

**ANALISIS SENTIMENT PEMILU DI TWITTER DENGAN
MENGUNAKAN ALGORITMA NAÏVE BAYES**



**UNIVERSITAS
BUDI LUHUR**

Nama Kelompok 2:

- Hafiizh Taufiqul Hakim (2012500720)
- Pandu Aryo Utomo (2012500498)

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS BUDI LUHUR

2022

1. PENDAHULUAN

Indonesia merupakan salah satu negara yang menganut sistem demokrasi. Hal ini ditandai dengan diadakannya pemilihan presiden dan wakil presiden setiap periodenya. Seorang tokoh politik yang ingin maju sebagai calon presiden tentu akan melihat atau mempertimbangkan popularitas mereka berdasarkan opini dari masyarakat. Dahulu masyarakat mengungkapkan opini, kritik, dan sarannya melalui media cetak yang tidak semua orang mempunyai kemampuan menulis dan kesempatan menerbitkan tulisannya. Namun, perkembangan teknologi komunikasi saat ini telah merubah kecenderungan kebiasaan masyarakat dalam mengekspresikan opininya pada media sosial. Salah satu media sosial yang populer di kalangan pengguna internet saat ini adalah Twitter. Bisa dikatakan jika media sosial Twitter telah menjadi wadah masyarakat Indonesia untuk berkomunikasi dan menyelurkan pendapat. Terbukti status tweet para pengguna Twitter di Indonesia tidak hanya berisi kehidupan pribadi saja, banyak dari status yang mereka bagikan juga berisi komentar dan opini tentang pemilihan umum.

Sentiment analisis adalah studi komputasional dari opini-opini orang, sentiment melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif.

Penentuan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak terpakai. Oleh karena itu, diajukan penerapan metode pembelajaran mesin untuk mengklasifikasi polaritas opini dari sumber data yang sangat banyak tersebut. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari text mining, dalam hal ini adalah klasifikasi dokumen. Ada beragam teknik klasifikasi dokumen, di antaranya adalah Naïve Bayes classifier, Decision Trees, dan Support Vector Machines. Salah satu metode yang paling populer digunakan dalam pengklasifikasian dokumen sekarang ini adalah metode Naïve Bayes classifier. Metode Naïve Bayes classifier mempunyai kecepatan dan akurasi yang tinggi ketika diaplikasikan dalam basis data yang besar dan data yang beragam. Metode Naïve Bayes Classifier memiliki beberapa kelebihan antara lain, sederhana, cepat dan berakurasi tinggi.

Sehingga dalam menyelesaikan permasalahan tersebut dilakukan “Analisis Sentimen Pemilu di Twitter dengan Menggunakan Algoritma Naive Bayes”, Penelitian ini bertujuan untuk melakukan analisis opini masyarakat terhadap Pemilu dengan cara *opinion mining* pada media sosial Twitter berbasis sentiment analisis.

2. METODOLOGI PENELITIAN

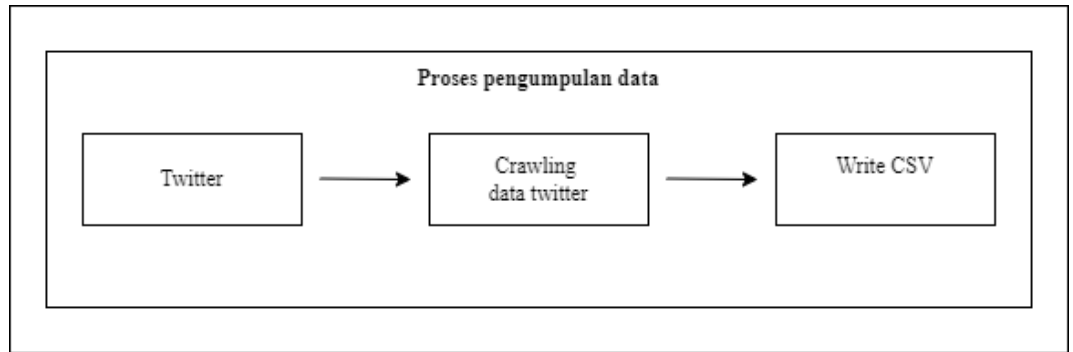
Beberapa metode untuk memperoleh data atau informasi dalam menyelesaikan permasalahan. Metode yang dilakukan tersebut antara lain :

- a) Sumber Data
Twitter menjadi sumber data untuk melakukan analisis sentiment ini, dikarenakan inti masalah yang ingin kita teliti yaitu tentang pemilu, tentunya kita harus menganalisis sentiment berdasarkan opini masyarakat terhadap pemilu tersebut.
- b) Pengumpulan Data
Setelah mendapatkan data berdasarkan pengguna twitter tentang pemilu langkah selanjutnya yaitu mengambil data yang sudah diberikan pengguna twitter dengan cara melakukan crawling tentang pemilu.
- c) Pre-processing
Tools yang digunakan dalam proses ini adalah Rapidminer. Sebelum data diolah menggunakan metode Naïve Bayes. dilakukan proses pre-processing. penelitian ini di terapkan text preprocessing untuk data yang akan di gunakan dalam proses analisis sentiment. Dalam penelitian ini digunakan metode Naïve Bayes.

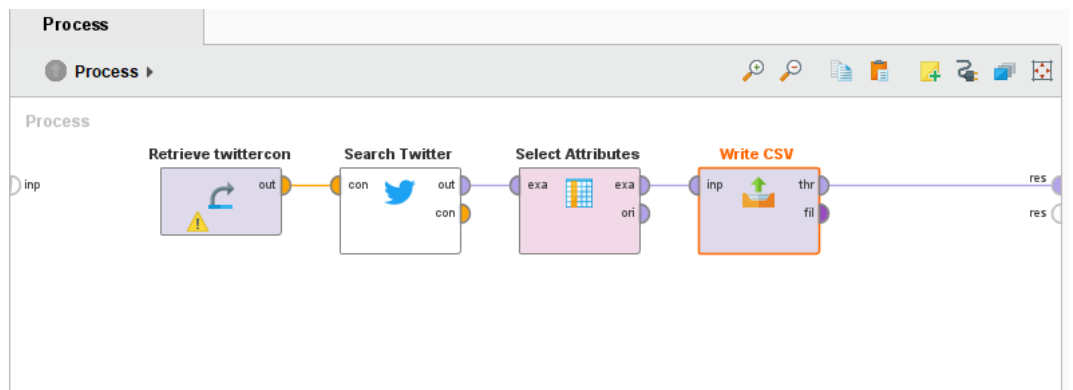
3. HASIL PEMBAHASAN

Dalam penelitian ini, akan dilakukan pengklasifikasian komentar dari twitter. Percobaan ini berguna untuk mengukur akurasi sistem dalam mengklasifikasi komentar. Pada percobaan ini, terdapat beberapa parameter pada evaluasi data. Parameter tersebut antara lain adalah menghitung akurasi. Untuk dapat menghitung nilai akurasi tersebut, dibutuhkan beberapa nilai yang lain seperti True Positif, True Negatif, Pred Positif, Pred Negatif. Hasil eksperimen akan menunjukkan nilai dari setiap evaluasi tersebut. Pada percobaan ini digunakan 199 data yang diperoleh secara acak dari hasil crawling melalui API twitter.

- a) Pengumpulan Data
Dalam penelitian ini data yang diambil adalah komentar para pengguna twitter tentang pemilu, Setelah mendapatkan data tersebut langkah selanjutnya yaitu dilakukan crawling data terhadap data opini dalam bentuk tweet yang diberikan pengguna twitter. Jumlah data yang diambil yaitu 199 komentar, dimana data yang kita proses akan kita ambil informasi yang terkandung didalamnya yaitu komentar negatif atau komentar positif. Agar memudahkan dalam mengelola data maka data perlu kita berikan analisis sentiment secara manual dengan membaca maksud dari kalimat yang ada dalam sentimen tersebut, sehingga dapat diberikan penilaian bahwa sentiment tersebut merupakan komentar negatif atau positif.



Gambar 1 : Proses Pengumpulan data



Gambar 2 : Process Crawling data Twitter

ExampleSet (Search Twitter)

Open in: [Turbo Prep](#) [Auto Model](#)

Filter (199 / 199 examples): all

	From-User-Id	To-User	To-User-Id	Language	Source	Text	Geo-Locatio...	Geo-Lo
es...	17128975	?	-1	in	<a href="http://...	Partai Ummat Sesumbar Bakal Kalahkan...	?	?
ir ...	86012022	?	-1	in	<a href="http://...	HNW, Waket MPR dari @PKSejahtera Tol...	?	?
3n...	120968478	?	-1	in	<a href="http://...	Abjad memperbaiki keadaan ini cukup jel...	?	?
UD...	1075709206...	?	-1	in	<a href="http://...	RT @hipohan: Ini sengaja dibiarkan mas...	?	?
er	1579593577...	?	-1	in	<a href="http://...	Jaga Pemilu 2024 tetap damai dan tentra...	?	?
isel	1532183433...	?	-1	in	<a href="http://...	Hallo #sahabatbawaslu	?	?
	1232520551...	?	-1	in	<a href="http://...	Mencari celah utk menunda pemilu, berb...	?	?
	121935956	?	-1	in	<a href="http://...	RT @keuangannews_id: Terbongkar, Pe...	?	?
jd...	278330740	?	-1	in	<a href="http://...	RT @keuangannews_id: Terbongkar, Pe...	?	?
ank	3032828762	?	-1	in	<a href="http://...	RT @diyoex: LIHAT Tanggal!	?	?
	121935956	?	-1	in	<a href="http://...	RT @hipohan: Ini sengaja dibiarkan mas...	?	?
er	1579593577...	?	-1	in	<a href="http://...	Stop propaganda negatif Pemilu 2024 htt...	?	?
ian...	1253430722	?	-1	in	<a href="http://...	RT @aniesbaswedan: (9/10) Kemarin, L...	?	?
n A...	631006376	k1lua_HxH	1603541839...	in	<a href="http://...	@k1lua_HxH @zainalamochtar Mudah2a...	?	?

ExampleSet (199 examples, 1 special attribute, 11 regular attributes)

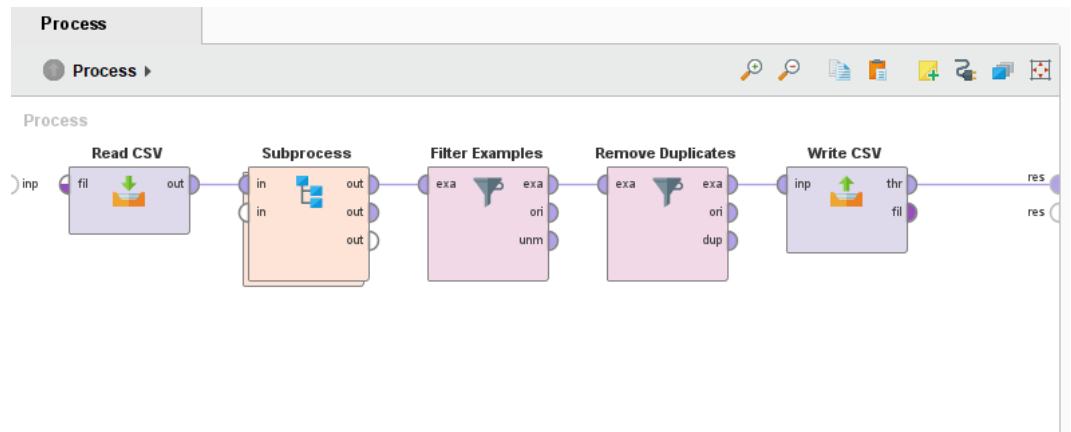
Gambar 3 : Hasil data set dari melakukan crawling dengan menggunakan Rapidminer (Data Kotor)

Keterangan :

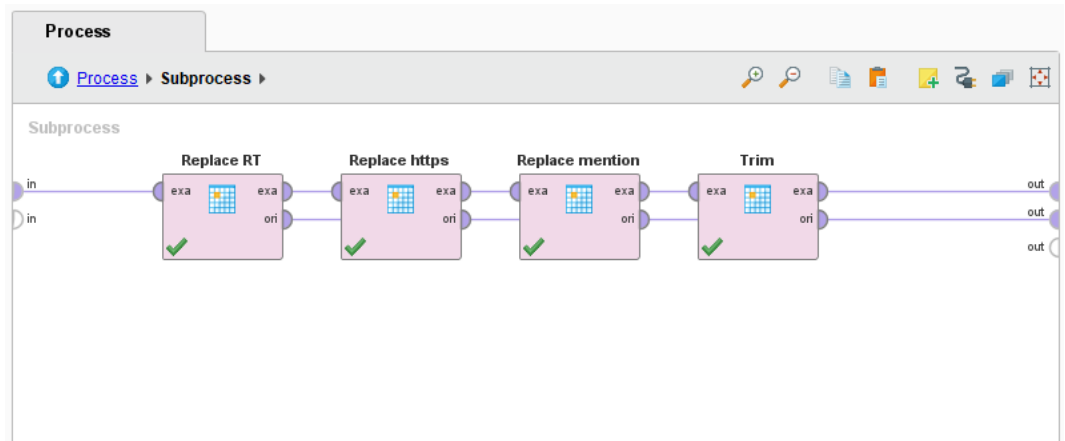
- Retrieve twittercon : melakukan koneksi pada twitter
- Search Twitter : Pencarian komentar untuk crawling data
- Select Attributes : Untuk memilih satu atau sekumpulan data yang memiliki suatu informasi tertentu
- Write CSV : Menyimpan/mengekspor data dengan format csv

b) Data Cleanning

Penelitian ini bertujuan untuk menyeleksi data dan membuang data yang berpotensi mengurangi akurasi machine learning.



Gambar 4 : Process data cleanning



Gambar 5 : subprocess

ExampleSet (Remove Duplicates)		
<div>Open in Turbo Prep Auto Model</div> <div>Filter (86 / 86 examples): all</div>		
Row No.	Text	Id
2	Jangan dikira...	1609469695...
3	KPU meneta...	1609469581...
4	Alhamdulillah...	1609469235...
5	Lolos Jadi Pe...	1609469212...
6	Syukuran Par...	1609468579...
7	Kalau Mema...	1609468393...
8	PKS menyam...	1609468062...
9	Di tahun 202...	1609468059...
10	PAN menyam...	1609468039...
11	Bila pemilu di...	1609468011...
12	Terima kasih ...	1609467936...
13	2023, Fahri H...	1609467598...
14	Bapak kelebi...	1609467410...
15	Wasekjen DP...	1609467262...
16	gue mah bod...	1609467036...

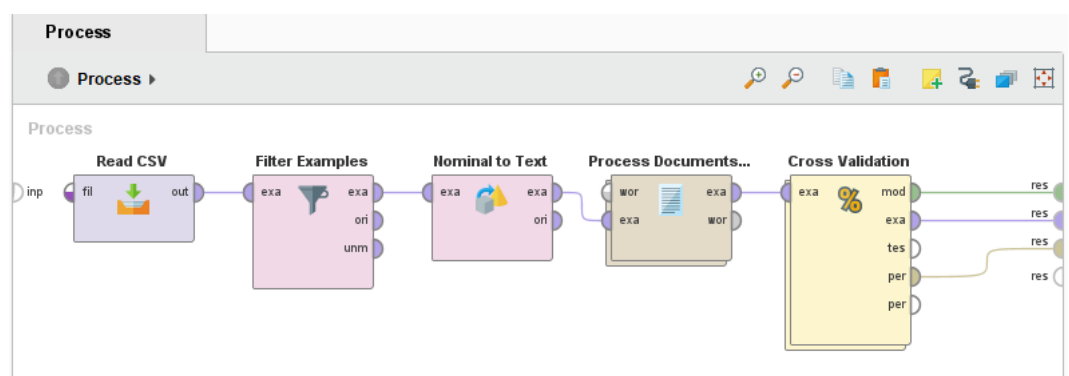
ExampleSet (86 examples, 0 special attributes, 2 regular attributes)

Gambar 6 : Hasil data set yang sudah dicleanning (data bersih)

Keterangan :

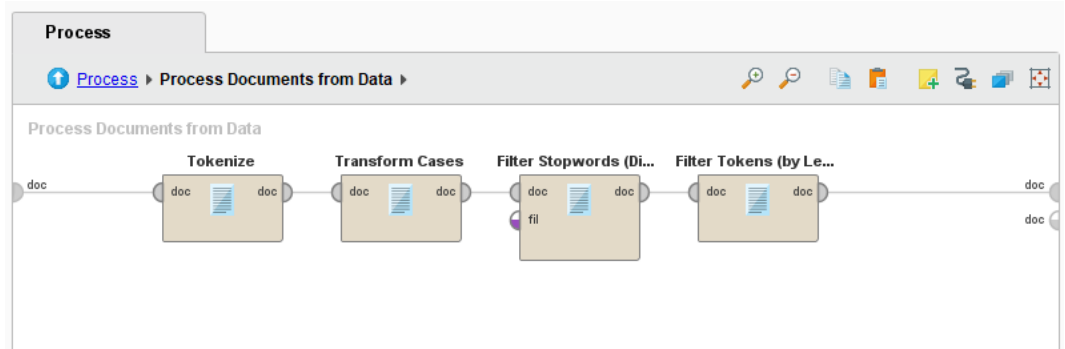
- Read CSV : Mengimport atau memproses data yang berformat csv
- Subprocess : Untuk menyimpan rantai operator dari proses cleaning
- Filter Examples : Digunakan untuk menghilangkan data yang mengandung missing value
- Remove Duplicates : Digunakan untuk menghilangkan data yang sama atau duplicates
- Write CSV : Menyimpan/mengekspor data dengan format csv

c) Pengolahan Data



Gambar 7 : Process pengolahan data

1) Tahap Pre-processing Data
Tahap pre-processing meliputi :



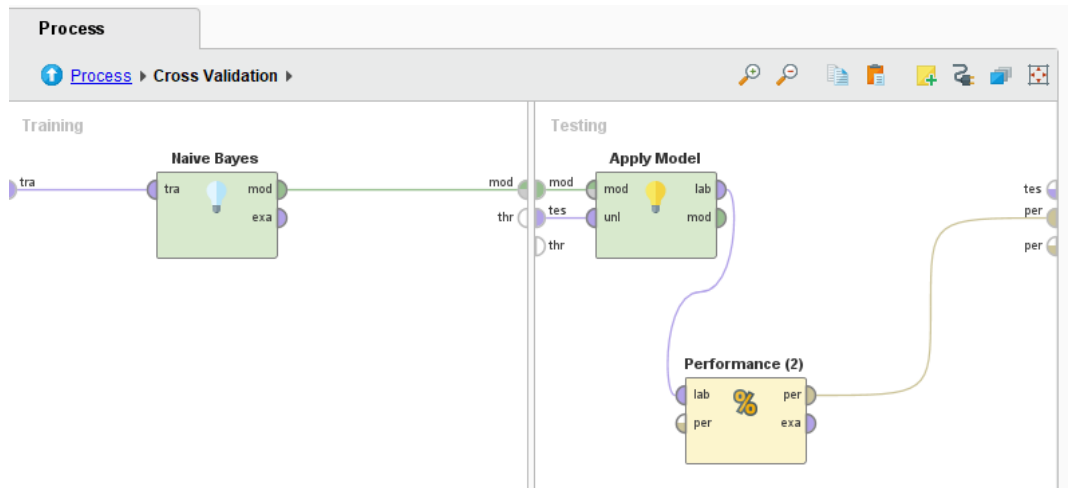
Gambar 8 : Tahap Pre-Processing

Keterangan :

- Tokenize : Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan panjang teks.
- Transform Cases : Dengan fitur transform cases kita dapat secara otomatis mengubah semua huruf pada teks menjadi huruf kecil semua atau menjadi huruf kapital semua, pada penelitian ini semua huruf dirubah kedalam huruf kecil karena teks berupa opini masyarakat terhadap pemilu yang sebagian besar merupakan huruf kecil semua.
- Filter Stopword : Dengan fitur Stopword maka teks sebelum diklasifikasikan dihilangkan dulu teks yang tidak berhubungan dengan analisa sentiment yang akan diteliti sehingga dimensi teks akan berkurang tanpa mengurangi isi sentimen dari komentar tersebut.
- Filter Tokens : operator ini untuk memfilter token berdasarkan panjang dari sejumlah karakter

2) Tahap Cross Validation

Langkah berikutnya dilakukan proses cross validation menggunakan metode Naïve Bayes yang meliputi proses training dan testing (apply model dan performance).Setelah itu dilakukan proses pemodelan dengan menggunakan metode naïve bayes untuk mendapatkan nilai accuracy, precision, recall berdasarkan performance.



Gambar 9 : Tahap Cross Validation

Terdiri dari 2 tahap ,yaitu tahap training dan testing. Pada tahap training, data dilatih menggunakan metode Naive Bayes. Kemudian pada tahap testing, data diproses melalui Apply Model dan Performance Test. Dari proses cross validation yaitu training dan testing, Metode Naïve Bayes dari nilai akurasi yaitu sebagai berikut.

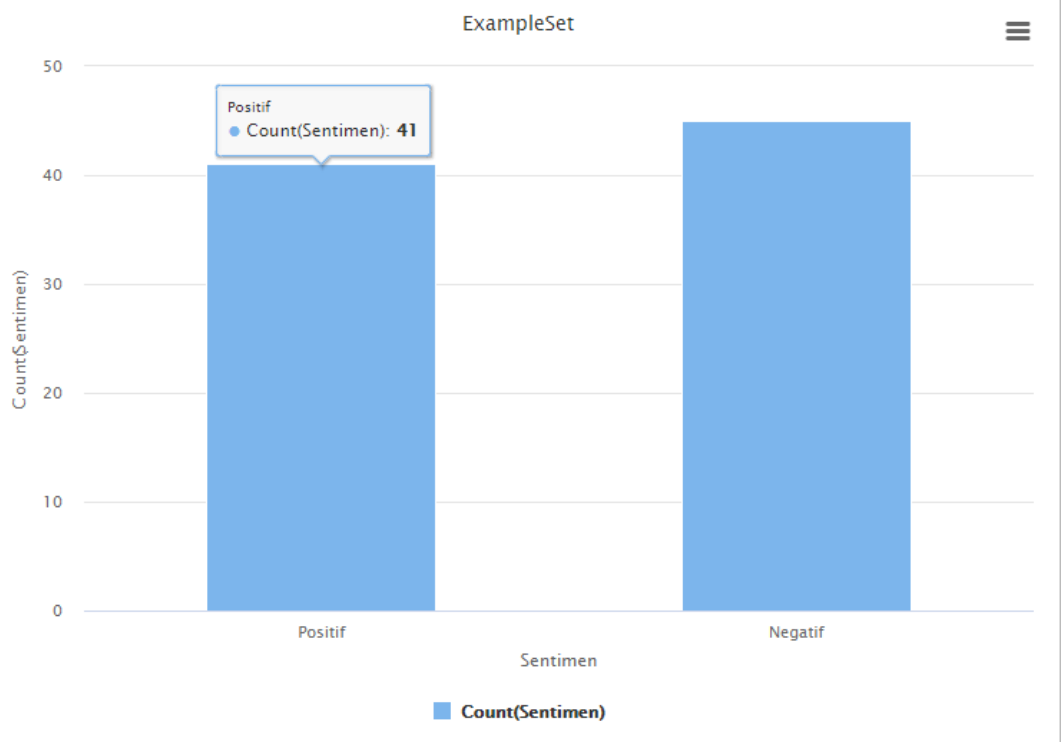
☒ Table View ☐ Plot View

accuracy: 60.56% +/- 13.42% (micro average: 60.47%)

	true Positif	true Negatif	class precision
pred. Positif	28	21	57.14%
pred. Negatif	13	24	64.86%
class recall	68.29%	53.33%	

Gambar 10 : Hasil nilai accuracy dengan metode Naive Bayes

Nilai akurasi yang didapatkan Metode Naïve Bayes. Pemilu adalah 60.56% dengan margin +/-13.42% dengan nilai rata-rata mikro sebesar 60.47%. Dengan demikian nilai accuracy maksimal adalah 73,98% dan nilai accuracy minimal adalah 47,14%. Jadi, dari proses cross validation yaitu training dan testing, metode Naïve bayes, tentang pemilu menghasilkan nilai accuracy 60.56%.



Gambar 11 : Hasil Visualisasi nilai sentiment dengan menggunakan metode Naive Bayes

4. KESIMPULAN

Berdasarkan analisis dan pengujian yang dilakukan hasil dari implementasi algoritma Naïve Bayes pada penelitian ini menunjukkan bahwa tingkat keakurasian untuk pemilu yaitu 60.56%. Hasil akurasi pada penelitian dinilai cukup baik. Analisis sentimen dengan algoritma Naïve Bayes berhasil mengklasifikasikan sentimen masyarakat di Twitter tentang pemilu. Hasil klasifikasi menunjukkan sentimen negatif lebih mendominasi dari sentiment positif, terdapat 45 ulasan Negatif dan 41 ulasan Positif. Dapat disimpulkan masih banyak orang yang masih ragu terhadap pemilu dilihat dari dominasi sentiment yang lebih tinggi negatif dari sentiment positif.