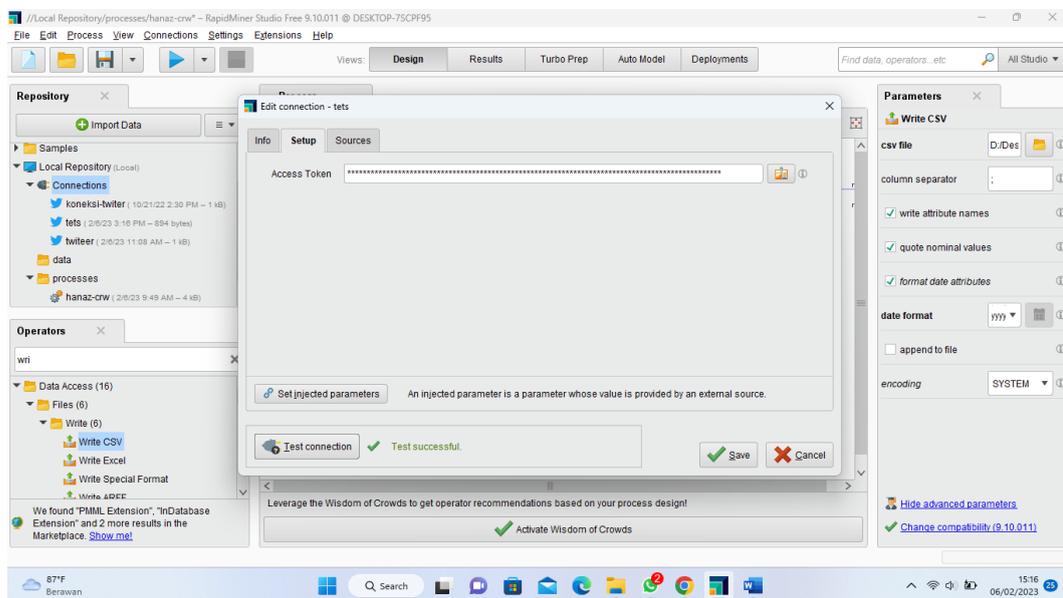


BAB IV

HASIL DAN PEMBAHASAN

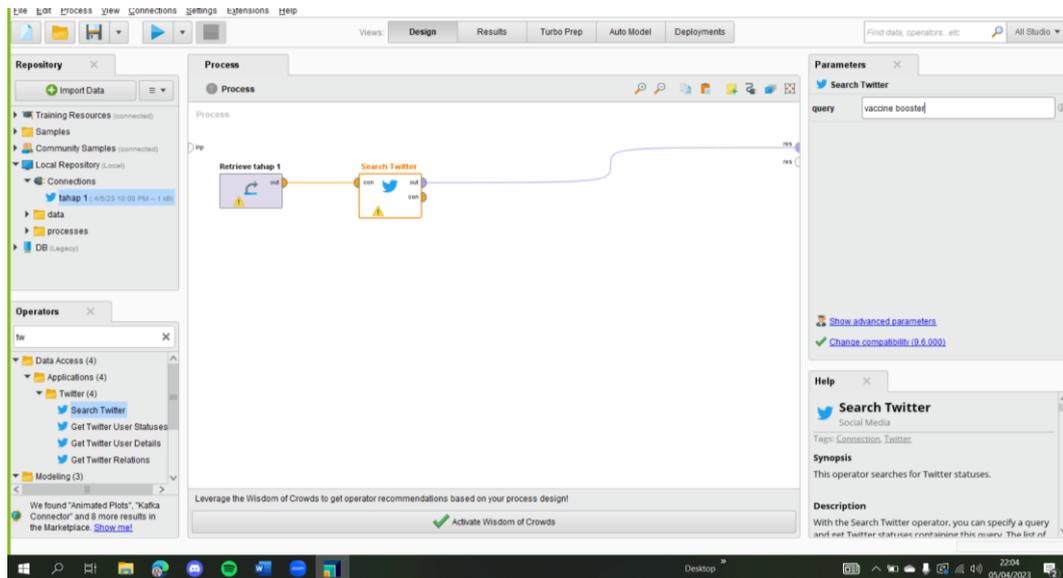
4.1 Proses Pengambilan data (*Crawling*)

Pengambilan data membutuhkan akun *twitter* karena proses perizinan membutuhkan akun agar mendapat *access token*. Tool *rapid miner* sebagai alat pengambilan data dibantu dengan operator search twitter yang memiliki connection antara aplikasi *twitter* dan *rapid miner*. Connection dibuat dengan menggunakan *access token* yang didapatkan lalu dilakukan *test connection* untuk menguji apakah *twitter* sudah terhubung dengan *rapid miner*. Hasil *test connection* dapat dilihat pada gambar 4.1



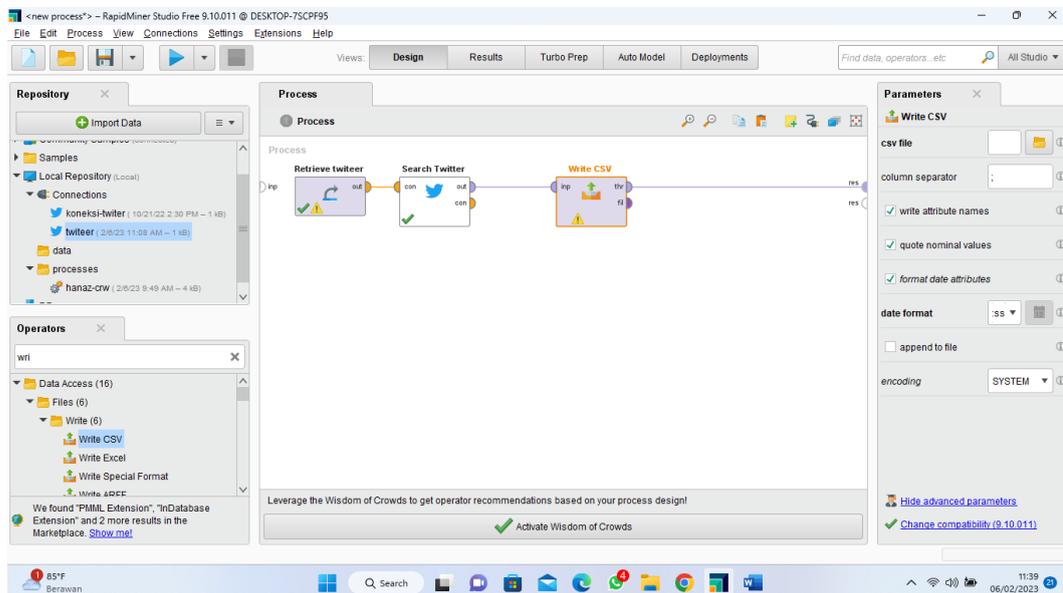
Gambar 4.1 Hasil *test connection*.

Setelah proses *test connection* berhasil dilakukan maka selanjutnya melakukan proses *crawling* data menggunakan tool *rapid miner* akan disimpan kedalam format *.csv*. Data yang didapatkan berdasarkan **10000 data mentah**. Proses pengambilan data dengan *rapid miner* dapat dilihat pada gambar 4.2.



Gambar 4.2 Proses pengambilan data (*crawling*)

Pengambilan data dilakukan dengan cara *crawling*, dalam melakukan *crawling* memanfaatkan operator *search twitter*. Kemudian untuk menyimpan data yang telah diperoleh kedalam bentuk *.csv* menggunakan operator *write csv*. Proses penyimpanan data *crawling* dapat dilihat pada gambar 4.3.



Gambar 4.3 Proses penyimpanan data (*crawling*)

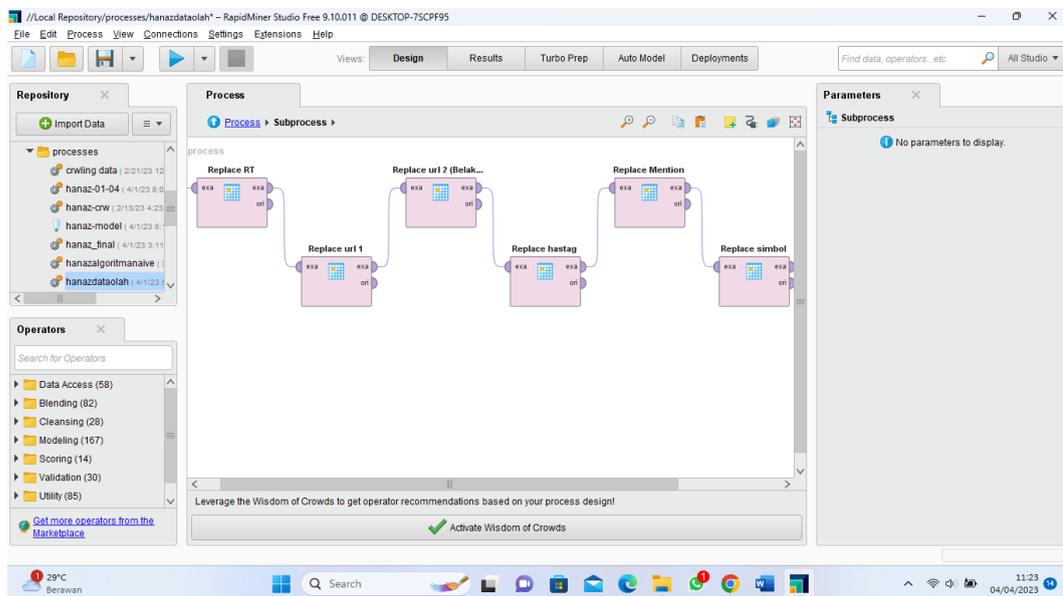
Setelah melakukan proses *crawling* data maka akan mendapat hasil berupa data mentah yang berjumlah 10000 data seperti pada gambar 4.4 dibawah ini.

Id	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1621810932...	Feb 4, 2023 5...	NHS England	551426814	?	-1	en	<a href="http...	Around 400,0...
1622455705...	Feb 6, 2023 1...	MrRobot3	1435761561...	rycunni	797404860	en	<a href="http...	@rycunni Thi...
1622453625...	Feb 6, 2023 1...	Pro-Noun's D...	355226992	?	-1	en	<a href="http...	RT @WelshR...
1622452591...	Feb 6, 2023 1...	Advaita	101713043	?	-1	en	<a href="http...	Stay warm all...
1622452049...	Feb 6, 2023 1...	Nohemi Melg...	134985143	?	-1	en	<a href="http...	RT @DrAlfred...
1622451412...	Feb 6, 2023 1...	Guest	1557430542...	?	-1	en	<a href="http...	RT @NBSap...
1622451122...	Feb 6, 2023 1...	Boian	9342300066...	jedstur	15720823	en	<a href="http...	@jedstur @la...
1622448855...	Feb 6, 2023 1...	Hunter in the ...	40326161	?	-1	en	<a href="http...	RT @hugh...
1622448783...	Feb 6, 2023 1...	Amber	3231865519	?	-1	en	<a href="http...	Need din b d...
1622448567...	Feb 6, 2023 1...	usslibertycal2	336641169	?	-1	en	<a href="http...	Third booster...
1622447930...	Feb 6, 2023 1...	Greg Fischer	1586219225...	MotherJones	18510860	en	<a href="http...	@MotherJon...
1622447027...	Feb 6, 2023 1...	KAGDUDE	1520411951...	StatesPoll	4547869279	en	<a href="http...	@StatesPoll...
1622446560...	Feb 6, 2023 1...	steve ast	7430959216...	?	-1	en	<a href="http...	Already regret...
1622446232...	Feb 6, 2023 1...	RenamedPM	1158690547...	ajthompson13	1483912172	en	<a href="http...	@ajthomps...

Gambar 4.4 Data (*crawling*)

a. Preprocessing Data

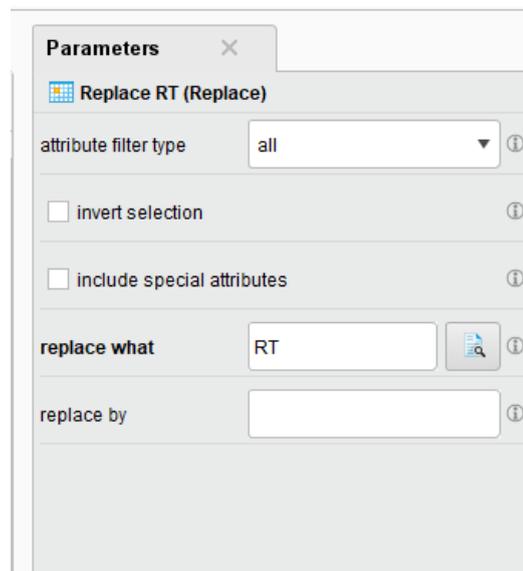
Tools yang digunakan dalam proses ini adalah *rapid miner*. Sebelum data diolah menggunakan algoritma naïve bayes dilakukan proses *pre-processing*. Pada tahap preprocessing hanya menggunakan kolom text, kemudian akan dilakukan beberapa proses untuk mendapatkan data acuan yang siap untuk di proses. Preprocessing data dilakukan untuk menghilangkan kata, url, *hashtag*, *mention*, data duplikat dan simbol yang tidak digunakan. Proses alur *preprocessing* data dan hasil data dapat dilihat pada gambar 4.5 dan gambar 4.6.



Gambar 4.5 Alur proses *preprocessing* data

b. Hapus RT

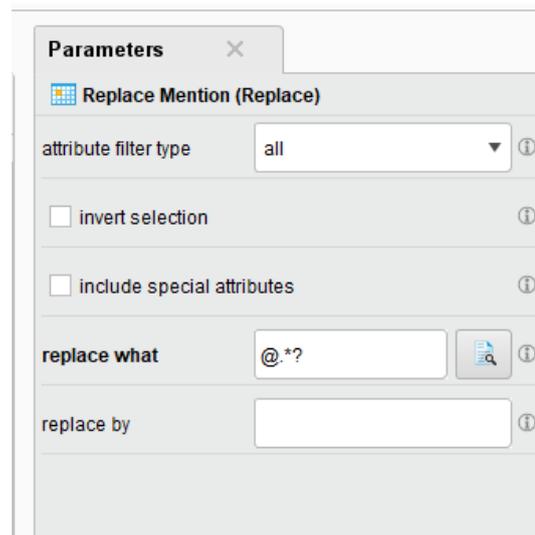
Tahap yang pertama adalah menghapus kata RT, parameter dalam operator proses ini dapat dilihat pada gambar 4.6



Gambar 4.6 Parameter hapus RT

c. Hapus *mention*

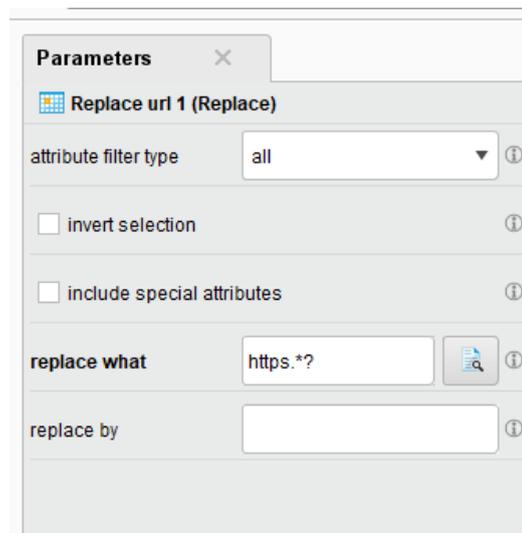
Tahap selanjutnya adalah menghapus *mention*, parameter operator dalam proses ini dapat dilihat pada gambar 4.7



Gambar 4.7 Parameter hapus *mention*

d. Hapus *url*

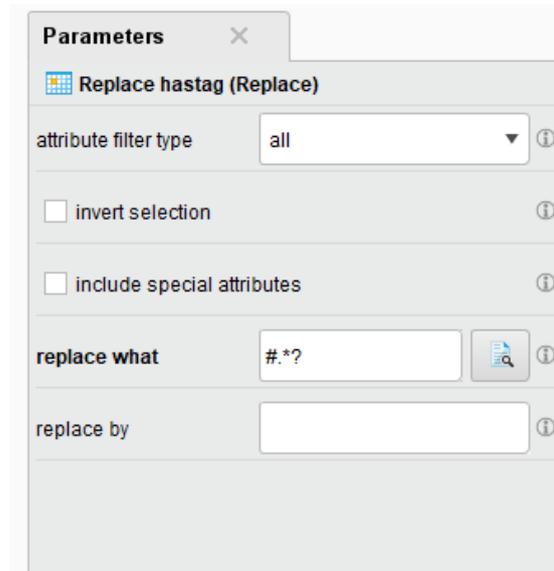
Tahap selanjutnya adalah menghapus *url*, parameter operator dalam proses ini dapat dilihat pada gambar 4.8



Gambar 4.8 Parameter hapus *url*

e. Hapus *hashtag*

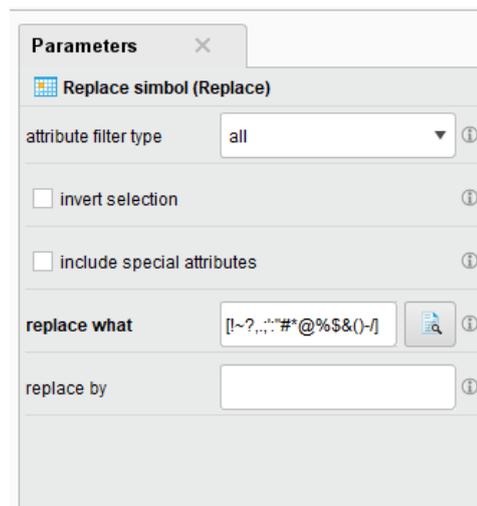
Tahap selanjutnya adalah menghapus *hashtag*, parameter operator dalam proses ini dapat dilihat pada gambar 4.9



Gambar 4.9 Parameter hapus *hashtag*

f. Hapus simbol

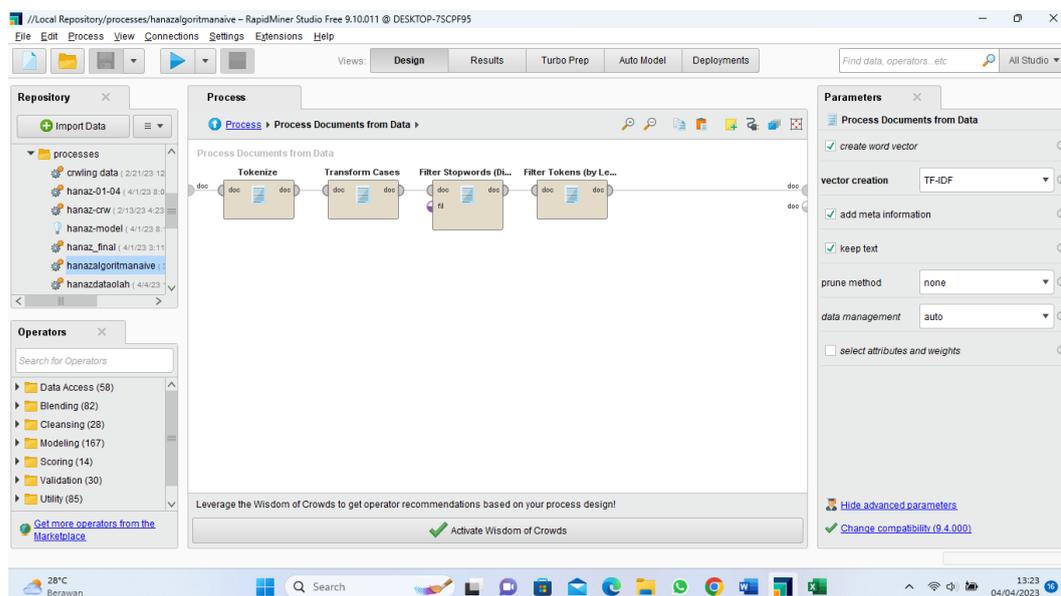
Tahap selanjutnya adalah menghapus simbol, parameter operator dalam proses ini dapat dilihat pada gambar 4.10



Gambar 4.10 Parameter hapus simbol

g. *Preprocessing Pada Teks*

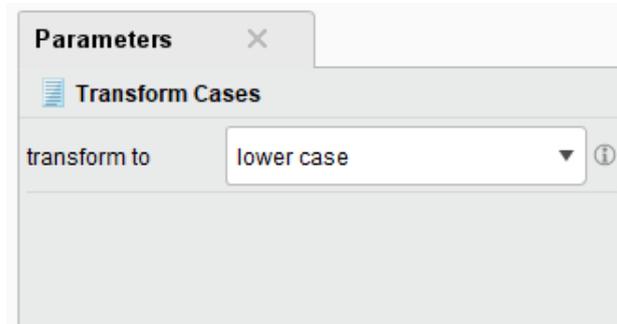
Tahapan berikutnya yaitu melakukan *preprocessing* dokumen dari data menggunakan operator tokenize, transform case, filter token dan filter stopwords. Alur preprocessing pada dokumen dapat dilihat pada gambar 4.11



Gambar 4.11 Alur proses *preprocessing* pada dokumen.

h. *Case folding*

Transform case merupakan operator *case folding*. Proses ini akan merubah seluruh kata dalam dokumen menjadi uppercase atau lowercase, untuk penelitian ini menggunakan lowercase. *Case folding* dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital, oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (huruf kecil). Parameter dalam operator proses ini dapat dilihat pada gambar 4.12.

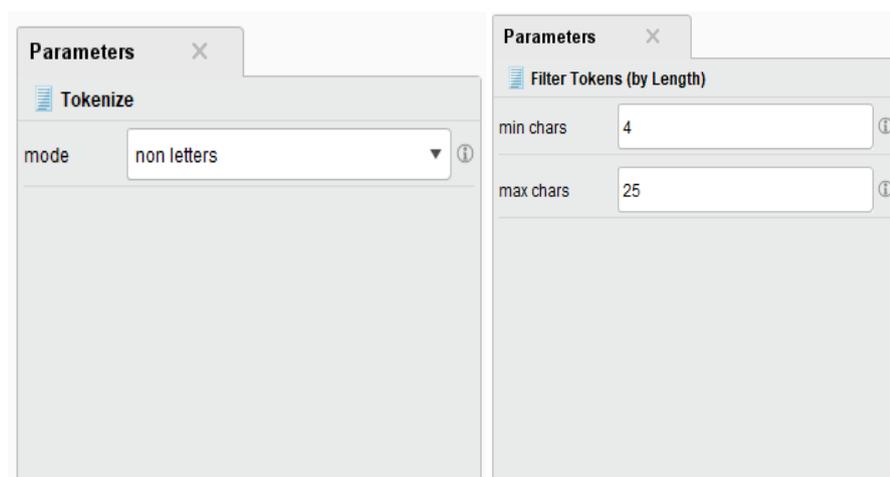


Gambar 4.12 Parameter *transform case*

Pada gambar 4.12 terlihat parameter yang digunakan dalam proses *transform case*, seluruh teks dalam dataset akan diubah menjadi lowercase.

i. Tokenizing

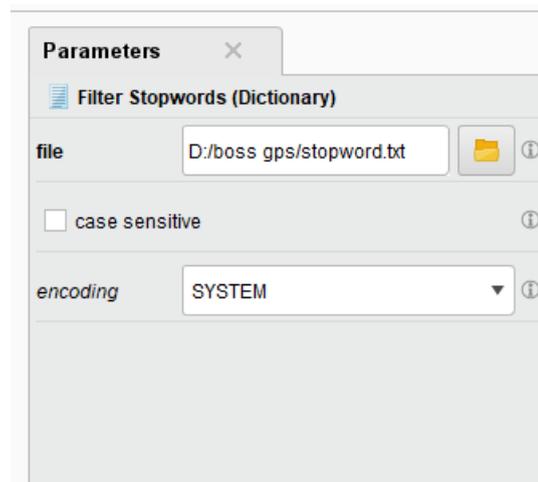
Tahap *tokenizing* merupakan tahap yang berfungsi untuk memisahkan antar kata yang ada didalam dataset. Pada teks dalam data terdiri dari beberapa kata yang saling terhubung dan dipisahkan dengan spasi. Untuk mempermudah dalam pemrosesan sebuah teks maka masing-masing kata dalam sebuah kalimat tersebut harus dipisahkan. *Tokenizing* merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna. Parameter operator *tokenizing* dapat dilihat pada gambar 4.13.



Gambar 4.13 Parameter *tokenizing*

j. Stopword removal

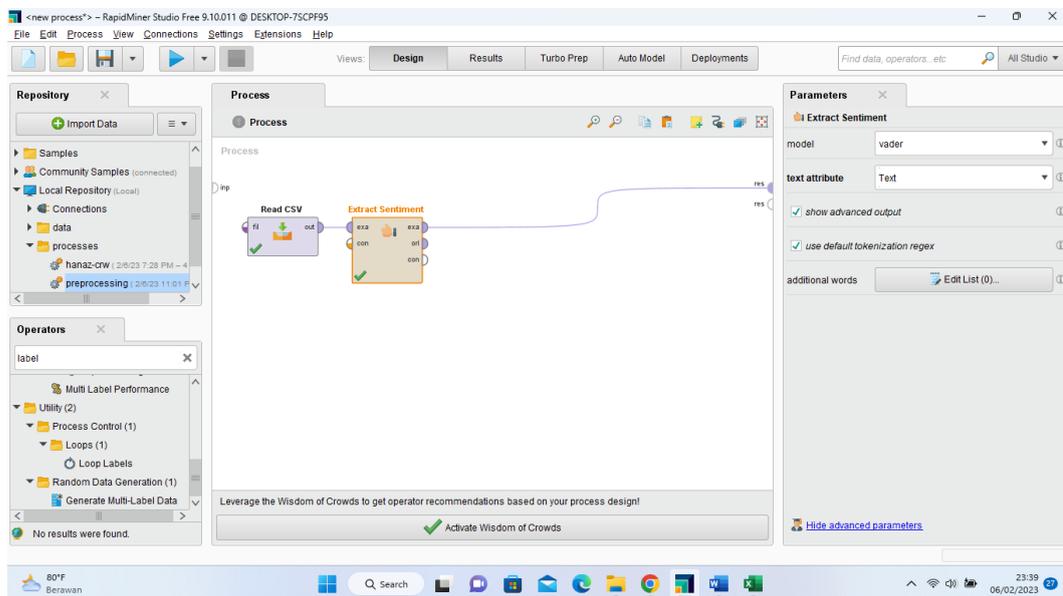
Stopword removal merupakan proses menghilangkan kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan. Kata yang dihilangkan adalah kata yang ada dalam dictionary yang berisi daftar kata (stopword list). Parameter operator proses *filter stopwords* ini dapat dilihat pada gambar 4.14.



Gambar 4.14 Parameter *filter stopwords*

k. Pelabelan Data

Pada proses pelabelan ini dilakukan pengelompokan data dengan cara pemberian label kelas menggunakan operator *extract sentiment*, adapun kelas yang digunakan adalah kelas berlabel positif, negative, dan netral. Kelas berlabel positif ditandai dengan nilai sentimen lebih dari 0, kelas berlabel netral ditandai nilai sentimen 0 sedangkan kelas negative memiliki nilai kurang dari 0. Proses pelabelan data dan hasil pelabelan dapat dilihat pada gambar 4.13 dan 4.14.



Gambar 4.15 Proses pelabelan data

Row No.	Score	Text
1	-0.590	The COVID-19 vaccine offer is changing. After 12 February, booster doses will no longer be offered. Future booster vac...
2	0.359	RISHI SUNAK Signed UK gov deal with Moderna before Xmas of £1 billion while refusing to say how much profit the ha...
3	0.205	According to the pharmacist covid figures are highest now since the first lockdown and the government are stopping gl...
4	0.077	If you want one go and get one NOW
5	0	Govt in LS. Health News, ET HealthW...
6	0.205	Top Thai authorities & advisors to the King have been in discussions with Prof Bhakdi & are preparing to have the Pfiz...
7	0.333	Meaning that Pfizer will become responsible for all vaccine injuries
8	-1.564	ericareport King of Thailand is breaking its contracts with Pfizer and seeking damages after several Royal Family mem...
9	-0.308	After 12 February, booster doses will no longer be offered.
10	0	Last call for adults under 50 to get Covid booster
11	0.205	Professor Sir Chris Whitty says the booster vaccine is a 'very important part of immunity to COVID' and is urging people...
12	0	By Michelle Roberts Digital health editor
13	-0.410	LPSmuts katnav01 HayleyMortimer Dr Robert Malone is the very scientist and pioneer of the MRNA Gene altering vacci...
14	0.538	My dad, 93, has had every possible COVID vaccine and booster. While many unvaccinated people have died from COVI...
15	-0.410	Have you had your COVID-19 autumn booster and your flu vaccine?

ExampleSet (1,170 examples, 1 special attribute, 1 regular attribute)

Gambar 4.16 Hasil nilai pelabelan data.

I. Pembentukan dataset

Dataset yang digunakan berjumlah 3458 baris *record tweet* yang akan digunakan sebagai data training sebanyak 1999 data, sedangkan data testing sebanyak 1459 data. *Sample* dari data *training* dan data testing yang digunakan dapat dilihat pada tabel 4.15 dan tabel 4.16 berikut.

Tabel 4.1 *Sample data training*

Text	sentiment
Sources I drew on in my speech today Florida vaccine data Estimation of number needed to vaccinate to prevent a COVID19 hospitalisation 12	POSITIF
Dr Anthony Fauci said that Americans will need booster dose of Covid vaccine every year like annual subscription	POSITIF
A 2yearold little girl received both the new Moderna mRNA Bivalent Booster shot and a Flu vaccine the same day The next day she was dead The little girl only lived long enough to see one Christmas she passed away on December 16 2022	NEGATIF
Sources I drew on in my speech today •Florida vaccine data •Estimation of number needed to vaccina...	POSITIF
Yes you it heard right... To keep 1 adult in their 50's out of hospital due to covid 43600 people needed to have a booster...	POSITIF
It may be due to boosterdid u get Pfizer or other mRNA vaccine before	NETRAL
Heres the net load of sinister shit for the sheep to conform to Why dont they have an alert for the vaccine booster time	NEGATIF

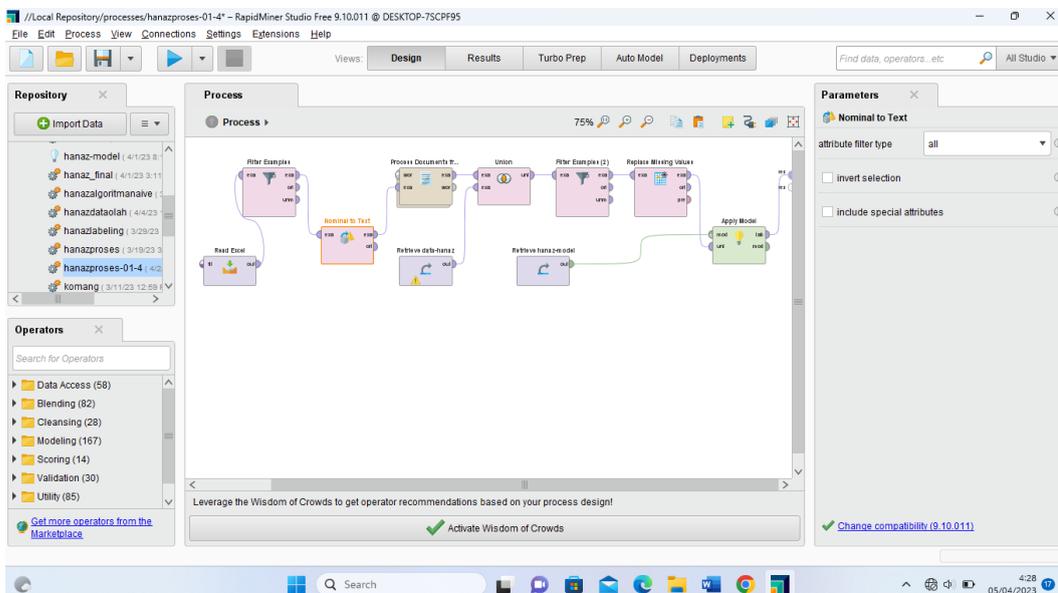
Tabel 4.2 *Sample data testing*

Text	sentiment
Vaccine effectiveness against SARSCoV2 infection with the Omicron or Delta variants following a twodose or booster BNT162b2 or mRNA1273 vaccination series	?
Do you still need to get your COVID19 autumn booster The booster programme is coming to an end soon but its not too late walkin appointments are available at our community vaccine clinic this Friday 3 Feb More vaccine options	?

FDA vaccine advisers ‘disappointed’ and ‘angry’ that early data about new Covid19 booster shot wasn’t presented for review last year CNN	?
Why is pushing you to get flu vaccine and at the same time when the says it incre...	?
Does this mean 1849 year olds can opt to have C19 booster until 12 Feb even if they dont have qualifying health condition	?
if it was mRNA yes if it had no long term safety data yes if it was a booster that only lasted 2 months and tested only on mice yes I would refuse that type of injection its not a vaccine really is it BillGates	?
One jab is unvaccinated UNLESS it is a one jab vaccine some are A full vaccination is both initial vaccinations and a booster Maybe some few really frail elderly can’t be vaccinated but it seems rare	?

m. Klasifikasi analisis sentimen

Setelah melakukan *preprocessing* data dan dokumen teks, tahap berikutnya adalah melakukan klasifikasi analisis sentimen pada data testing yang belum memiliki label. Tahap ini merupakan proses untuk memberikan pelatihan dan mengimplementasikan algoritma naïve bayes pada data testing yang belum memiliki label sebanyak 1459 *record* data. Proses klasifikasi pemberian label pada data testing dapat dilihat pada gambar 4.17.

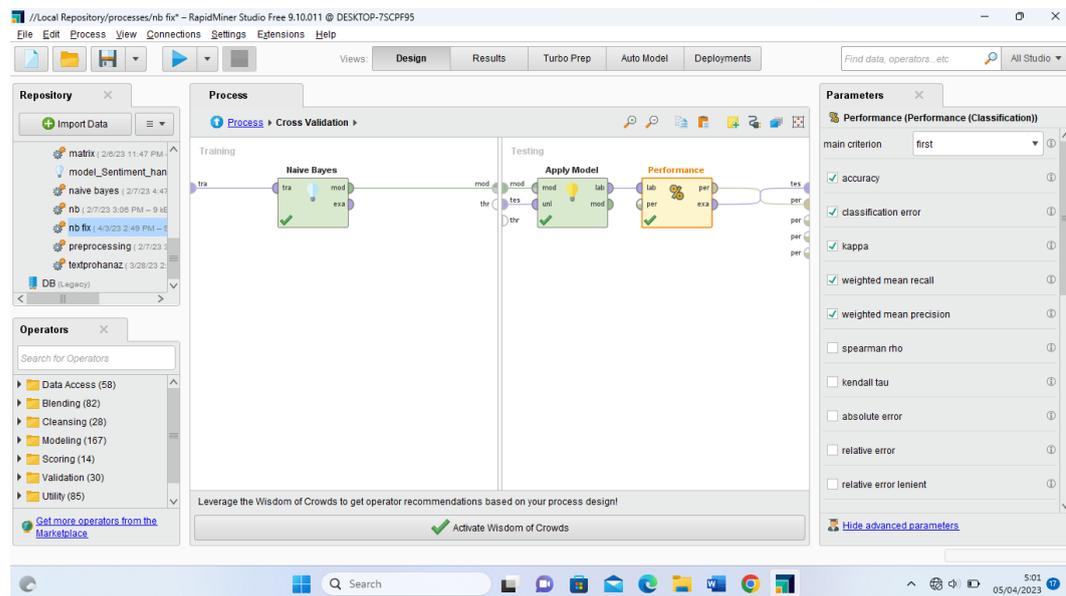


Gambar 4.17 Proses pelabelan data.

Dari proses klasifikasi pemberian label pada data testing yang berjumlah 1459 *record* data, menghasilkan *tweet* kelas positif sebanyak 647, *tweet* netral 187 dan *tweet* negatif sebanyak 625.

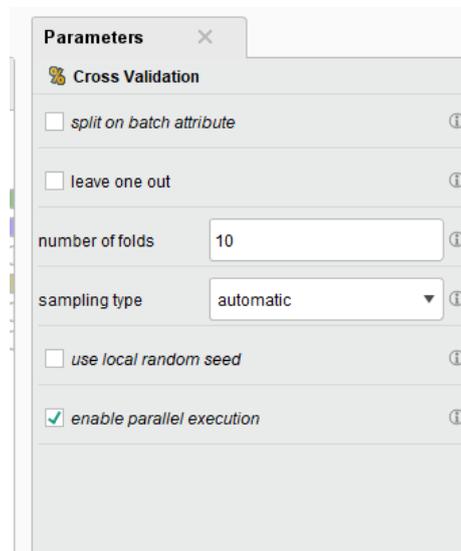
n. Evaluasi klasifikasi analisis sentimen

Setelah proses pemberian label data testing dilakukan, maka tahap selanjutnya adalah melakukan evaluasi terhadap data testing yang sudah terlabeli. Tujuan dilakukan tahap evaluasi ini untuk mengukur tingkat performa dari proses pemberian label dengan algoritma naïve bayes. Dalam melakukan tahap evaluasi ini menggunakan teknik *cross validation*.



Gambar 4.18 Proses *cross validation*

Tahapan yang dilakukan pada *cross validation* dapat dilihat pada gambar 4.18, mulai dari tahap klasifikasi naïve bayes, apply model dan juga *performance*. Parameter dalam operator *cross validation* dapat dilihat pada gambar 4.19.

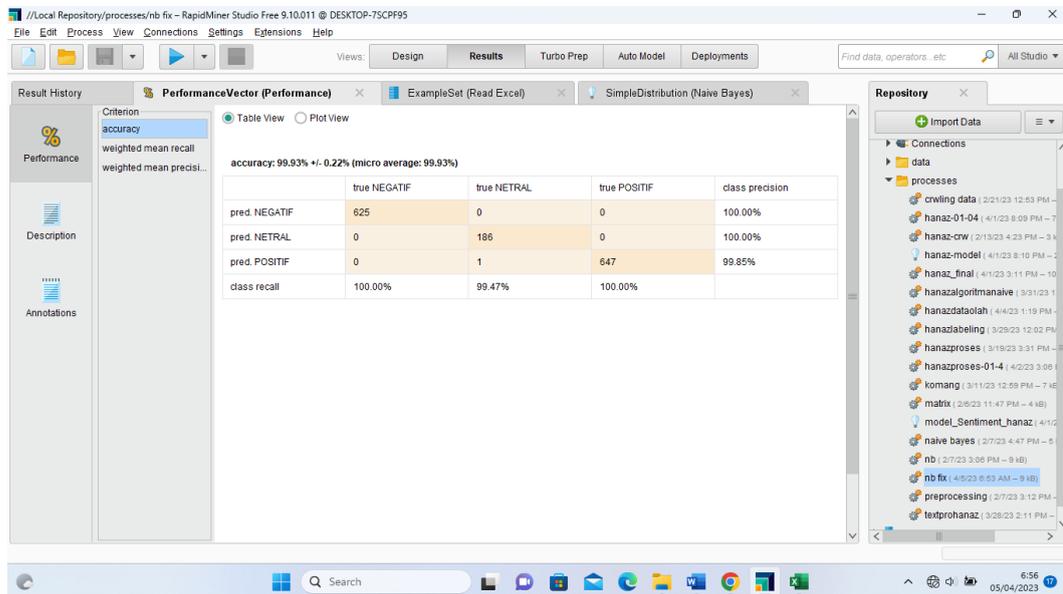


Gambar 4.19 Parameter *cross validation*

Pada gambar 4.19 dapat dilihat bahwa parameter dalam *cross validation* menggunakan validasi silang sebanyak 10 kali dengan tipe sampel data secara otomatis.

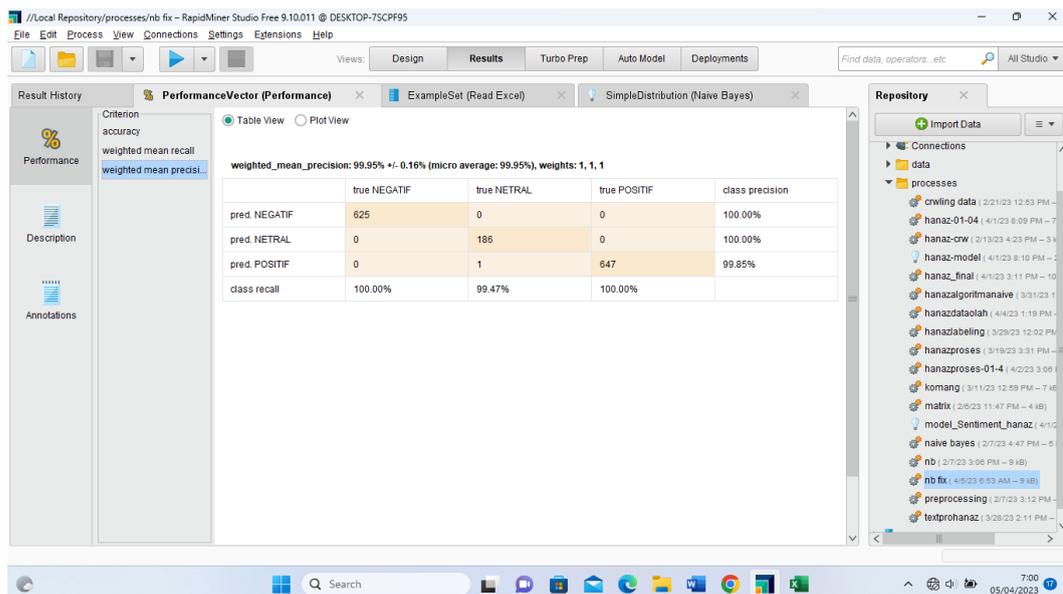
Setelah proses *cross validation* selesai tahap berikutnya adalah menjalankan proses *cross validation* tersebut guna mengetahui tingkat performa yang dihasilkan, tingkat performa yang diukur adalah akurasi, presisi dan recall.

Akurasi adalah jumlah dokumen yang diklasifikasi dengan benar, ketika hasil prediksi dari model dengan sentimen yang telah ditentukan, baik *true positive* maupun *true negatif*. Hasil dari nilai akurasi dapat dilihat pada gambar 4.15.



Gambar 4.20 Hasil nilai akurasi

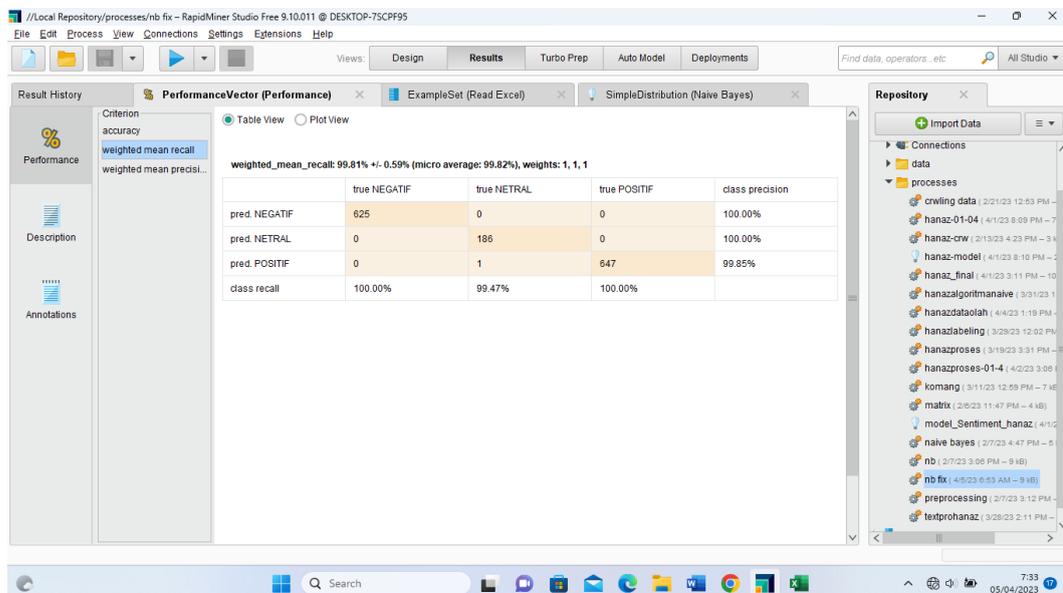
Pesisi adalah klasifikasi *true positive* dari semua data yang diprediksi sebagai kelas *positive*. Hasil dari nilai presisi dapat dilihat pada gambar 4.21.



Gambar 4.21 Hasil nilai presisi

Sedangkan recall merupakan banyaknya data yang relevansi dalam koleksi yang dihasilkan oleh system. Dengan kata lain, recall ialah jumlah data yang memiliki klasifikasi *true positive* dari semua dokumen yang benar-benar positif (termasuk

false negative). Hasil dari nilai recall dapat dilihat pada gambar 4.22.



Gambar 4.22 Hasil nilai recall