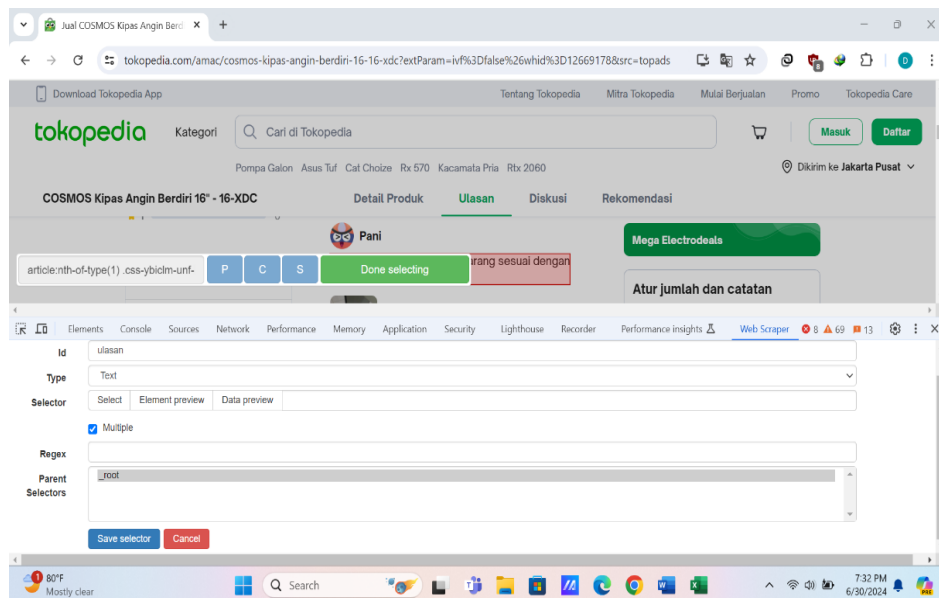


BAB IV

HASIL DAN PEMBAHASAN

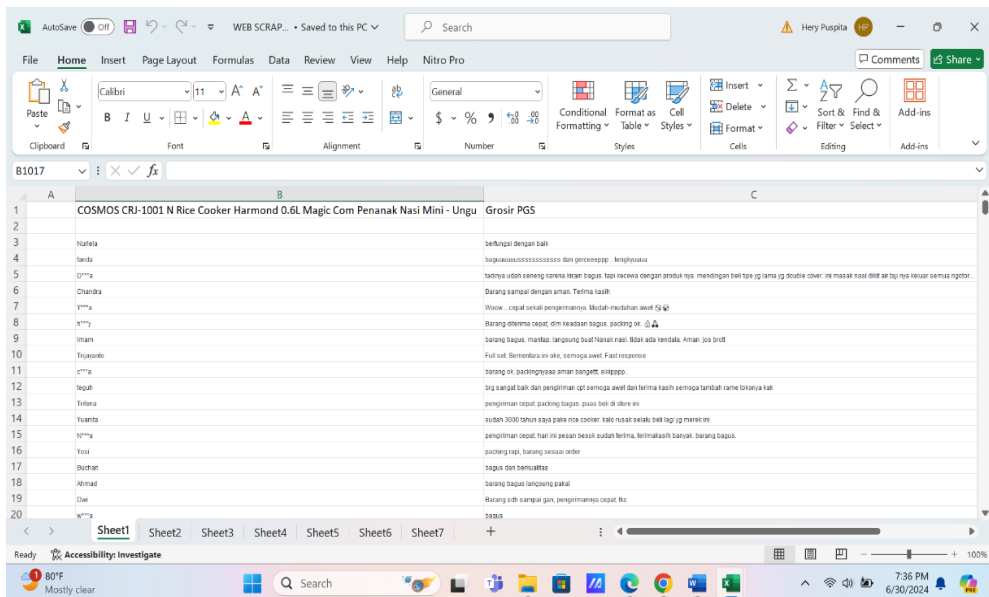
4.1 Seleksi Data

Tahap pengumpulan data ini dilakukan dengan Teknik *web scrapping* dengan bantuan *tools web scrapper* dari ekstensi *google chrome*. Berikut Langkah awal mengunjungi situs resmi Tokopedia melalui link *google.com* kemudian dilakukan *scrapping* dengan target ulasan pelanggan pada produk Cosmos yang ada di Tokopedia.



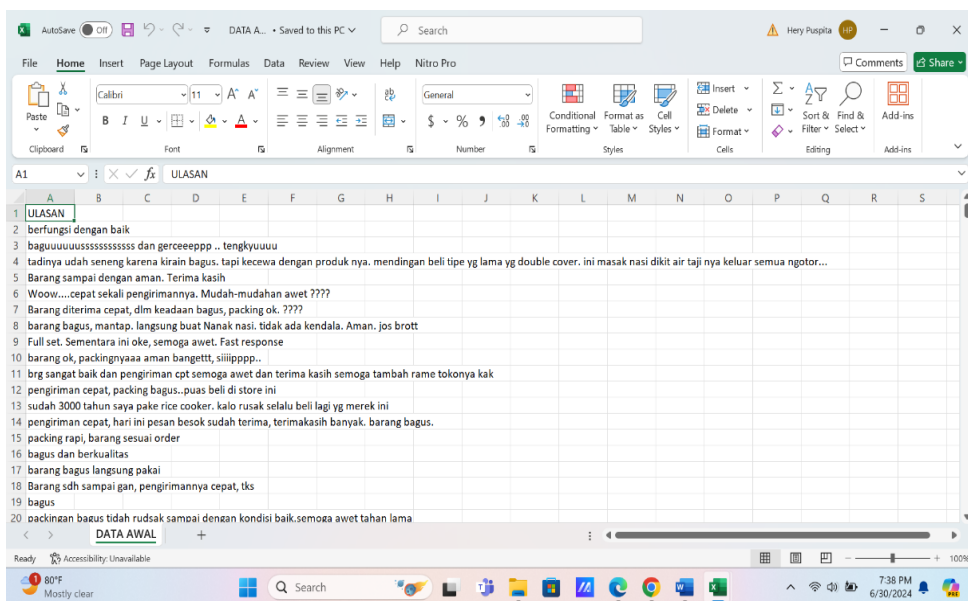
Gambar 4.1 Proses Scrapping Ulasan Produk Cosmos

Data yang berisi ulasan pengguna aplikasi Tokopedia diambil dari bulan Desember 2022 sampai bulan Desember 2023 data yang diambil sebanyak 1009 data ulasan. Terdapat dua atribut pada saat pengambilan data melalui hasil scrapping, Dimana terdapat atribut username dan ulasan. Berikut adalah dataset hasil *scrapping* data.



Gambar 4.1.2 Data Hasil Scrapping

Kemudian pada tahap ini dilakukan pemilihan atribut yang akan digunakan, sebelumnya pada tahap pengambilan data melalui *scrapping*. Terdapat dua atribut yaitu *username* dan ulasan. Selanjutnya dilakukan penghapusan atribut yang digunakan yaitu hanya ulasan.

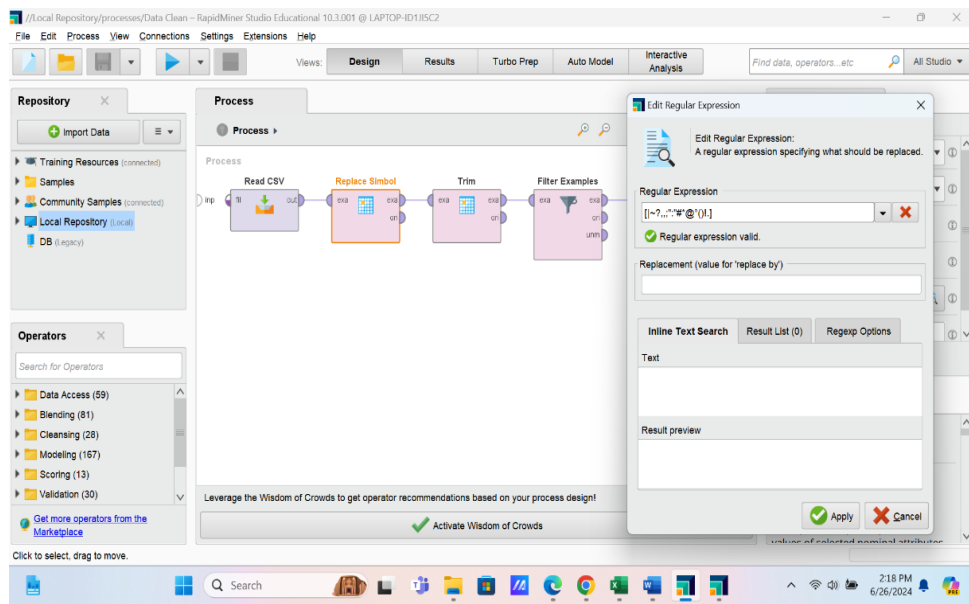


Gambar 4.1.3 Atribut Ulasan

4.2 Cleaning Data

Berikut adalah proses pembersihan data yang digunakan pada beberapa operator untuk penghapusan *symbol*, karakter, dan angka.

Pada penghapusan *symbol* ini operator yang digunakan yaitu operator *replace*. Pada operator *replace* pertama dinamakan *replace symbol* satu digunakan untuk menghapus simbol-simbol.



Gambar 4.2 Proses Cleaning Data

Berikut adalah hasil pada proses *cleaning* dapat dilihat pada tabel dibawah ini.

Tabel 4.2 Proses Cleaning Ulasan

Sebelum	Sesudah
baguuuuuusssssssssss dan gerceeeppp .. Tengkyuuuu	baguuuuuusssssssssss dan gerceeeppp tengkyuuuu
Woow....cepat sekali pengirimannya. Mudah-mudahan awet 🙄🙄	Woowcepat sekali pengirimannya Mudah-mudahan awet
Barang diterima cepat, dlm keadaan bagus, packing ok. 👍👍	Barang diterima cepat dlm keadaan bagus packing ok

Sebelum	Sesudah
barang bagus, mantap. langsung buat Nanak nasi. tidak ada kendala. Aman. jos brott	barang bagus mantap langsung buat Nanak nasi tidak ada kendala Aman jos brott
ga nyangka bisa beli rice cooker mini berkualitas dengan harga segini 🙏💕🙏	ga nyangka bisa beli rice cooker mini berkualitas dengan harga segini
Packingnya rapih, barangnya ori sesuai dengan pesanan. Terimakasih. 🙏	Packingnya rapih barangnya ori sesuai dengan pesanan Terimakasih
tadinya udah seneng karena kirain bagus. tapi kecewa dengan produknya. mendingan beli tipe yg lama yg double cover. ini masak nasi dikit air taji nya keluar semua ngotor...	tadinya udah seneng karena kirain bagus tapi kecewa dengan produknya mendingan beli tipe yg lama yg double cover ini masak nasi dikit air taji nya keluar semua ngotor

4.3 Labelling Data

Setelah semua data dibersihkan melalui *text processing* diatas tahap selanjutnya adalah melakukan pelabelan data secara manual. Dimana proses yang dilakukan dengan membaca satu persatu dari masing-masing ulasan dengan melihat kata-kata yang mengandung emosi maka akan dilabeli sentimen negatif begitu sebaliknya jika terdapat kata-kata pujian maka akan dilabeli sentiment positif. Berikut data yang sudah dibersihkan dan dilabeli secara manual.

Tabel 4.3 Hasil Pelabelan

Sebelum	Sesudah
berfungsi dengan baik	Positif
baguuuuuusssssssssss dan gerceeeppp tengkyuuuu	Positif

Sebelum	Sesudah
tadinya udah seneng karena kirain bagus tapi kecewa dengan produk nya mendingan beli tipe yg lama yg double cover ini masak nasi dikit air taji nya keluar semua ngotor	Negatif
Woowcepat sekali pengirimannya Mudah-mudahan awet	Positif
Barang diterima cepat dlm keadaan bagus packing ok	Positif
Ini bahan pancinya bukan harmond yg anti gores Mohon judul produk disesuaikan	Negatif
produk sdh diterima kondisi nya aman packing nya jg aman dan baik	Positif

4.4 Tokenize

Pada tahap selanjutnya dilakukan tokenisasi dimana proses pemecah dokumen menjadi bagian-bagian kata sebelum dan sesudah proses *tokenize* dapat dilihat pada table dibawah ini.

Tabel 4.4 Hasil ProsesTokenize

Sebelum	Sesudah
berfungsi dengan baik	berfungsi – dengan – baik
baguuuuuusssssssss dan gerceeeppp tengkyuuuu	Baguuuuuusssssssss – dan – gerceeeppp – tengkyuuuu
tadinya udah seneng karena kirain bagus tapi kecewa dengan produk nya mendingan beli tipe yg lama yg double cover ini masak nasi dikit air taji nya keluar semua ngotor	tadinya – udah – seneng – karena – kirain – bagus – tapi – kecewa – dengan – produk - nya - mendingan – beli – tipe – yg – lama – yg – double – cover – ini – masak – nasi – dikit – air – taji – nya – keluar – semua – ngotor

Sebelum	Sesudah
Woowcepat sekali pengirimannya Mudah-mudahan awet	Woowcepat – sekali – pengirimannya – Mudah-mudahan – awet
Barang diterima cepat dlm keadaan bagus packing ok	Barang – diterima – cepat – dlm – keadaan – bagus – packing – ok
Ini bahan pancinya bukan harmond yg anti gores Mohon judul produk disesuaikan	Ini – bahan – pancinya – bukan – harmond – yg – anti – gores – Mohon – judul – produk – disesuaikan
produk sdh diterima kondisinya aman packing nya jg aman dan baik	produk – sdh – diterima – kondisinya – aman – packing – nya – jg – aman – dan – baik

4.5 Case Folding

Pada tahap *case folding* yaitu proses mengubah teks menjadi huruf kecil (*lowercase*) dari dokumen yang awalan berupa (*uppercase*) huruf besar. Dimana tujuan dari proses *case folding* ini untuk menghilangkan perbedaan huruf kecil dan besar dalam suatu dokumen teks. Berikut adalah hasil dari tahap *Case Folding* dapat dilihat dari 32able dibawah ini.

Tabel 4.5 Hasil Proses Case Folding

Sebelum	Sesudah
berfungsi dengan baik	berfungsi dengan baik
baguuuuuusssssssss dan gerceeeppp tengkyuuuu	baguuuuuusssssssss dan gerceeeppp tengkyuuuu
tadinya udah seneng karena kirain bagus tapi kecewa dengan produknya mendingan beli tipe yg lama yg double cover ini masak nasi	tadinya udah seneng karena kirain bagus tapi kecewa dengan produknya mendingan beli tipe yg lama yg double cover ini masak nasi

Sebelum	Sesudah
dikit air taji nya keluar semua ngotor	dikit air taji nya keluar semua ngotor
Woowcepat sekali pengirimannya Mudah-mudahan awet	woowcepat sekali pengirimannya mudah-mudahan awet
Barang diterima cepat dlm keadaan bagus packing ok	barang diterima cepat dlm keadaan bagus packing ok
Ini bahan pancinya bukan harmond yg anti gores Mohon judul produk disesuaikan	ini bahan pancinya bukan harmond yg anti gores mohon judul produk disesuaikan
produk sdh diterima kondisi nya aman packing nya jg aman dan baik	produk sdh diterima kondisi nya aman packing nya jg aman dan baik

4.6 Stopwords Removal

Selanjutnya pada tahap ini filter *stopword* biasa disebut *stopword removal* dimana pada tahap *text processing* dilakukan penghapusan kata-kata yang tidak memiliki makna namun tidak berpengaruh pada suatu kalimat seperti “ada” “yang” “aku”. Berikut dibawah ini hasil dari tahapan *Stopword Removal*.

Tabel 4.6 Hasil Proses Stopword Removal

Sebelum	Sesudah
berfungsi dengan baik	berfungsi
baguuuuuusssssssss dan gerceeeppp tengkyuuuu	baguuuuuusssssssss gerceeeppp tengkyuuuu
tadinya udah seneng karena kirain bagus tapi kecewa dengan produk nya mendingan beli tipe yg lama yg double cover ini masak nasi dikit air taji nya keluar semua ngotor	udah seneng bagus kecewa produk nya mendingan beli tipe double cover masak nasi dikit air taji nya ngotor

Sebelum	Sesudah
woowcepat sekali pengirimannya mudah-mudahan awet	woowcepat pengirimannya mudah-mudahan awet
barang diterima cepat dlm keadaan bagus packing ok	barang diterima cepat bagus packing ok
ini bahan pancinya bukan harmond yg anti gores mohon judul produk d disesuaikan	bahan pancinya harmond anti gores mohon judul produk d disesuaikan
produk sdh diterima kondisi nya aman packing nya jg aman dan baik	produk diterima kondisi nya aman packing nya aman

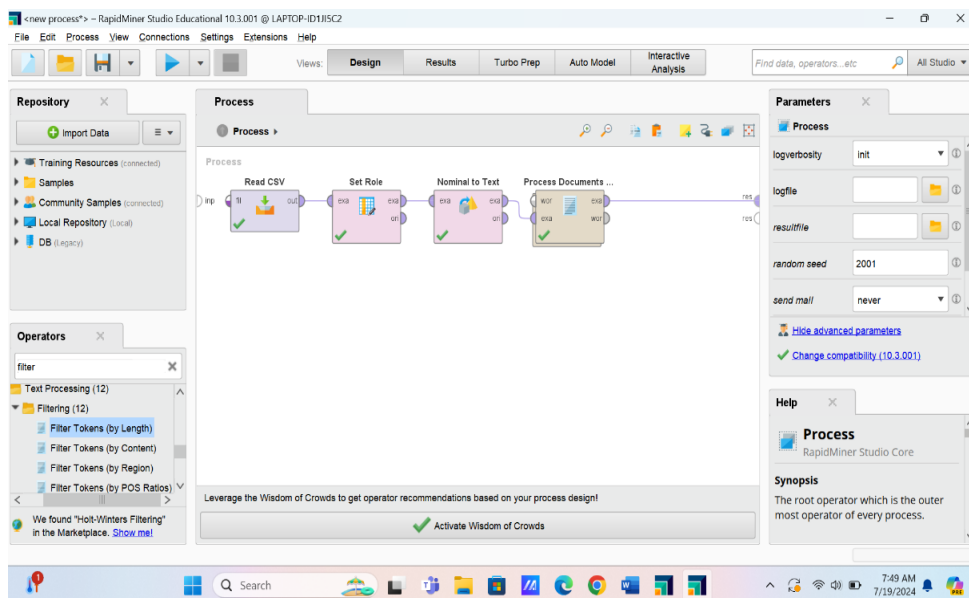
4.7 Filter token (by length)

Pada tahap ini dilakukan penghapusan sejumlah kata atau kata singkatan yang minimal karakternya sudah ditentukan sebelum digunakan, pada penelitian ini peneliti menggunakan Panjang minimal 4 karakter dan maksimal 25 karakter. Berikut adalah tabel sebelum dan sesudah dilakukan tahapan *Filter token (by length)*.

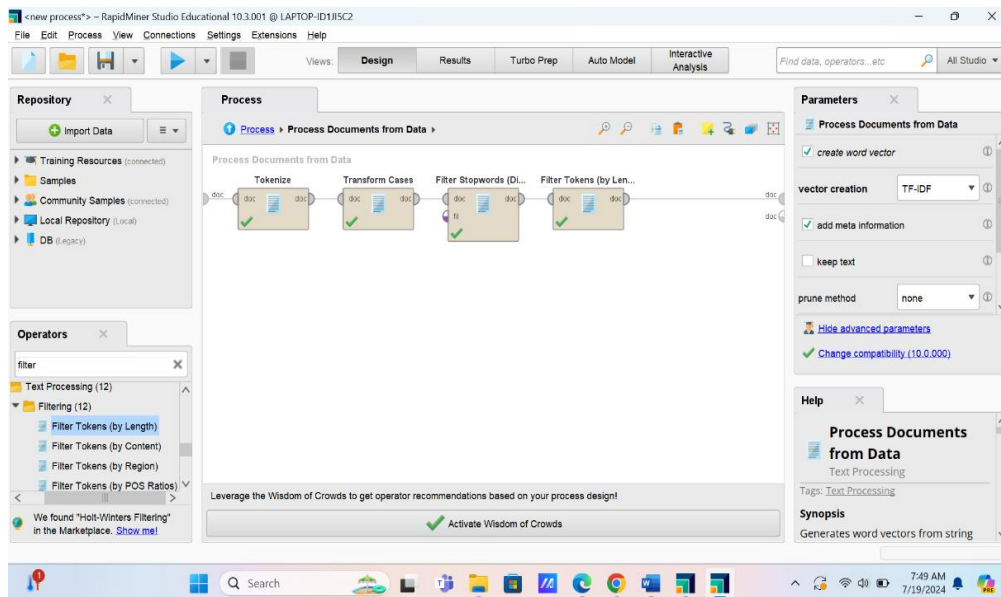
Tabel 4.7 Hasil Proses Filter Token (by length)

Sebelum	Sesudah
berfungsi dengan baik	berfungsi
baguuuuuusssssssss dan gerceeeppp tengkyuuuu	baguuuuuusssssssss gerceeeppp tengkyuuuu
tadinya udah seneng karena kirain bagus tapi kecewa dengan produk nya mendingan beli tipe yg lama yg double cover ini masak nasi dikit air taji nya keluar semua ngotor	kecewa produk mendingan beli tipe double cover masak nasi dikit air ngotor
woowcepat sekali pengirimannya mudah-mudahan awet	pengirimannya mudah-mudahan awet

Sebelum	Sesudah
barang diterima cepat dlm keadaan bagus packing ok	barang diterima cepat bagus packing ok
ini bahan pancinya bukan harmond yg anti gores mohon judul produk disesuaikan	Bahan harmond anti gores mohon judul produk disesuaikan
produk sdh diterima kondisi nya aman packing nya jg aman dan baik	produk diterima kondisi aman packing aman



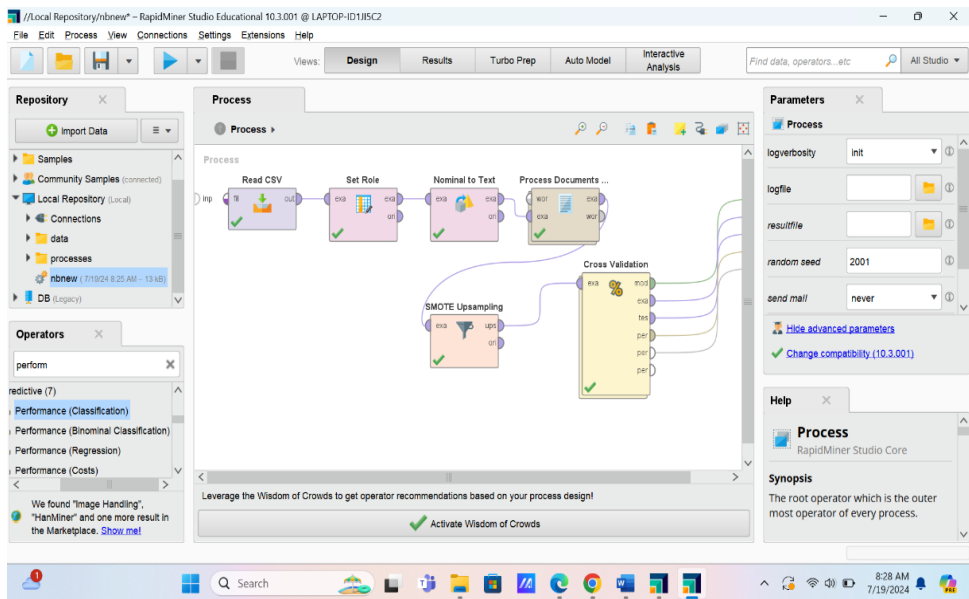
Gambar 4.7.1 Proses Processing



Gambar 4.7.2. Teknik Processing

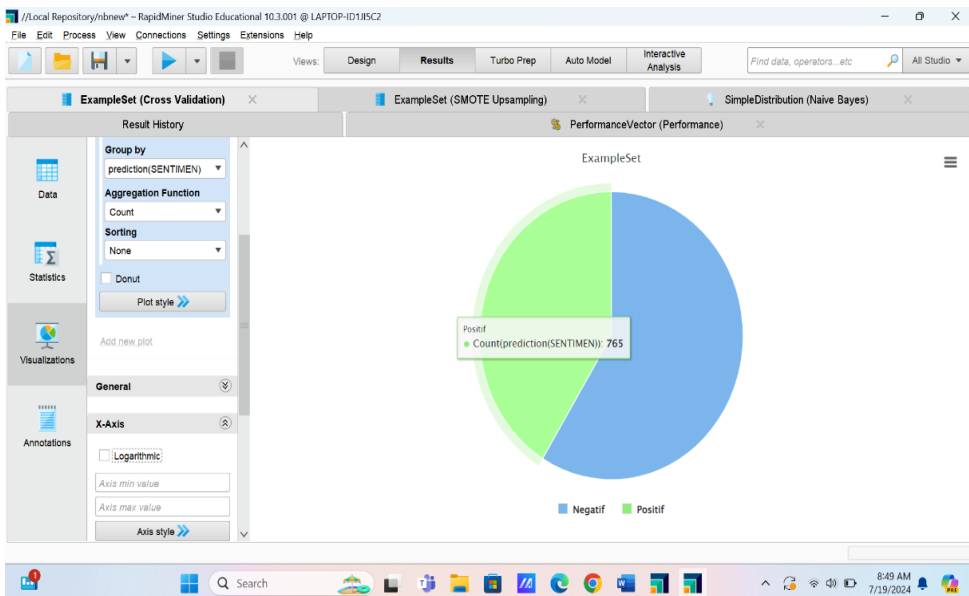
4.8 Klasifikasi Naïve Bayes

Pada tahap ini dilakukan model klasifikasi naïve bayes yang mana terdapat beberapa operator diantaranya *Read Excel* yang berfungsi menampung data train lalu dihubungkan dengan Set Role yang mana untuk mengubah target role nya menjadi label kemudian dihubungkan dengan *Nominal to Text* lalu ke proses *Document From Data* yang didalamnya terdapat *sub proses seperti tokenize, case folding, stopword removal, dan filter token (by length)*. Kemudian *upsampling* untuk menyeimbangkan data kemudian *cross validation* untuk menguji model naïve bayes dengan pengujian 1-10 kali.



Gambar 4.8.1 Proses Model Naïve Bayes

Berdasarkan hasil dari Klasifikasi Naïve Bayes didapatkan Klasifikasi prediksi sentimen positif sebesar 765 teks dan klasifikasi prediksi sentimen negatif sebesar 1089 teks dari total teks 1854.



Gambar 4.8.2 Diagram Sentimen

Berikut presentasi sentimen positif dan sentimen negatif dari hasil klasifikasi dengan hasil sentiment positif dibagi seluruh dokumen dan sentimen negatif dibagi seluruh dokumen. Berikut persentasi sentimen positif dan sentiment negatif.

Persentasi sentimen positif

$$\frac{765}{1854} \times 100\% = 41\%$$

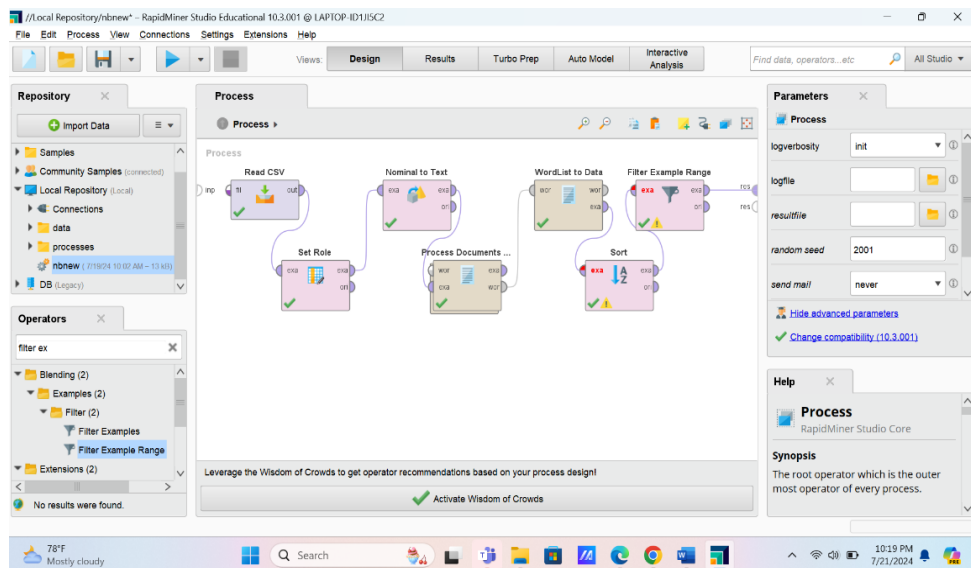
Persentasi sentimen negatif

$$\frac{1089}{1854} \times 100\% = 59\%$$

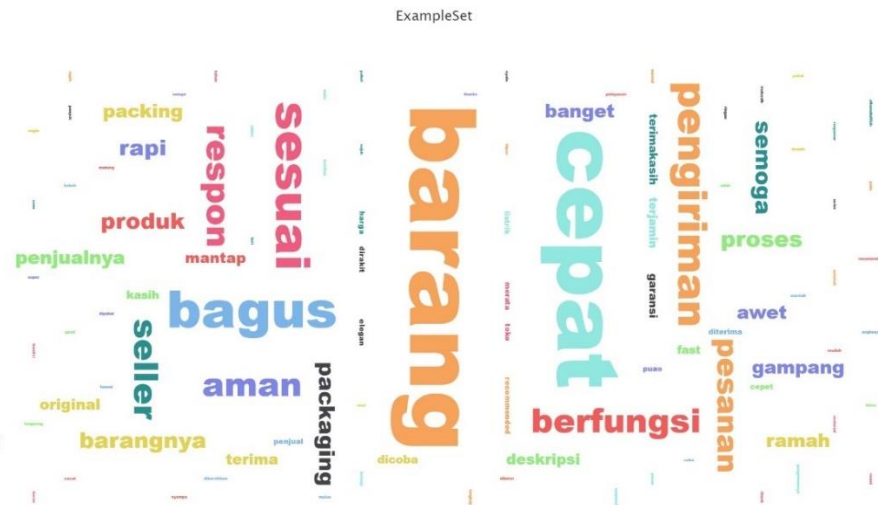
Diperoleh hasil persentasi sentimen dengan jumlah ulasan negatif lebih banyak dari pada jumlah sentiment positif. Jumlah sentiment positif sebesar 41%, dan jumlah sentimen negatif sebesar 59%. Artinya dari keseluruhan hasil klasifikasi terdapat hasil prediksi sentimen negatif lebih besar dari pada sentimen positif.

4.9 Wordcloud

Pada tahap ini ditampilkan bobot kata yang dominan muncul, divisualisasikan dalam bentuk *wordcloud*, dan kata yang sering muncul pada teks adalah kata “Barang”. Dapat dilihat pada gambar dibawah berikut.



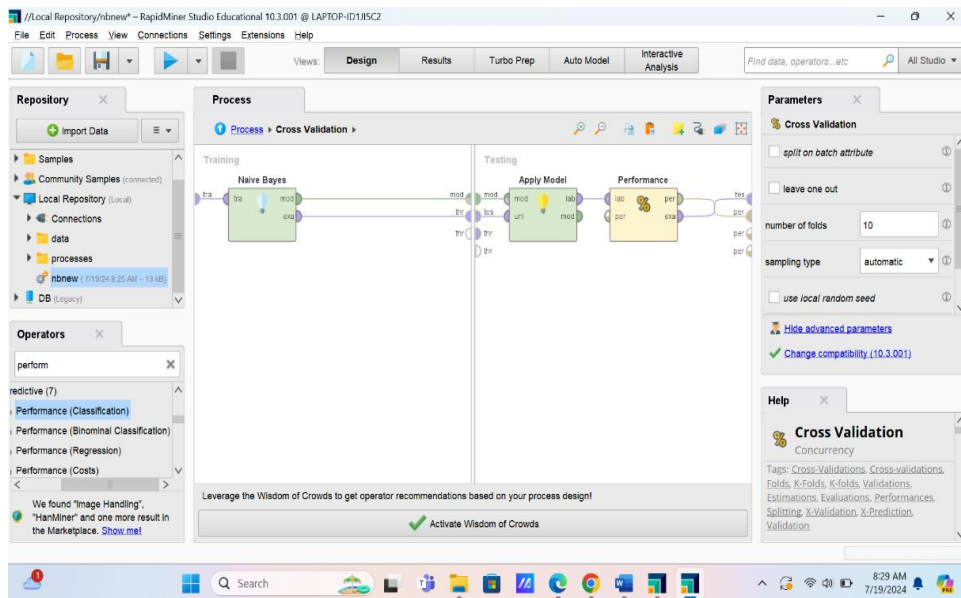
Gambar 4.9.1 Proses Wordcloud



Gambar 4.9.2 Wordcloud

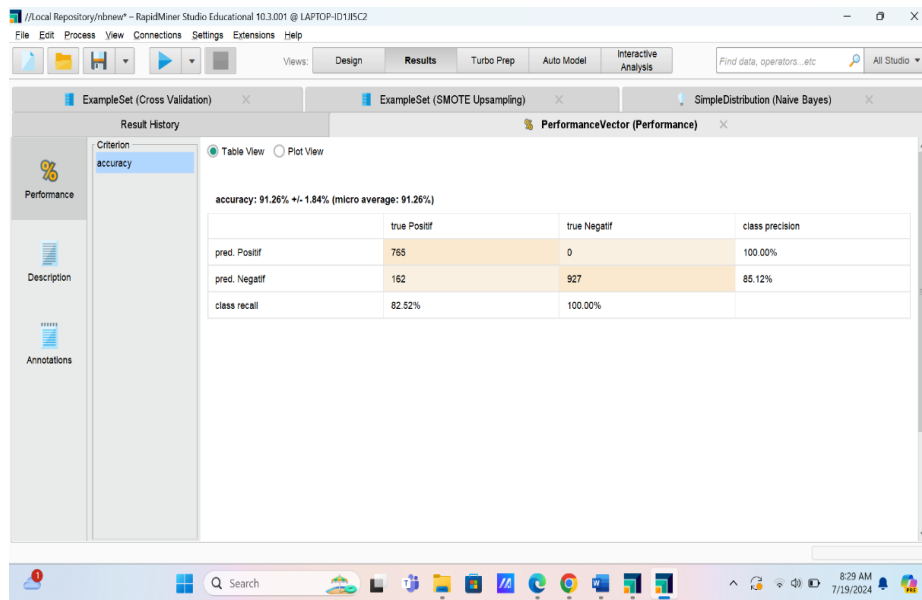
4.10 Evaluasi

Tahap ini dilakukan untuk menghitung evaluasi dari model naïve bayes dengan menggunakan *cross validation* yang didalamnya dihubungkan dengan *apply model* dan *performance* untuk dihitung terhadap hasil dari naïve bayes.



Gambar 4.10.1 Proses Validasi Naïve Bayes

Berdasarkan hasil evaluasi model bahwa hasil klasifikasi terhadap model naïve bayes didapatkan hasil *accuracy* sebesar 91,26%, bisa dilihat pada gambar dibawah ini.



Gambar 4.10.2 Hasil Accuracy Naïve Bayes

Langkah selanjutnya peneliti menggunakan perhitungan manual *confusion matrix* dimana dilakukan untuk menghitung *accuracy*, *precision*, dan *recall*. Adapun rumus *Confusion Matrix* dilihat pada rumus dibawah ini.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} = \frac{765+927}{765+927+0+162} = \frac{1692}{1854} \times 100\% = 91,26\%$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{765}{765+0} = \frac{765}{765} \times 100\% = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{765}{765+162} = \frac{765}{927} \times 100\% = 82,52\%$$

$$\text{F1 - Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Keterangan :

TP : nilai *True Positive*

FP : nilai *False Positive*

FN : nilai *False Negative*

TN : nilai *True Negative*.

Berdasarkan pada tabel *confusion matrix* bahwa terdapat data yang terklasifikasi benar sebanyak 765 data dan banyak juga data yang benar namun terklasifikasi atau prediksi negatif sebanyak 927 data, dan salah prediksi terdapat 0 data yang masuk dalam kategori negatif yang seharusnya positif, kemudian terdapat data 162 data salah prediksi atau klasifikasi yang masuk dalam kategori positif namun seharusnya negatif.

Pada hasil perhitungan manual *confusion matrix* diatas menunjukkan hasil *accuracy* sebesar 91,26% *precision* 100% dan *recall* sebesar 82,52%. Dari hasil perhitungan tersebut sama dengan hasil perhitungan dari rapid miner diatas.