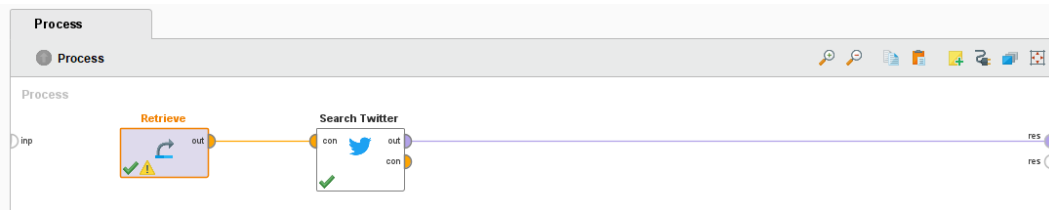


BAB IV HASIL DAN PEMBAHASAN

4.1 Fase Pengumpulan Data

Tahap pertama adalah proses pengumpulan data pada *Twitter* atau disebut *crawling* data. *Crawling* data pada *Twitter* yaitu suatu proses mengunduh data berupa user atau *tweet* dari server *Twitter* dengan bantuan *Application Programming Integration* (API) (Eka Sembodo et al., 2016). Pada proses ini kata kunci yang digunakan yaitu “kenaikan tiket Borobudur” dengan total data sebanyak 591 *tweet*. Proses *crawling* data pada tahapan pengumpulan data dapat dilihat pada gambar 4.1 dan 4.2 berikut ini:



Gambar 4.1 Proses Crawling Data Twitter

Row No.	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1	?	pemerintah te...	?	?	?	?	?	?
2	Jun 14, 2022 ...	detikcom	69183155	?	-1	in	<a href="https...	Menteri PU
3	?	?	?	?	?	?	?	?
4	Jun 20, 2022 ...	mnctrijaya su...	1438732688...	?	-1	in	<a href="http://...	Gubernur .
5	Jun 20, 2022 ...	Binsar Ritonga	1321801888...	?	-1	in	<a href="http://...	Lah drpd n
6	Jun 20, 2022 ...	Eka	3771801192	?	-1	in	<a href="http://...	RT @gald:
7	Jun 19, 2022 ...	@??	1417332125...	?	-1	in	<a href="http://...	RT @Jero
8	Jun 19, 2022 ...	Sugeng Novi...	488563266	amakmal_37	2204919582	in	<a href="http://...	@amakma
9	Jun 19, 2022 ...	Nur Hasanud...	1410146144...	?	-1	in	<a href="http://...	RT @kolor
10	Jun 19, 2022 ...	CSIRT Bulel...	1113798518...	?	-1	in	<a href="https...	Tiket Masu
11	Jun 19, 2022 ...	Achmad N	1487349455...	?	-1	in	<a href="http://...	RT @kolor
12	Jun 19, 2022 ...	Mochall	1488375140...	?	-1	in	<a href="http://...	RT @kolor

Gambar 4.2 Hasil Crawling Data Twitter

4.2 Fase Preprocessing Data

Preprocessing adalah salah satu tahap yang paling penting di dalam *text mining* (Alasadi & Bhaya, 2017) yang digunakan untuk melakukan perubahan data mentah menjadi data yang siap digunakan dan lebih terstruktur dengan pembersihan dan penyeragaman data karena data yang dikumpulkan biasanya masih berupa data kotor. Adapun tahap preprocessing yang diterapkan dalam penelitian ini adalah *cleansing*, *tokenization*, *transform cases*, *stopword removal*, *filter token (by length)*, dan *labelling*.

4.2.1 Cleansing

Cleansing adalah proses yang bertujuan untuk menghilangkan berbagai informasi yang tidak dibutuhkan dalam proses analisis sentimen baik berupa link (*http*, *https*, *pic.twitter*), *hashtag*, *username* (dituliskan *@username*), angka, maupun karakter spesial lainnya untuk memperoleh hasil analisis yang lebih baik seperti pada tabel 4.1 berikut ini.

Tabel 4.1 Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Cleansing*

Proses	Teks
Teks sebelum dilakukan cleansing	RT @Bambang_DP: Setelah melihat video ini, sepertinya tiket masuk candi borobudur menjadi masuk akal. https://t.co/atjgGnUdCT
Teks sesudah dilakukan cleansing	Setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal

Sumber: Hasil Penelitian (2023)

4.2.2 Tokenization

Tokenization adalah proses pemecahan setiap kalimat yang terdapat pada data menjadi potongan-potongan kata. Caranya adalah dengan menjadikan

spasi sebagai acuan untuk memisah setiap katanya. Hasil proses *tokenization* dapat dilihat pada tabel 4.2 berikut ini.

Tabel 4.2 Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Tokenization*

Proses	Teks
Teks sebelum dilakukan <i>tokenization</i>	Setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal
Teks sesudah dilakukan <i>tokenization</i>	Setelah Melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal

Sumber: Hasil Penelitian (2023)

4.2.3 Transform Cases

Transform Cases adalah proses untuk merubah seluruh huruf yang mengandung huruf besar menjadi huruf kecil sehingga dapat berhubungan dengan sentimen. Hasil proses *transform cases* dapat dilihat pada tabel 4.3 berikut ini.

Tabel 4.3 Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Transform Cases*

Proses	Teks
Teks sebelum dilakukan <i>transform cases</i>	Setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal
Teks sesudah dilakukan <i>transform cases</i>	setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal

Sumber: Hasil Penelitian (2023)

4.2.4 Stopword Removal

Pada tahap ini, operator yang digunakan adalah *filter stopwords (dictionary)* karena dataset yang digunakan berbahasa Indonesia. Pada proses ini dimasukkan data daftar kata-kata yang termasuk *stopwords* kemudian *file* nya diupload dalam operator *filter stopwords (dictionary)*. Pada tahap ini, kata-kata yang tidak relevan akan dihapus seperti kata setelah, ini, sepertinya, jadi, merupakan kata-kata yang tidak memiliki makna tersendiri jika dipisahkan dengan kata lain dan tidak terkait dengan kata sifat yang berhubungan sentimen. Hasil dari proses *filter stopwords (dictionary)* dapat dilihat pada tabel 4.4 berikut ini.

Tabel 4.4 Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Filter Stopword (Dictionary)*

Proses	Teks
Teks sebelum dilakukan <i>filter stopwords (dictionary)</i>	setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal
Teks sesudah dilakukan <i>filter stopwords (dictionary)</i>	melihat video tiket masuk candi borobudur menjadi masuk akal

Sumber: Hasil Penelitian (2023)

4.2.5 Filter Token (By Length)

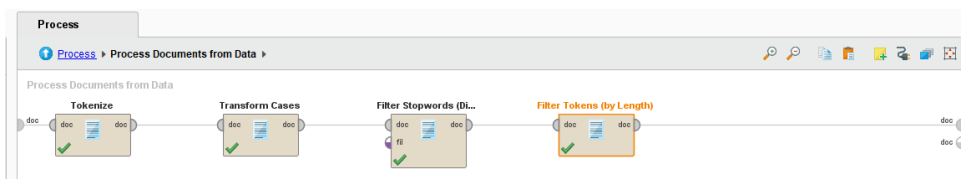
Dalam proses ini, kata-kata yang memiliki huruf panjang kurang dari 4 atau lebih dari 25 akan dihapus, seperti kata “di”, “dan”, “dengan”, “oleh”, dan sebagainya yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen. Hasil dari proses *filter token (by length)* dapat dilihat pada tabel 4.5 berikut ini.

Tabel 4.5 Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Filter Token (By Length)*

Proses	Teks
Teks sebelum dilakukan <i>filter token (by length)</i>	Setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal
Teks sesudah dilakukan <i>filter token (by length)</i>	Setelah melihat video sepertinya tiket masuk candi borobudur jadi menjadi masuk akal

Sumber: Hasil Penelitian (2023)

Tahapan *preprocessing* yang digunakan untuk tahapan awal pengolahan data dapat dilihat pada gambar 4.3 berikut ini:



Gambar 4.3 Tahap *Preprocessing*

Pada gambar 4.3 menjelaskan tentang *preprocessing*, dimana tahapan tersebut terdapat dalam operator *Process Document From Files*. Tahapan *preprocessing* yang digunakan yaitu *tokenize*, *transform cases*, *filter stopwords (dictionary)*, *filter tokens (by length)*.

4.2.6 Labelling

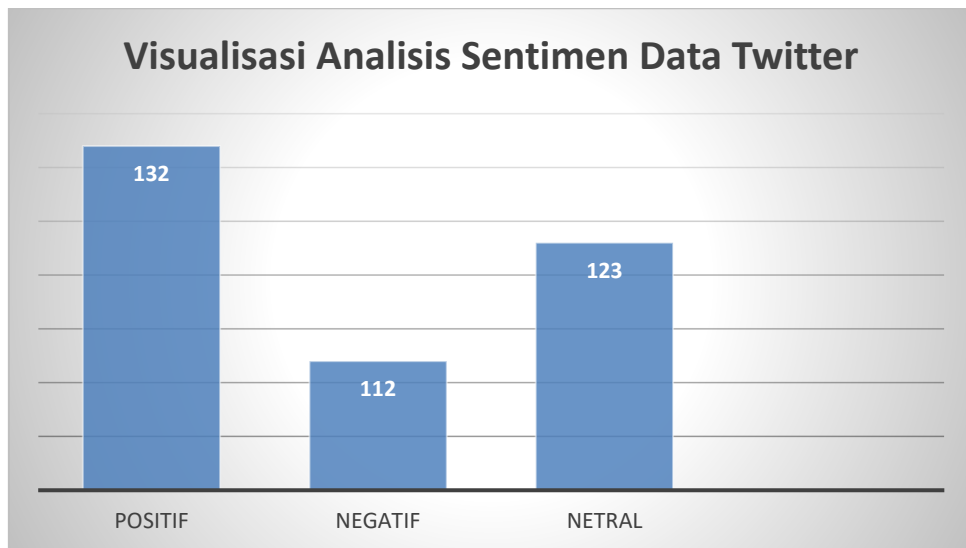
Labelling merupakan pemrosesan terhadap hasil data setelah dilakukan *cleansing*, dimana hasilnya diberikan perhitungan pelabelan dari komentar yang diambil, sehingga mendapatkan klasifikasi yaitu label positif, negatif, dan label netral. Dari 591 *tweet*, setelah dilakukan *preprocessing* data dapat diperoleh 367 data *twitter*. Berikut ini adalah hasil pelabelan dari 367 data *twitter* yang dapat dilihat pada tabel 4.6 berikut ini.

Tabel 4.6 Data Tweet yang sudah diberikan Label

Data	Sentimen	Text
1	Positif	Pengumuman tarif tiket Candi Borobudur yang rencananya mencapai ribu kini tidak jadi naik nih guys Meski dibatalkan pemerintah telah menyiapkan sejumlah aturan untuk pengunjung
2	Netral	Menteri PUPR Basuki Hadimuljono menyatakan pemerintah telah memutuskan untuk membatalkan kenaikan tiket naik Candi Borobudur
3	Positif	Sah Tiket Candi Borobudur Batal Naik Segini Harga Terbaru
4	Netral	Gubernur Jawa Tengah Ganjar Pranowo menyatakan Menteri Koordinator Bidang Kemaritiman dan Investasi Luhut Binsar Pandjaitan sepakat menunda rencana kenaikan harga tiket naik ke area stupa di Candi Borobudur Masih ada beberapa langkah sebelum akhirnya diberlakukan kenaikan tarif
...
366	Positif	Nuri menjelaskan bahwa rencana menaikkan harga tiket masuk ke Candi Borobudur menjadi hal yang menggemparkan masyarakat Rencana penetapan harga tiket ini juga tidak dilandaskan pada penelitian
367	Negatif	Sungguh angkuh pernyataanx coba sya ada dsna tentu sya akan tnya apa kompetensix menetapkan harga tiket masuk borobudur Dst Dsb

Sumber: Hasil Penelitian (2023)

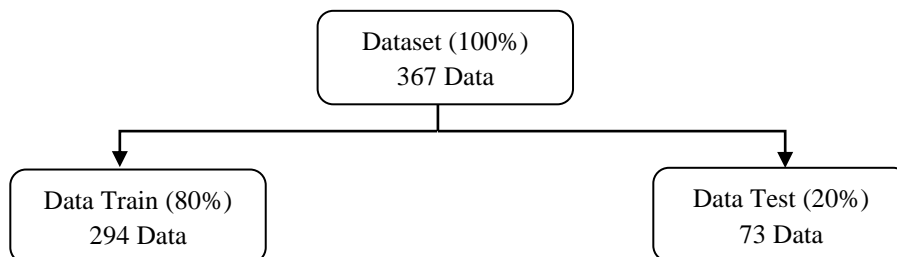
Berikut adalah visualisasi hasil pelabelan dari 367 data *twitter* yang dapat dilihat pada tabel 4.4 dibawah ini.



Gambar 4.4 Visualisasi Analisis Sentimen Data Twitter

4.3 Fase Pembagian Data

Pada tahapan ini akan dilakukan pembagian dataset yang telah memiliki label untuk mendapatkan data *train* dan data *test*. Pembagian dataset menggunakan perbandingan 80% data *train* dan 20% data *test*. Hal ini digunakan karena berdasarkan penelitian sebelumnya hasil perbandingan 80% data *train* dan 20% data *test* mendapatkan hasil yang baik (Juhaeni & Wibowo, 2022; Rifki & Imelda, 2022). Total data *tweet* adalah 367 data. Data ini menghasilkan 294 data *train* dan 73 data *test*. Berikut adalah pembagian data yang dapat dilihat pada gambar 4.5 dibawah ini.



Gambar 4.5 Pembagian Data

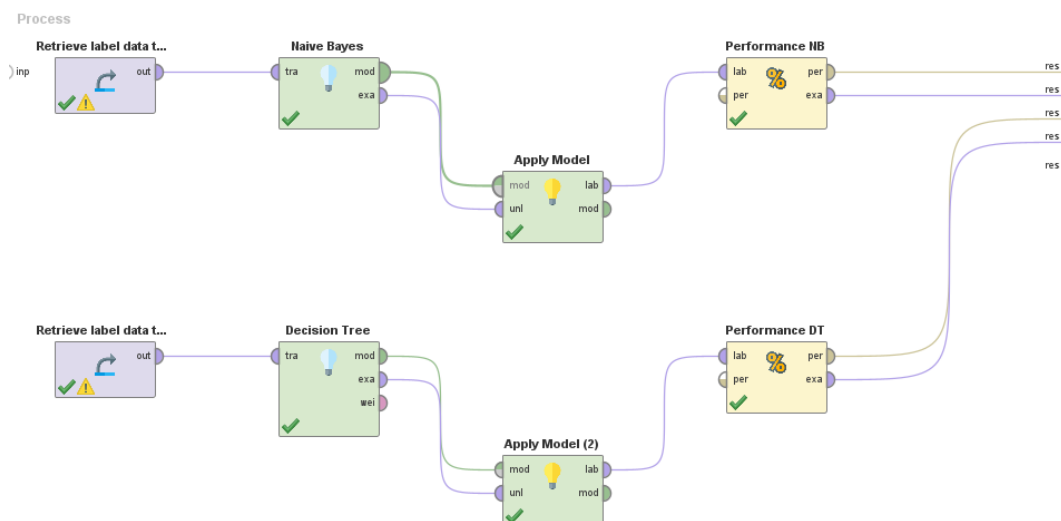
4.4 Fase Pemodelan

Pada tahap ini akan dilakukan pemilihan teknik mining dengan menentukan algoritma yang akan digunakan. *Tool* yang digunakan adalah *RapidMiner* versi 10.1 pada gambar 4.6 dibawah ini.



Gambar 4.6 Rapid Miner Studio Versi 10.1

Hasil pengujian model yang dilakukan adalah mengklasifikasikan sentimen positif, negatif, dan netral menggunakan algoritma *Naive Bayes* dan *Decision Tree*. Berikut adalah desain model *Rapidminer* yang digunakan pada gambar 4.7 dibawah ini:



Gambar 4.7 Desain Model Perbandingan Algoritma

4.5 Fase Pengujian Model

Pengujian merupakan salah satu proses yang penting untuk mengevaluasi dan menganalisis seberapa besar tingkat akurasi. Pada penelitian ini menggunakan pengujian pada akurasi, presisi dan recall. Data yang digunakan untuk melakukan pengujian sebanyak 20% data *test* sebesar 73 data dan 80% data *train* sebesar 294 data dari total keseluruhan data dari 367 data.

4.5.1 Pengujian Data Test

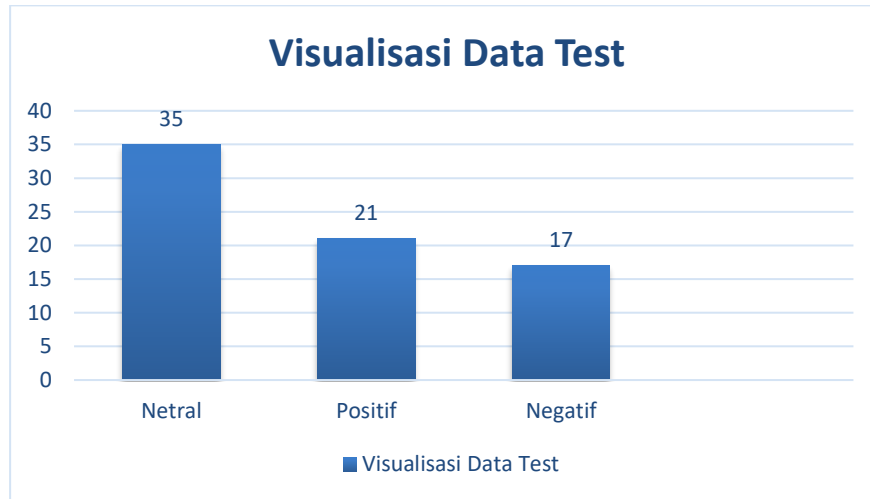
Pada tabel dibawah ini merupakan hasil pengujian data yang telah dibuat pada tabel 4.7 berikut.

Tabel 4.7 Data Test Hasil Pengujian

Data	Sentimen	Text
1	Positif	Pengumuman tarif tiket Candi Borobudur yang rencananya mencapai ribu kini tidak jadi naik nih guys Meski dibatalkan pemerintah telah menyiapkan sejumlah aturan untuk pengunjung
2	Netral	Menteri PUPR Basuki Hadimuljono menyatakan pemerintah telah memutuskan untuk membatalkan kenaikan tiket naik Candi Borobudur
3	Positif	Sah Tiket Candi Borobudur Batal Naik Segini Harga Terbaru
4	Netral	Gubernur Jawa Tengah Ganjar Pranowo menyatakan Menteri Koordinator Bidang Kemaritiman dan Investasi Luhut Binsar Pandjaitan sepakat menunda rencana kenaikan harga tiket naik ke area stupa di Candi Borobudur Masih ada beberapa langkah sebelum akhirnya diberlakukan kenaikan tarif
...
72	Positif	Harga tiket masuk ke Candi Borobudur sebenarnya bukan untuk mencari keuntungan dari biaya perawatan Tapi membatasi atau mengendalikan pengunjung agar tidak merusak situs
73	Netral	woncheolhan haha untung dah pernah

Sumber: Hasil Penelitian (2023)

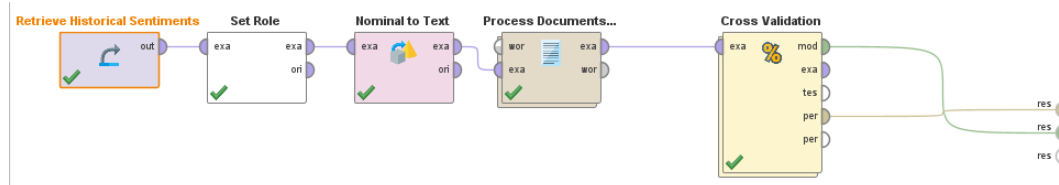
Berikut adalah visualisasi data *test* yang dapat dilihat pada tabel 4.4 dibawah ini.



Gambar 4.8 Visualisasi Data Test

4.5.1.1 Hasil Pengujian Model *Naïve Bayes*

Berikut ini desain proses pengujian model metode *Naive Bayes* yang digunakan dapat dilihat pada gambar 4.9 dibawah ini:



Gambar 4.9 Pengujian Model *Naive Bayes*

Berdasarkan hasil pengujian diatas dapat dilihat hasilnya pada tabel 4.8 dengan *confusion matrix* sebagai berikut:

Tabel 4.8 *Confusion Matrix Naive Bayes*

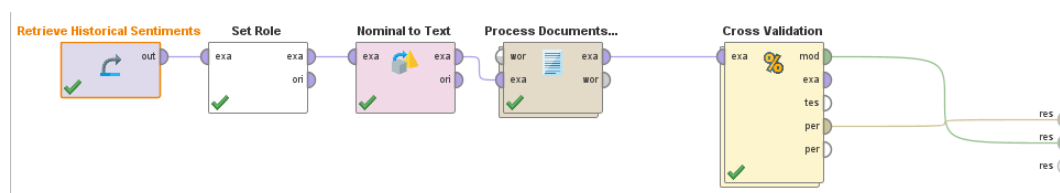
Accuracy: 61.07% +/- 22.09% (micro average: 61.64%)				
	True Positif	True Netral	True Negatif	Class Precision
Pred. Positif	9	6	3	50.00%
Pred. Netral	7	28	6	68.29%
Pred. Negatif	5	1	8	57.14%
Class Recall	42.86%	80.00%	47.06%	

Sumber: Hasil Penelitian (2023)

Berdasarkan pada tabel dengan *confusion matrix* diatas dengan menggunakan metode *naive bayes* untuk mengecek data *test* hasil yang di dapatkan yaitu tingkat akurasi 61.07% +/- 22.09% (micro average: 61.64%). Data kelas positif yang sesuai prediksi positif yaitu 9 data, data positif yang diprediksi netral yaitu 7 data, dan data positif yang diprediksi negatif yaitu 5 data. Data kelas netral yang sesuai ke dalam prediksi netral yaitu 28 data, data netral yang diprediksi positif yaitu 6 data, dan data netral yang diprediksi negatif yaitu 1 data. Data kelas negatif yang sesuai ke dalam prediksi negatif yaitu 8 data, data negatif yang diprediksi positif yaitu 3 data, dan data negatif yang diprediksi netral yaitu 6 data.

4.5.1.2 Hasil Pengujian Model *Decision Tree*

Berikut ini desain proses pengujian model metode *Decision Tree* yang digunakan yaitu pada gambar 4.10 dibawah ini:



Gambar 4.10 Pengujian Model *Decision Tree*

Berdasarkan hasil pengujian diatas dapat dilihat hasilnya pada tabel 4.9 dengan *confusion matrix* sebagai berikut:

Tabel 4.9 *Confusion Matrix Decision Tree*

Accuracy: 45.18% +/- 8.75% (micro average: 45.21%)				
	True Positif	True Netral	True Negatif	Class Precision
Pred. Positif	2	4	0	33.33%
Pred. Netral	18	31	17	46.97%
Pred. Negatif	1	0	0	0.00%
Class Recall	9.52%	88.57%	0.00%	

Sumber: Hasil Penelitian (2023)

Berdasarkan pada tabel dengan *confusion matrix* diatas dengan menggunakan metode *decision tree* untuk mengecek data *test* hasil yang di dapatkan yaitu tingkat akurasi 45.18% +/- 8.75% (micro average: 45.21%). Data kelas positif yang sesuai prediksi positif yaitu 2 data, data positif yang diprediksi netral yaitu 18 data, dan data positif yang diprediksi negatif yaitu 1 data. Data kelas netral yang sesuai ke dalam prediksi netral yaitu 31 data, data netral yang diprediksi positif yaitu 4 data, dan data netral yang diprediksi negatif yaitu 0 data. Data kelas negatif yang sesuai ke dalam prediksi negatif yaitu 17 data, data negatif yang diprediksi positif yaitu 0 data, dan data negatif yang diprediksi netral yaitu 0 data.

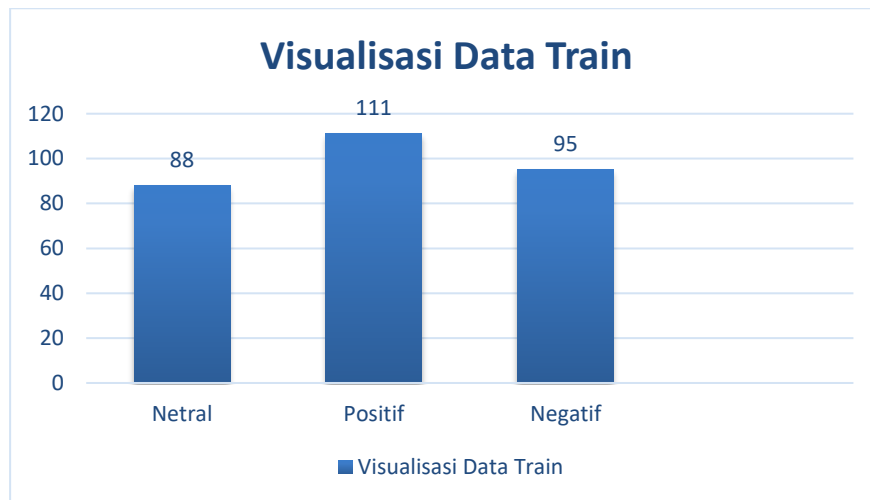
4.5.2 Pengujian Data Train

Pada tabel 4.10 dibawah ini merupakan hasil pengujian data yang telah dibuat.

Tabel 4.10 Data *Train* Hasil Pengujian

Data	Sentimen	Text
1	Netral	Sejalan dgn Protes Rencana Kenaikan Harga Tiket naik ke Candi...
2	Positif	Soal Harga Tiket Candi Borobudur Luhut Pejabat Negara Jangan Over Reacted
3	Netral	Mikir
4	Positif	Pantesan tiket candi borobudur naik
...
293	Positif	Nuri menjelaskan bahwa rencana menaikkan harga tiket masuk ke Candi Borobudur menjadi hal yang menggemparkan masyarakat Rencana penetapan harga tiket ini juga tidak dilandaskan pada penelitian
294	Negatif	Sungguh angkuh pernyataanx coba sya ada dsna tentu sya akan tnya apa kompetensix menetapkan harga tiket masuk borobudur

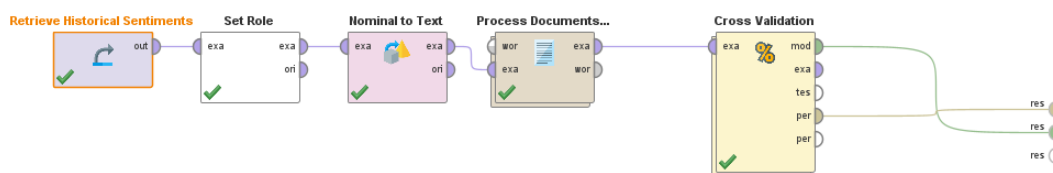
Sumber: Hasil Penelitian (2023)



Gambar 4.11 Visualisasi Data Train

4.5.2.1 Hasil Pengujian Model *Naïve Bayes*

Berikut ini desain proses pengujian model metode *Naive Bayes* yang digunakan yaitu pada gambar 4.12 dibawah ini:



Gambar 4.12 Pengujian Model *Naive Bayes*

Berdasarkan hasil pengujian diatas dapat dilihat hasilnya pada tabel 4.11 dengan *confusion matrix* sebagai berikut:

Tabel 4.11 *Confusion Matrix Naive Bayes*

Accuracy: 50.32% +/- 5.70% (micro average: 50.34%)				
	True Netral	True Positif	True Negatif	Class Precision
Pred. Netral	50	37	21	46.30%
True Positif	25	49	25	49.49%
Pred. Negatif	13	25	49	56.32%
Class Recall	56.82%	44.14%	51.58%	

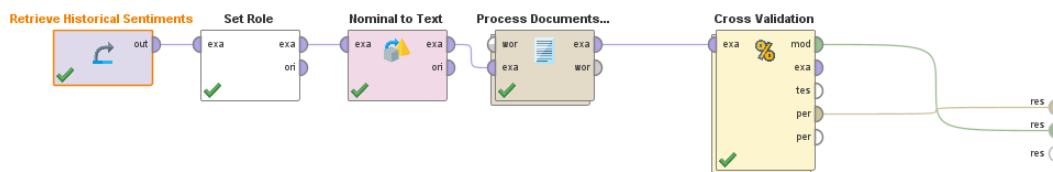
Sumber: Hasil Penelitian (2023)

Berdasarkan pada tabel dengan *confusion matrix* diatas dengan menggunakan metode *naive bayes* untuk mengecek data *train* hasil yang di dapatkan yaitu

tingkat akurasi 50.32% +/- 5.70% (micro average: 50.34%). Data kelas netral yang sesuai prediksi netral yaitu 50 data, data netral yang diprediksi positif yaitu 25 data, dan data netral yang diprediksi negatif yaitu 13 data. Data kelas positif yang sesuai ke dalam prediksi positif yaitu 49 data, data positif yang diprediksi netral yaitu 37 data, dan data positif yang diprediksi negatif yaitu 25 data. Data kelas negatif yang sesuai ke dalam prediksi negatif yaitu 49 data, data negatif yang diprediksi netral yaitu 21 data, dan data negatif yang diprediksi positif yaitu 25 data.

4.5.2.2 Hasil Pengujian Model *Decision Tree*

Berikut ini desain proses pengujian model metode *Decision Tree* yang digunakan yaitu pada gambar 4.13 dibawah ini:



Gambar 4.13 Pengujian Model *Decision Tree*

Berdasarkan hasil pengujian diatas dapat dilihat hasilnya pada tabel 4.12 dengan *confusion matrix* sebagai berikut:

Tabel 4.12 *Confusion Matrix Decision Tree*

Accuracy: 39.08% +/- 5.32% (micro average: 39.12%)				
	True Netral	True Positif	True Negatif	Class Precision
Pred. Netral	1	4	1	16.67%
Pred. Positif	87	103	83	37.73%
Pred. Negatif	0	4	11	73.33%
Class Recall	1.14%	92.79%	11.58%	

Sumber: Hasil Penelitian (2023)

Berdasarkan pada tabel dengan *confusion matrix* diatas dengan menggunakan metode *decision tree* untuk mengecek data *train* hasil yang di dapatkan yaitu tingkat akurasi 39.08% +/- 5.32% (micro average: 39.12%). Data kelas netral yang sesuai prediksi netral yaitu 1 data, data netral yang diprediksi positif yaitu 87

data, dan data netral yang diprediksi negatif yaitu 0 data. Data kelas positif yang sesuai ke dalam prediksi positif yaitu 103 data, data positif yang diprediksi netral yaitu 4 data, dan data positif yang diprediksi negatif yaitu 4 data. Data kelas negatif yang sesuai ke dalam prediksi negatif yaitu 11 data, data negatif yang diprediksi netral yaitu 1 data, dan data negatif yang diprediksi positif yaitu 83 data. Adapun perbandingan hasil komparasi akurasi algoritma yang telah digunakan yaitu pada tabel 4.13 sebagai berikut:

Tabel 4.13 Perbandingan Akurasi Algoritma

Data Testing	
Algoritma	Accuracy
<i>Naïve Bayes</i>	61.07%
<i>Decision Tree</i>	45.18%
Data Training	
Algoritma	Accuracy
<i>Naïve Bayes</i>	50.32%
<i>Decision Tree</i>	39.08%

Sumber: Hasil Penelitian (2023)

Dalam penelitian ini, hasil perhitungan metode *Naïve Bayes* pada data *testing* mendapatkan nilai akurasi 61.64% sedangkan hasil perhitungan *Decision Tree* mendapatkan nilai akurasi 45.18% artinya metode *Naïve Bayes* lebih baik daripada metode *Decision Tree*. Perhitungan metode *Naïve Bayes* pada data *training* mendapatkan nilai akurasi 50.34% sedangkan hasil perhitungan *Decision Tree* mendapatkan nilai akurasi 39.12% artinya metode *Naïve Bayes* lebih baik daripada metode *Decision Tree*. Berdasarkan Tabel 4.13, dapat disimpulkan bahwa akurasi *Naïve Bayes* mendapatkan akurasi yang lebih tinggi dibandingkan dengan *Decision Tree*. Model klasifikasi teks yang digunakan dapat memudahkan untuk mengetahui *tweet* positif, netral, dan negatif. Berdasarkan data *tweet* yang diolah menggunakan *tool Rapidminer*, data *tweet* akan terpisah menjadi kata-kata yang memiliki bobot pada setiap kata-katanya. Kata-kata tersebut akan digunakan untuk melihat kata-kata yang berhubungan dengan sentimen yang sering muncul dan memiliki bobot tertinggi dan dapat digunakan untuk mengetahui *tweet* tersebut positif, netral, ataupun negatif.