69897
mantap	1	0,69897

masalah	1	0,69897
mudah	1	0,69897
nilai	1	0,69897
ongkir	1	0,69897
operasional	1	0,69897
pakai	2	0,39794
search	1	0,69897
Shopee	5	0
transaksi	1	0,69897
update	1	0,69897

Tabel 3.7 merupakan perhitungan *invers document frequency* (IDF) dimana menggunakan persamaan (2) dapat dilihat pada Tabel 3.8.

Tabel 3.8 Perhitungan TF-IDF

Tf*Idf					
term/kata	d1	d2	d3	d4	d5
akun				0,69897	
always					0,69897
aplikasi	0,221849	0,221849			0,221849
bagus		0,221849	0,221849	0,221849	
belanja					0,69897
biaya					0,69897
buka	0,69897				
duh	0,69897				
favorit					0,69897
gampang			0,69897		
good			0,69897		
gratis		0,69897			
jaring				0,79588	
kasih					0,69897
kecewa				0,69897	
kenapaya				0,69897	
kesini					0,69897
live	0,69897				
login				0,69897	
luck			0,69897		
makasih			0,69897		

mantap		0,69897			
masalah				0,69897	
mudah			0,69897		
nilai					0,69897
ongkir		0,69897			
operasional					0,69897
pakai			0,39794		0,39794
search	0,69897				
Shopee			0	0	0
transaksi			0,69897		
update	0,69897				

Tabel 3.8 yaitu merupakan perhitungan hasil dari perkalian TF (*Term Frequency*) dengan IDF (*Invers Document Frequency*). Kemudian terdapat pembuatan model dengan menggunakan algoritma NBC dan SVM dengan 600 data dimana masing-masing dibagi menjadi 200 data positif, 200 data negatif dan 200 data netral.

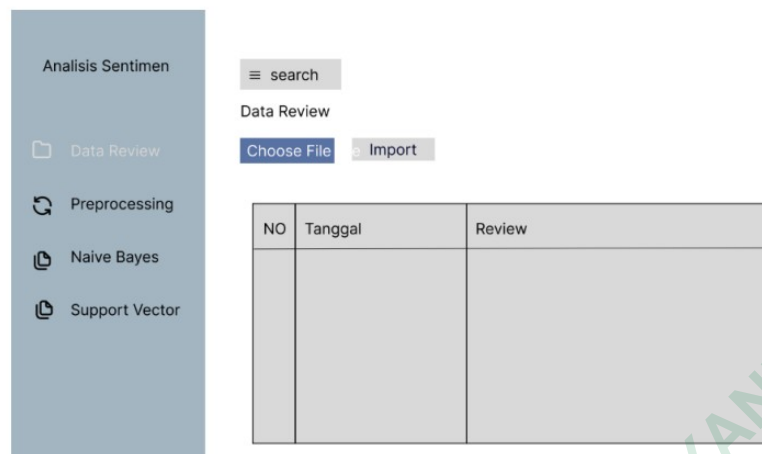
3.2.6 TESTING

Proses *testing* dilakukan guna mengetahui tingkat akurasi yang dihasilkan dari data yang dilabel dan di uji. Berdasarkan data yang diuji akan menggunakan metode NBC dan SVM menghasilkan klasifikasi dengan kategori positif, negatif dan netral. Hasil dari pengklasifikasian ini akan ditampilkan dalam bentuk tabel. Pada proses *testing* data yang digunakan adalah sebanyak 450 data dengan masing masing 150 data positif, 150 data netral dan 150 data negatif.

3.3 DESAIN INTERFACE

3.3.1 HALAMAN DATA REVIEW

Pada halaman data *review* digunakan untuk menambahkan data *review* yang akan digunakan untuk analisis. Dalam halaman ini juga dapat menampilkan data *review* yang akan digunakan. Tampilan halaman data *review* terdapat pada Gambar 3.2.



Gambar 3.2 Halaman *Data Review*

3.3.2 HALAMAN PREPROCESSING DATA

Halaman *preprocessing* data digunakan untuk melakukan processing data yang akan digunakan untuk penelitian. Hasil *preprocessing* sendiri akan digunakan untuk proses selanjutnya, hasil *preprocessing* juga akan ditampilkan pada halaman ini. Tampilan halaman *preprocessing* data dapat dilihat pada Gambar 3.3.

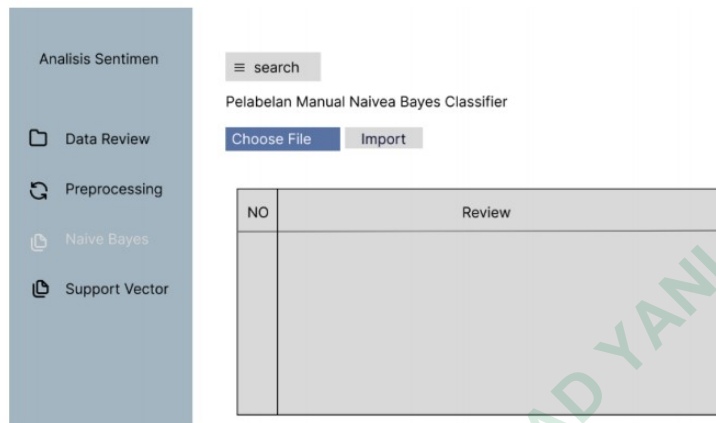


Gambar 3.3 Halaman *Preprocessing Data*

3.3.3 HALAMAN PELABELAN MANUAL

Halaman ini merupakan pelabelan manual yang digunakan untuk menampilkan hasil pelabelan manual. Proses pelabelan manual dilakukan dengan memberi kelas dan label pada data *review* secara manual yang

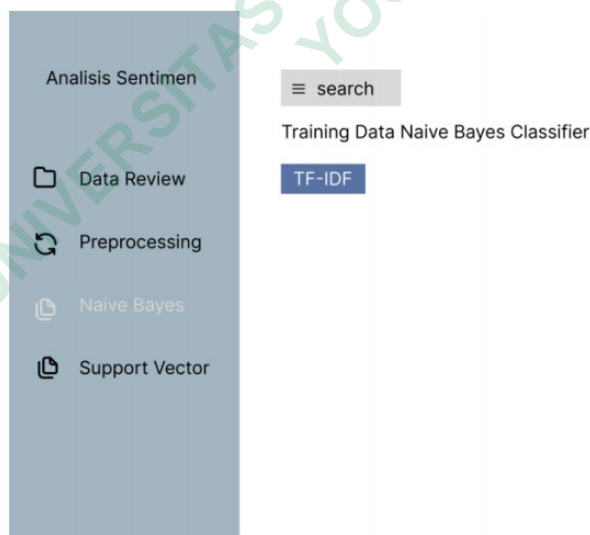
dilakukan pada Microsoft Excel. Tampilan halaman pelabelan manual terdapat pada Gambar 3.4.



Gambar 3.4 Halaman Pelabelan Manual

3.3.4 HALAMAN TRAINING DATA

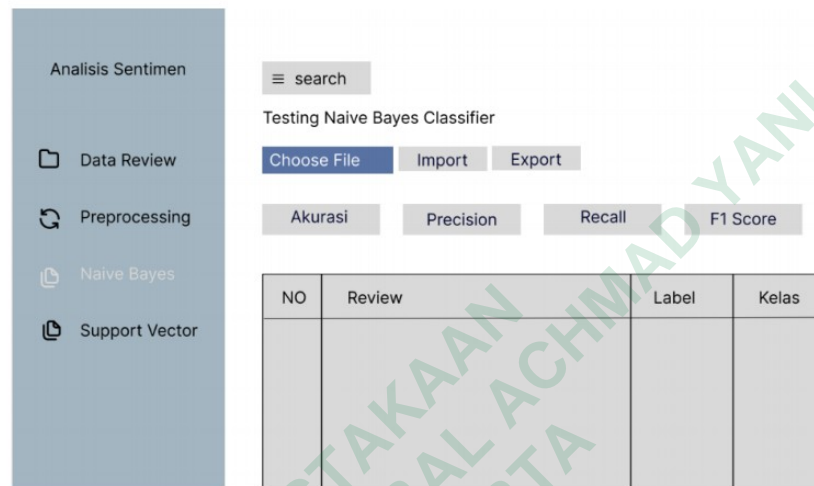
Halaman ini merupakan halaman *training* yang digunakan untuk menampilkan hasil TF-IDF dari data *training*. Proses pada *training* data menggunakan ekstraksi TF-IDF untuk menghasilkan model klasifikasi untuk menampilkan proses klasifikasi sentiment secara otomatis. Tampilan halaman *training* data dapat dilihat pada Gambar 3.5.



Gambar 3.5 Halaman *Training* Data

3.3.5 HALAMAN *TESTING DATA*

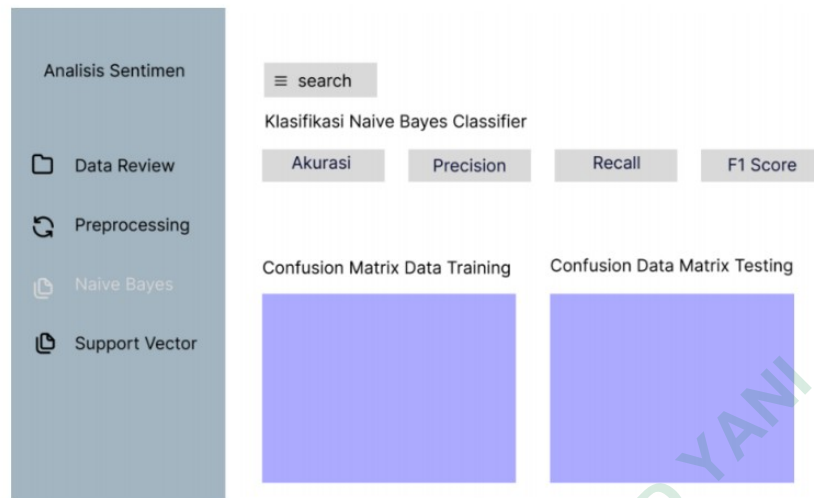
Halaman ini merupakan halaman *testing* yang digunakan untuk menampilkan hasil akurasi, *precision*, *recall* dan *F1 score* dari data. Halaman ini juga menampilkan hasil pelabelan yang dilakukan oleh sistem. Tampilan halaman *testing* data dapat dilihat pada Gambar 3.6.



Gambar 3.6 Halaman *Testing Data*

3.3.6 HALAMAN KLASIFIKASI

Pada halaman klasifikasi digunakan untuk menampilkan hasil dari perhitungan *confusion matrix training* dan *testing*. Proses klasifikasi digunakan untuk memprediksi kelas dan label data *review* yang sudah dibangun pada *training* data. Tampilan halaman klasifikasi dapat dilihat pada Gambar 3.7.



Gambar 3.7 Halaman Klasifikasi