

**LAPORAN PRAKTIKUM TEKS MINING**

**DATA MINING**



**DISUSUN OLEH :**

**Hana Hanifah - 10110024**

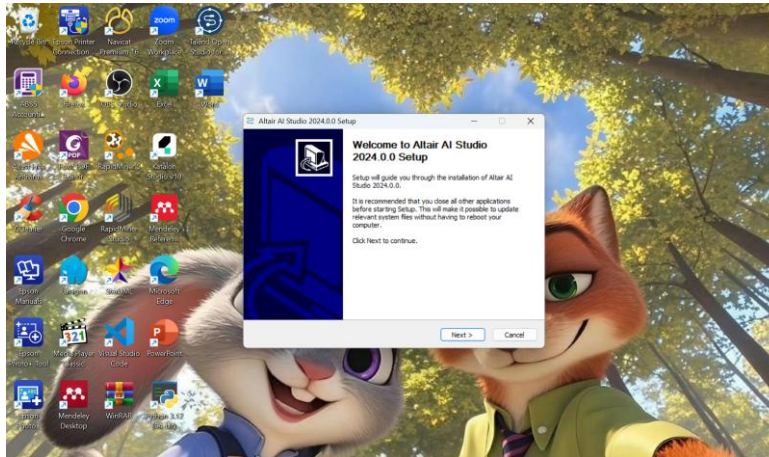
**PRODI D3 SISTEM INFORMASI**

**JURUSAN TEKNOLOGI INFORMASI DAN KOMPUTER**

**POLITEKNIK NEGERI SUBANG**

**2025**

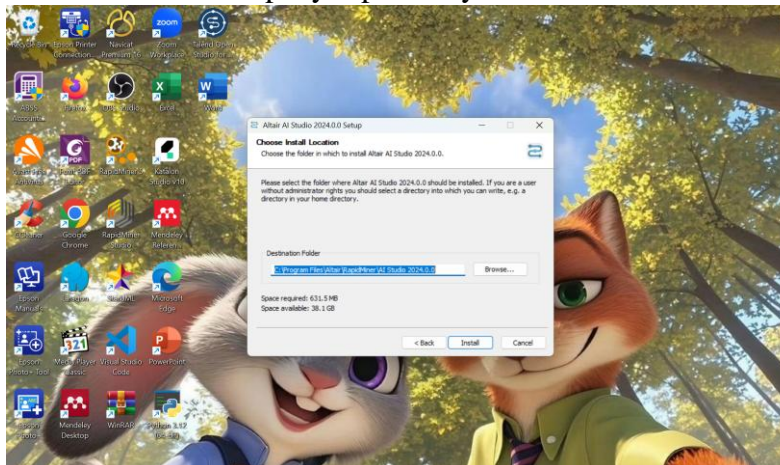
## 1. Kita Install Altair AI Studio



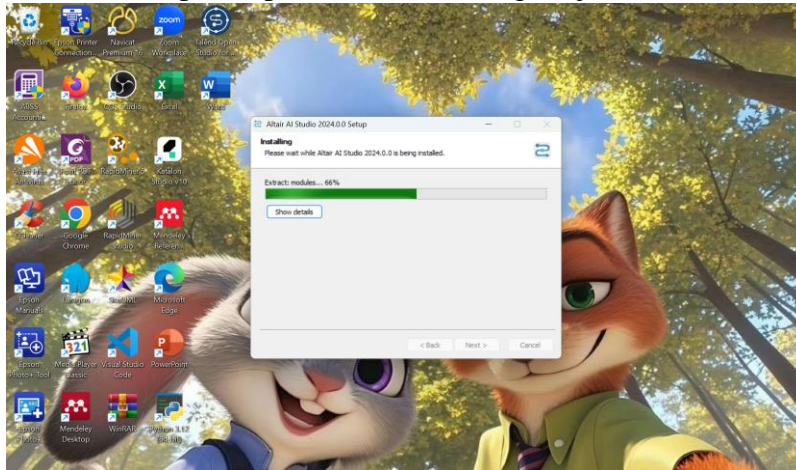
## 2. Lalu klik I Agree



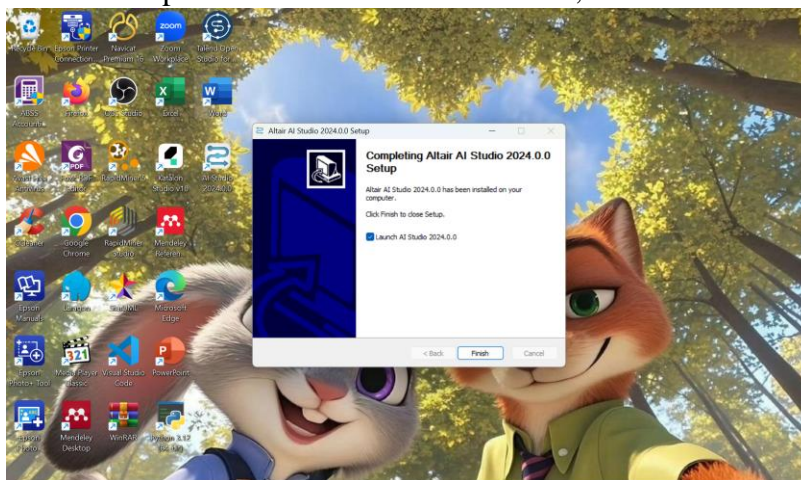
## 3. Lalu disini kita atur penyimpanan nya.



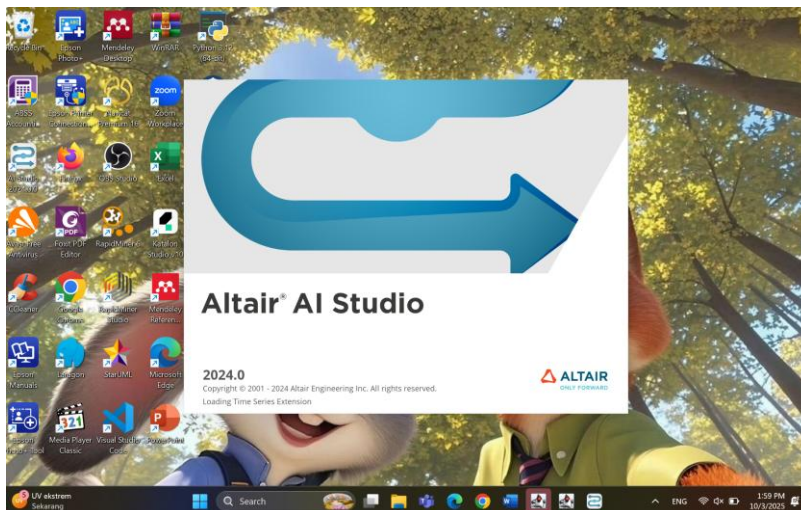
4. Lalu ini merupakan proses install sedang berjalan.



5. Lalu ini tampilan ketika install sudah selesai, lalu klik finish.

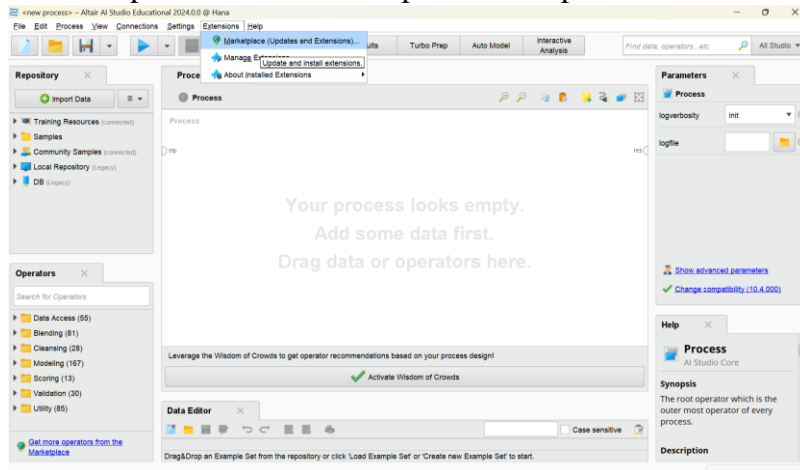


6. Lalu setelah itu kita buka software altair AI Studio

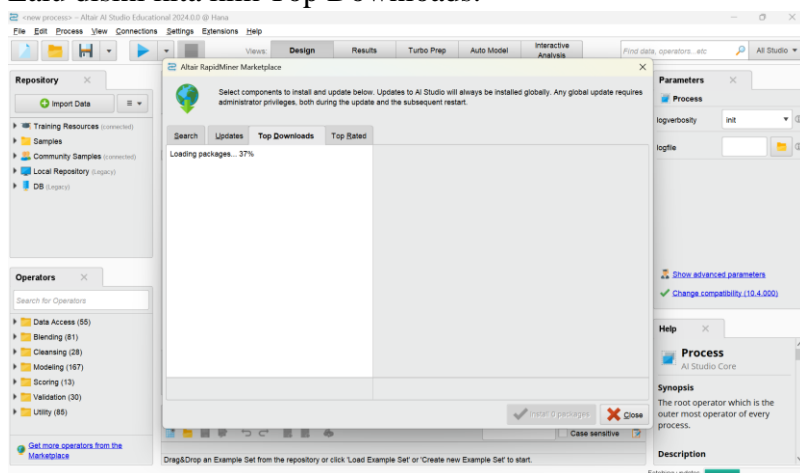




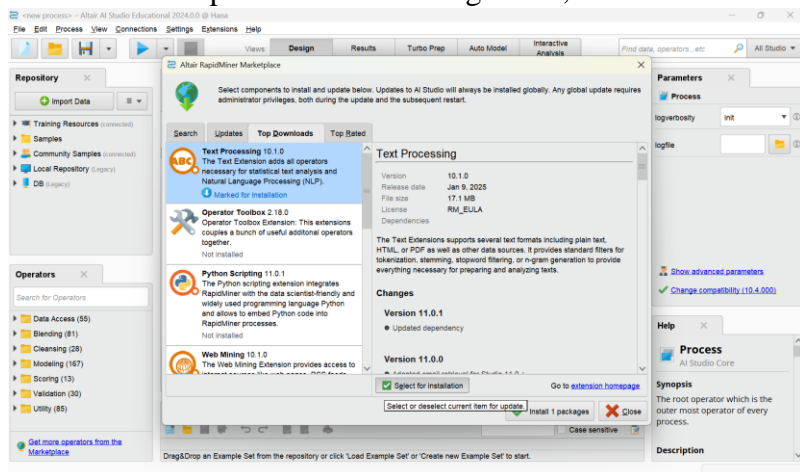
7. Lalu kita pilih Extensions lalu pilih Marketplace.



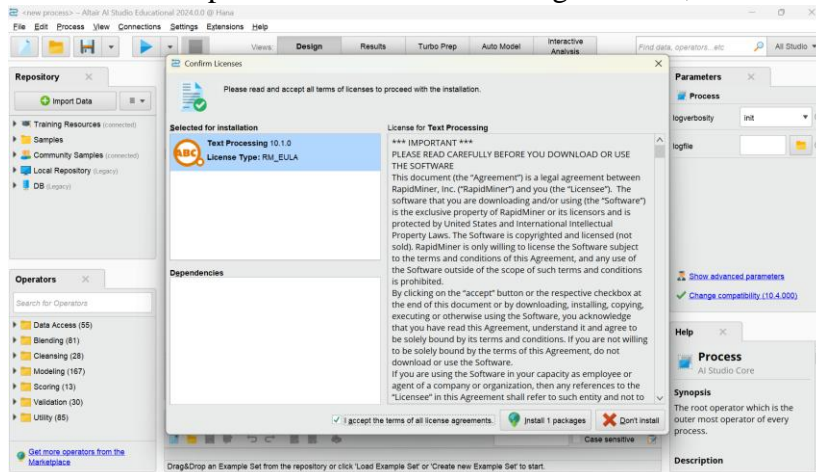
8. Lalu disini kita klik Top Downloads.



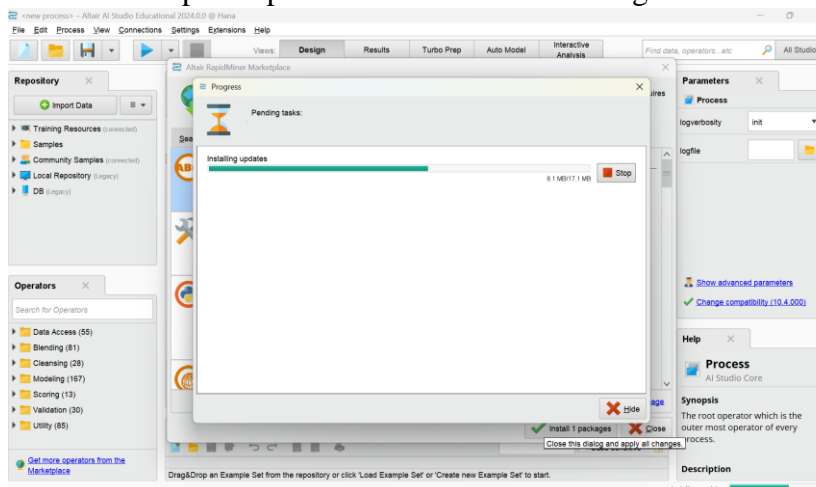
9. Lalu disini kita pilih Text Processing 10.10, lalu lakukan install.



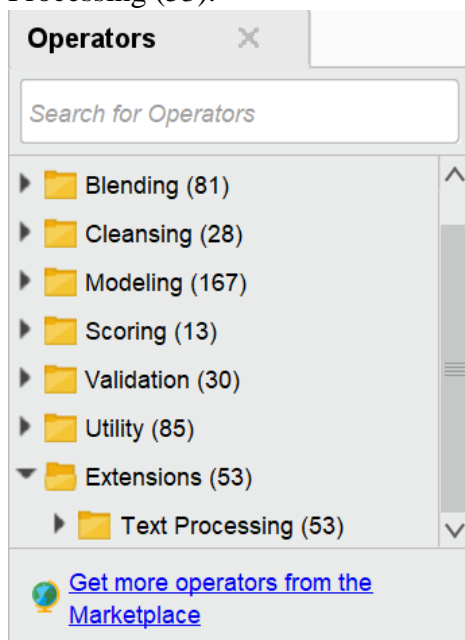
10. Lalu klik I accept the terms of all license agreements, lalu klik install 1 packages.



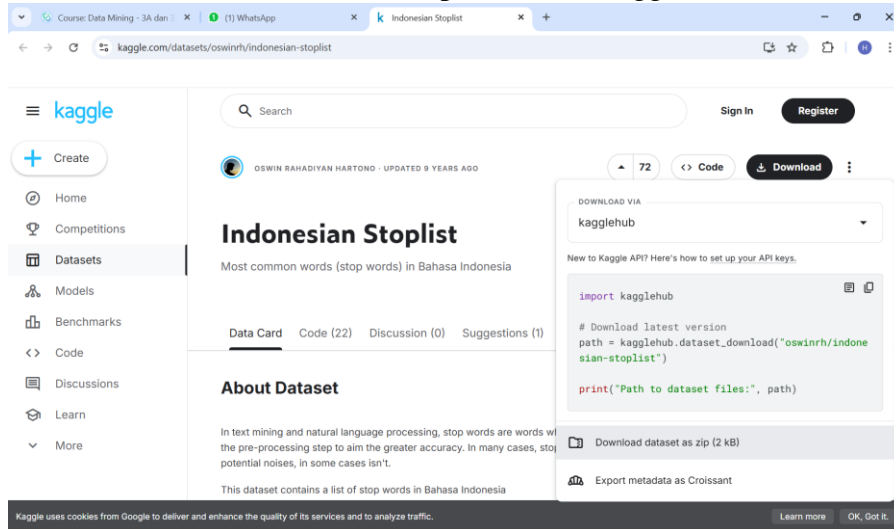
11. Lalu ini merupakan proses install Text Processing 10.10.



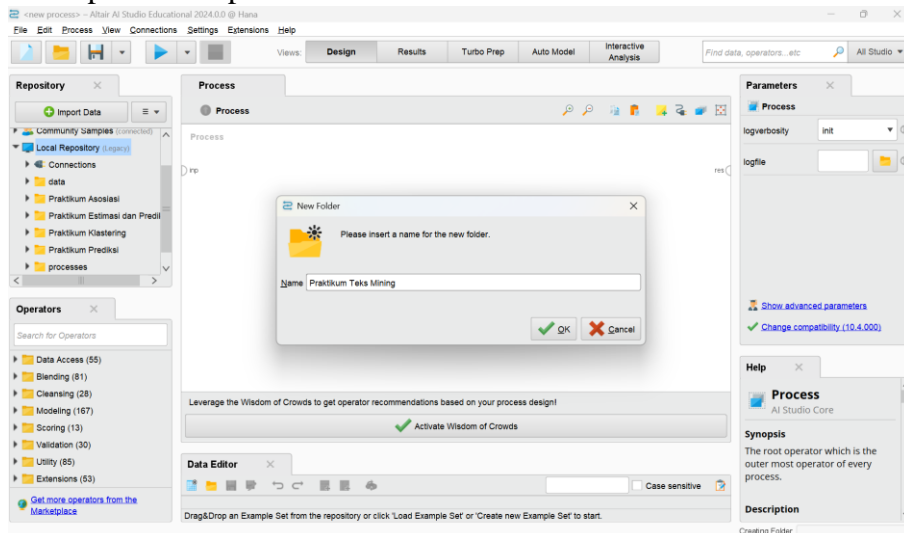
12. Lalu hasilnya sudah muncul di operator dimana Folder Extensions (53) serta Text Processing (53).



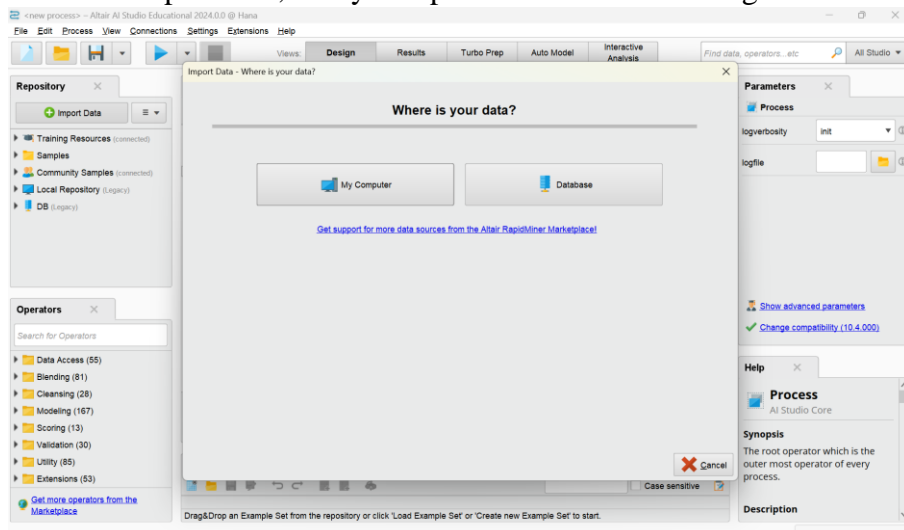
13. Lalu disini kita download dataset pada laman Kaggle.



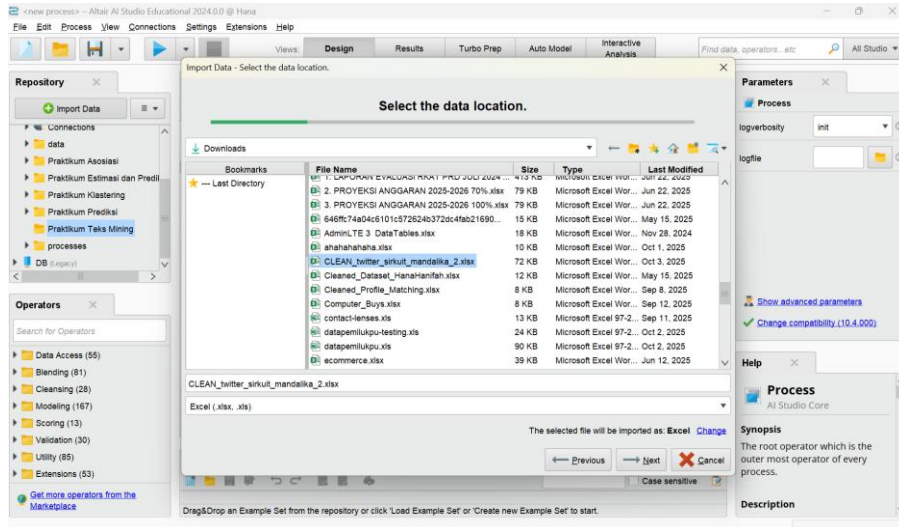
14. Lalu setelah download, disini saya membuat folder Praktikum Teks Mining agar data tersimpan dalam praktikum ini.



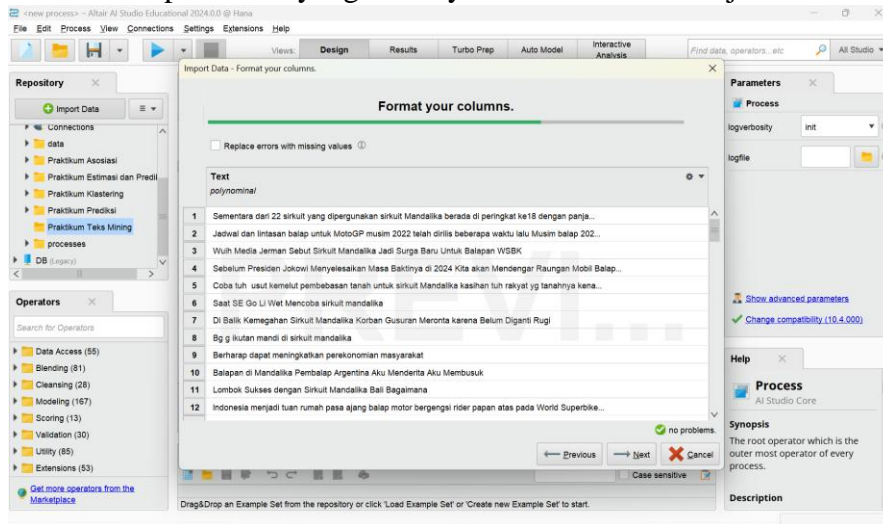
15. Lalu kita import data, di my computer dimana untuk mengolah data teks mining.



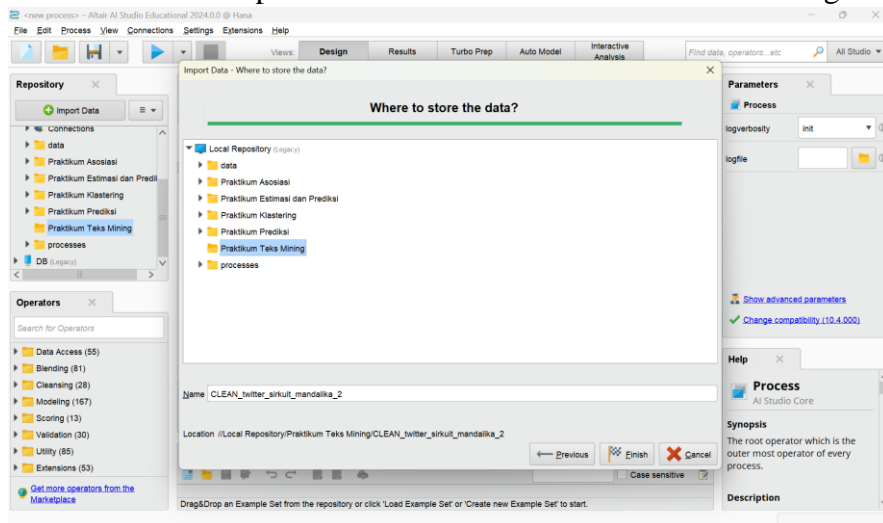
## 16. Lalu disini kita ambil data CLEAN\_twitter\_sirkuit\_mandalika\_2.xls



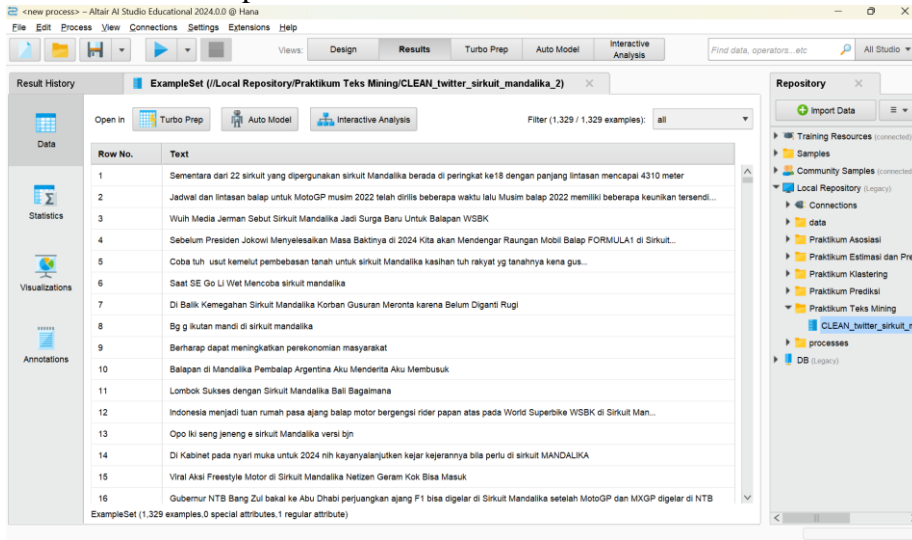
## 17. Lalu ini tampilan data yang nantinya akan kita olah menjadi teks mining.



## 18. Lalu disini kita simpan file tadi di folder Praktikum Teks Mining.

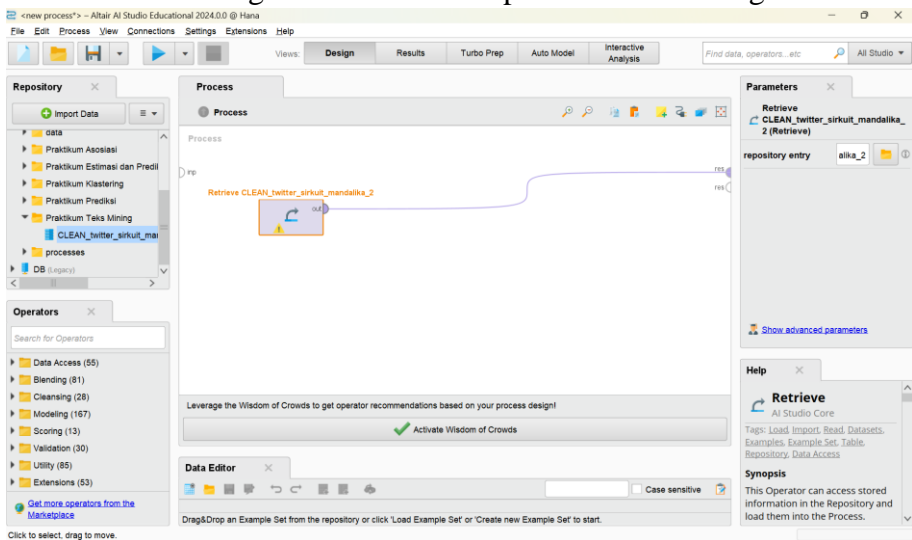


## 19. Lalu Ini hasil dari proses diatas.

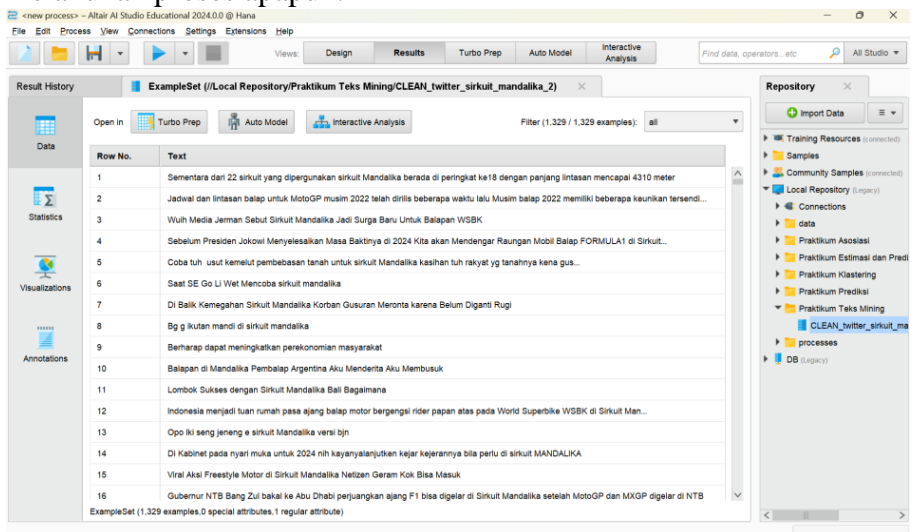


Row No.	Text
1	Sementara dari 22 sirkuit yang dipergunakan sirkuit Mandalika berada di peringkat ke18 dengan panjang lintasan mencapai 4310 meter
2	Jadwal dan lintasan balap untuk MotoGP musim 2022 telah dirilis beberapa waktu lalu Musim balap 2022 memiliki beberapa keunikan tersendiri...
3	Wuuh Media Jerman Sebut Sirkuit Mandalika Jadi Surga Baru Untuk Balapan WSBK
4	Sebelum Presiden Jokowi Menyelesaikan Masa Baktinya di 2024 Kita akan Mendengar Raungan Mobil Balap FORMULA1 di Sirkuit...
5	Coba tuh usut kemelut pembebasan tanah untuk sirkuit Mandalika kasihan tuh rakyat yg tanahnya kena gus...
6	Saat SE Go Li Wet Mencoba sirkuit mandalka
7	Di Balik Kemegahan Sirkuit Mandalika Korban Gusuhan Meronta karena Belum Digantri Rugi
8	Bg g iktan mandi di sirkuit mandalka
9	Berharap dapat meningkatkan perekonomian masyarakat
10	Balapan di Mandalika Pembalap Argentina Aku Mendenda Aku Membusuk
11	Lombok Sukses dengan Sirkuit Mandalika Bali Bagaimana
12	Indonesia menjadi tuan rumah pasha ajang balap motor bergengsi rider papan atas pada World Superbike WSBK di Sirkuit Man...
13	Opo iki seng jeneng e sirkuit Mandalika versi bjh
14	Di Kabinet pada nyari muka untuk 2024 nih kayanyalanjutkan keajaibannya bila perlu di sirkuit MANDALIKA
15	Viral Aksi Freestyle Motor di Sirkuit Mandalika Netizen Garam Kok Bisa Masuk
16	Gubernur NTB Bang Zul bakal ke Abu Dhabi perjuangkan ajang F1 bisa digelar di Sirkuit Mandalika setelah MotoGP dan MXGP digelar di NTB

## 20. Lalu disini kita drag file tadi ke dalam proses lalu sambungkan ke res.



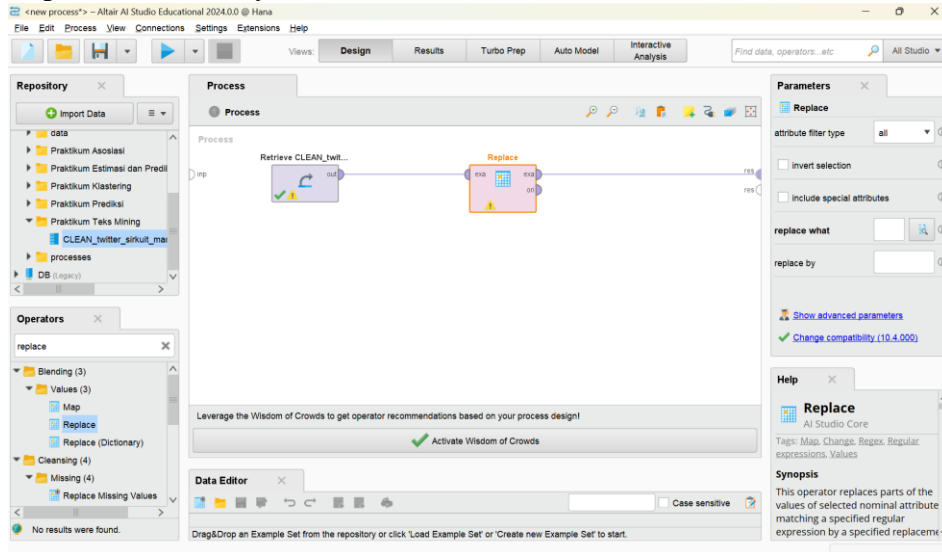
## 21. Lalu ini adalah hasil dari proses diatas, dimana data, masih sama karena kita belum melakukan proses apapun.



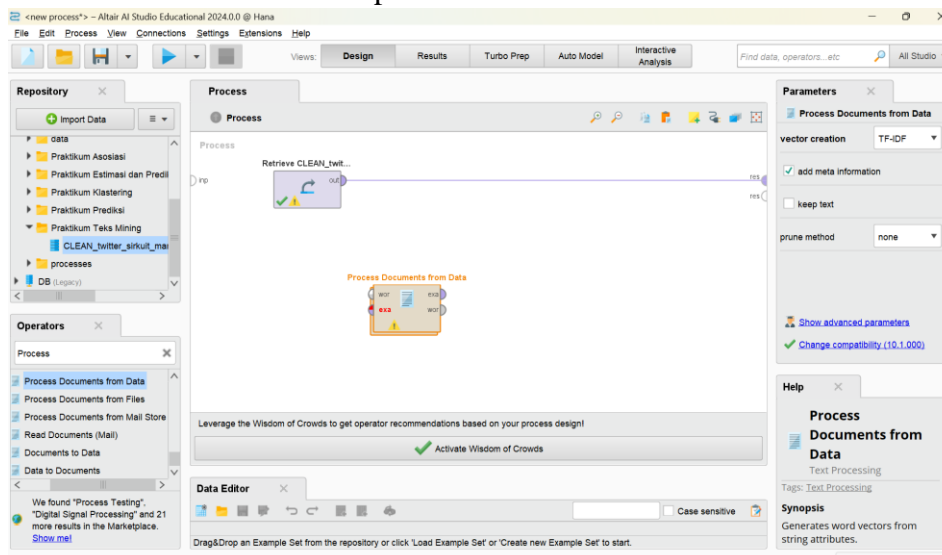
Row No.	Text
1	Sementara dari 22 sirkuit yang dipergunakan sirkuit Mandalika berada di peringkat ke18 dengan panjang lintasan mencapai 4310 meter
2	Jadwal dan lintasan balap untuk MotoGP musim 2022 telah dirilis beberapa waktu lalu Musim balap 2022 memiliki beberapa keunikan tersendiri...
3	Wuuh Media Jerman Sebut Sirkuit Mandalika Jadi Surga Baru Untuk Balapan WSBK
4	Sebelum Presiden Jokowi Menyelesaikan Masa Baktinya di 2024 Kita akan Mendengar Raungan Mobil Balap FORMULA1 di Sirkuit...
5	Coba tuh usut kemelut pembebasan tanah untuk sirkuit Mandalika kasihan tuh rakyat yg tanahnya kena gus...
6	Saat SE Go Li Wet Mencoba sirkuit mandalka
7	Di Balik Kemegahan Sirkuit Mandalika Korban Gusuhan Meronta karena Belum Digantri Rugi
8	Bg g iktan mandi di sirkuit mandalka
9	Berharap dapat meningkatkan perekonomian masyarakat
10	Balapan di Mandalika Pembalap Argentina Aku Mendenda Aku Membusuk
11	Lombok Sukses dengan Sirkuit Mandalika Bali Bagaimana
12	Indonesia menjadi tuan rumah pasha ajang balap motor bergengsi rider papan atas pada World Superbike WSBK di Sirkuit Man...
13	Opo iki seng jeneng e sirkuit Mandalika versi bjh
14	Di Kabinet pada nyari muka untuk 2024 nih kayanyalanjutkan keajaibannya bila perlu di sirkuit MANDALIKA
15	Viral Aksi Freestyle Motor di Sirkuit Mandalika Netizen Garam Kok Bisa Masuk
16	Gubernur NTB Bang Zul bakal ke Abu Dhabi perjuangkan ajang F1 bisa digelar di Sirkuit Mandalika setelah MotoGP dan MXGP digelar di NTB



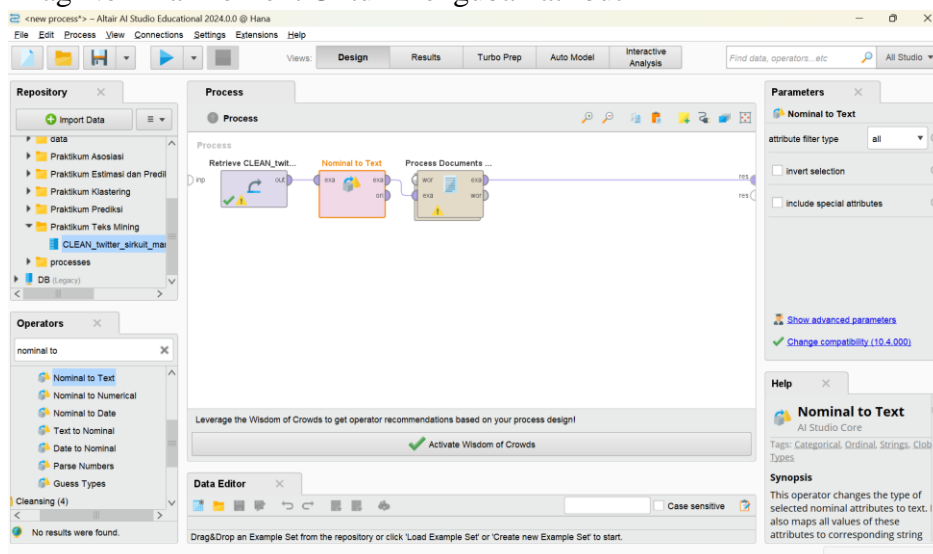
## 22. Replace untuk menyortir data.



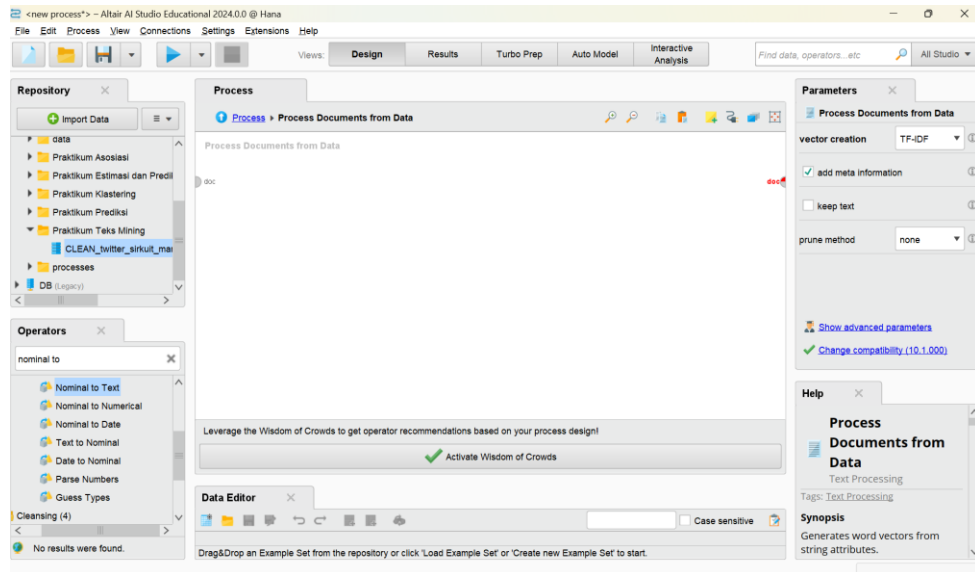
## 23. Lalu disini kita tambahkan operators Process Documents From Data



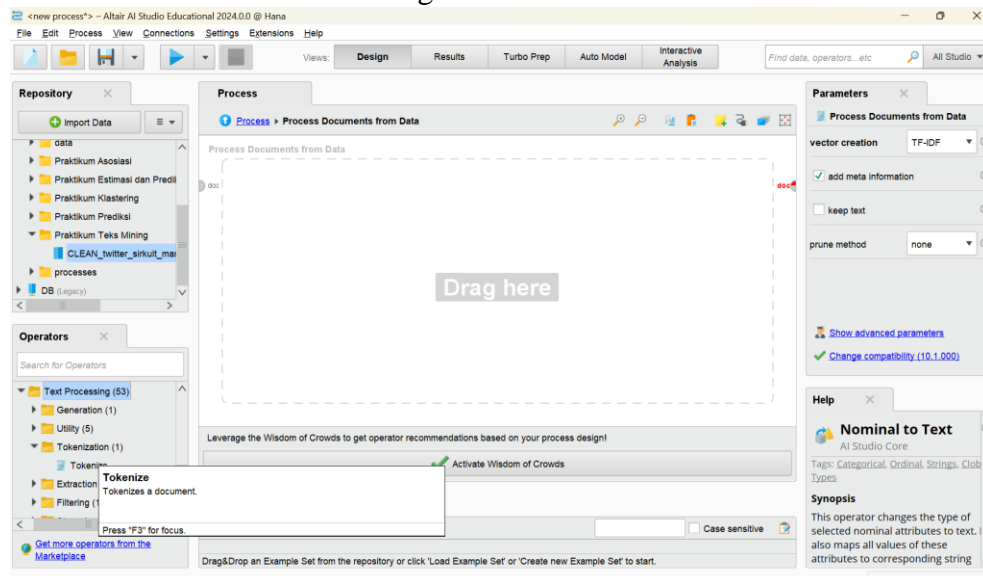
## 24. Drag Nominal To Text Untuk mengubah atribut



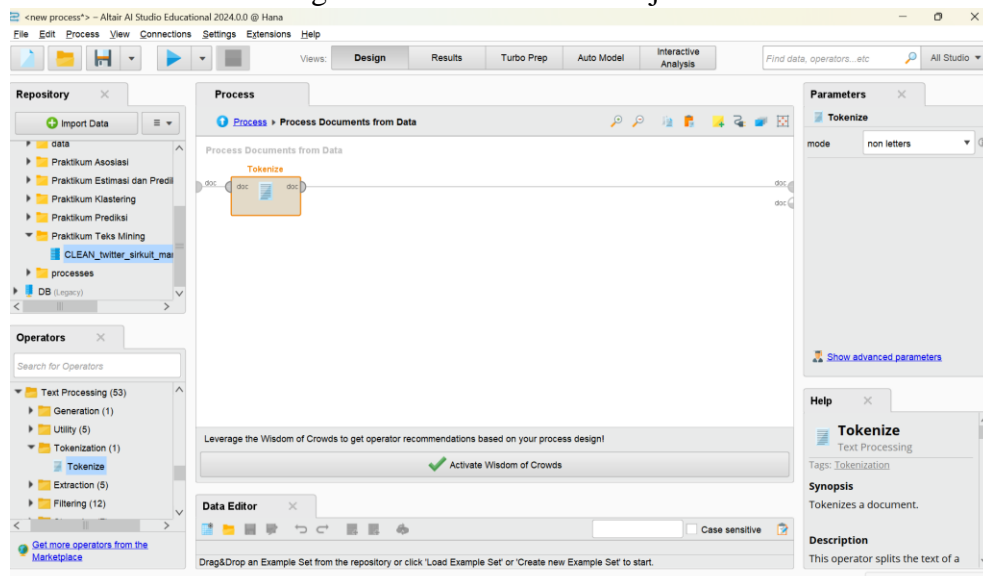
## 25. Double Klik



## 26. Tokenize untuk memisah”kata.



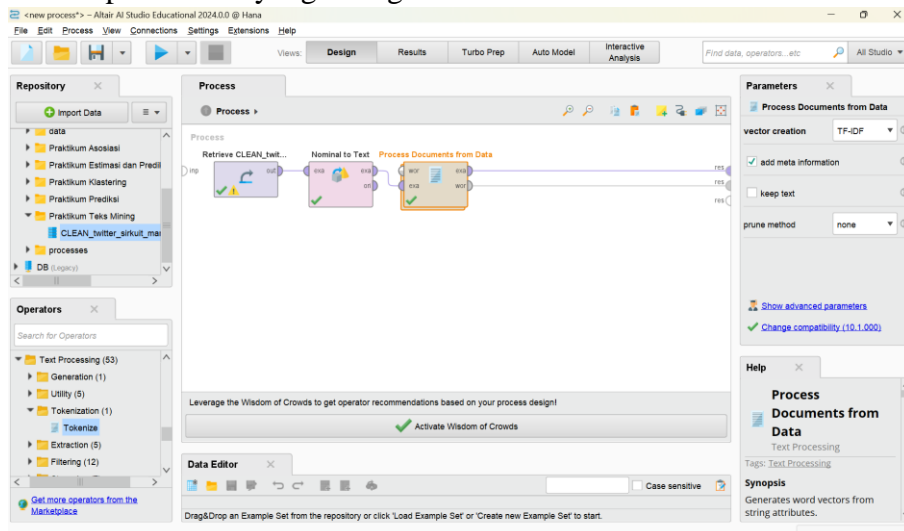
## 27. Lalu disini kita sambungkan. Lalu setelah itu kita jalankan



28. Lalu ini adalah hasil dari proses Tokenize dimana berfungsi untuk proses memecah urutan teks menjadi unit-unit yang lebih kecil yang disebut token. Token dapat berupa kata, frasa, atau symbol.

Row No.	A	ABS	ADA	ADAH	ADV	AJA	AJANG	AK
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0

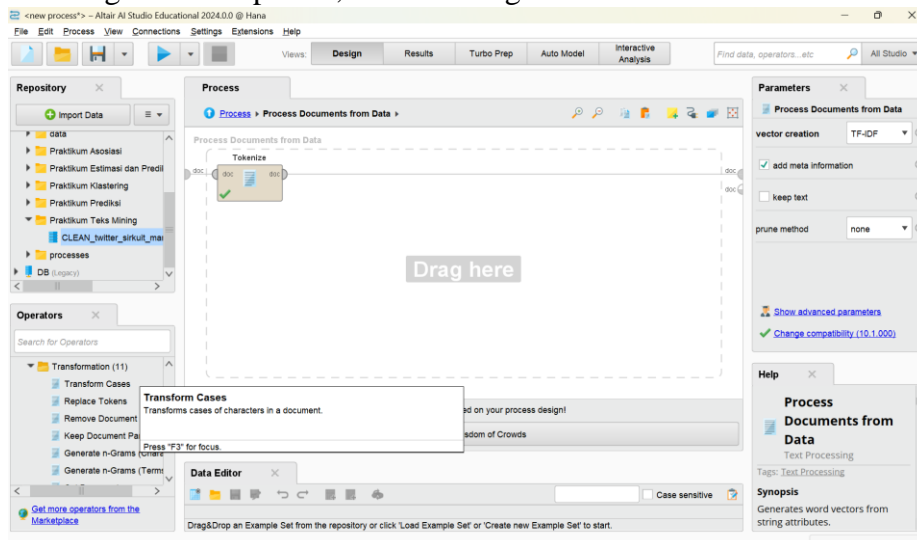
29. Menampilkan kata yang sering muncul



30. Lalu ini adalah hasil dari Tokenize.

Word	Attribute Name	Total Occurrences	Document Occurrences
A	A	1	1
ABS	ABS	2	1
ADA	ADA	2	2
ADAH	ADAH	1	1
ADV	ADV	1	1
AJA	AJA	1	1
AJANG	AJANG	1	1
AK	AK	1	1
ALLAH	ALLAH	1	1
AMBIL	AMBIL	1	1
ANTARA	ANTARA	2	2
ANTI	ANTI	1	1
ANYWAY	ANYWAY	1	1
APBN	APBN	1	1
ASPAL	ASPAL	2	1

31. Lalu ini double klik pada file document. Dan tambahkan operators Transform Cases lalu drag ke dalam proses, dan sambungkan ke Tokenize.

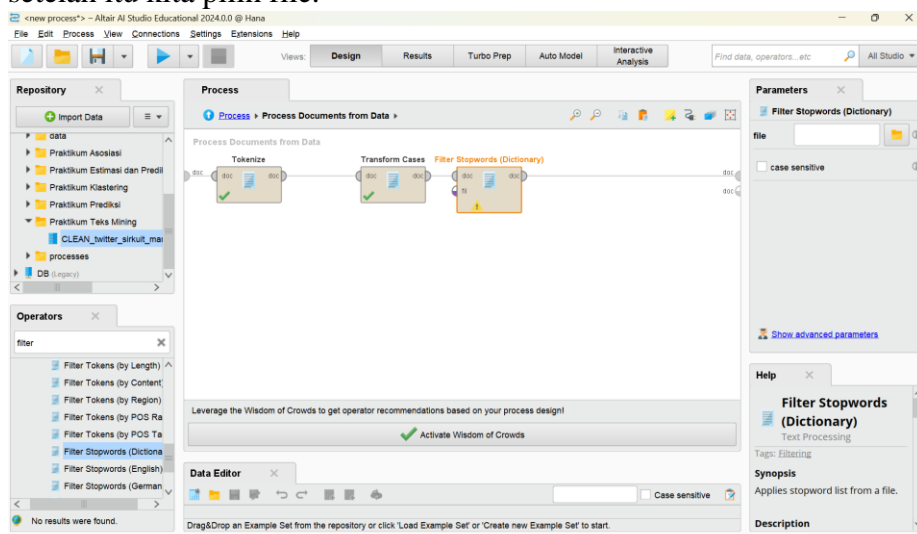


32. Lalu ini hasilnyaaa Proses Transform Cases.

The screenshot shows the 'Result History' panel in Altair AI Studio. It displays a table with the following data:

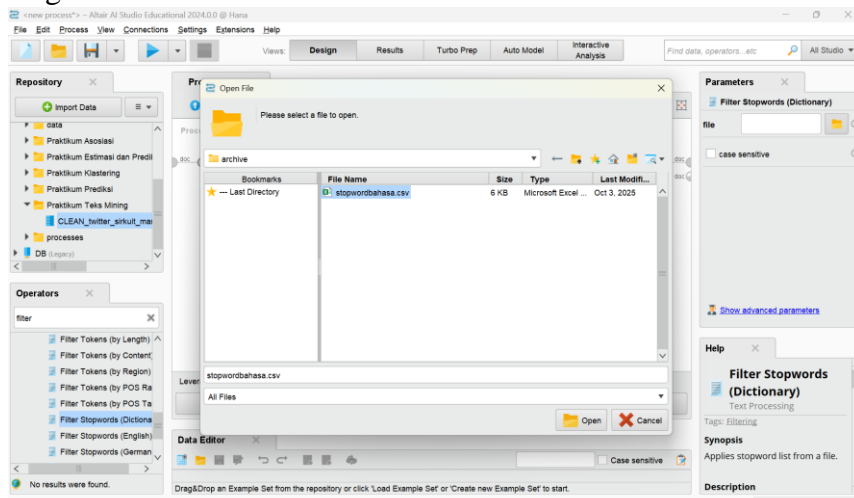
Word	Attribute Name	Total Occurrences	Document Occurrences
a	a	2	2
aamin	aamin	2	2
aan	aan	1	1
abal	abal	2	2
abang	abang	1	1
abis	abis	1	1
abiss	abiss	1	1
abs	abs	2	1
abu	abu	1	1
aburizal	aburizal	1	1
acara	acara	8	8
acungi	acungi	1	1
ad	ad	1	1
ada	ada	175	147
adaada	adaada	1	1

33. Lalu kita tambahkan lagi operators Filter Stopwords (Dectionary). Lalu klik pada setelah itu kita pilih file.

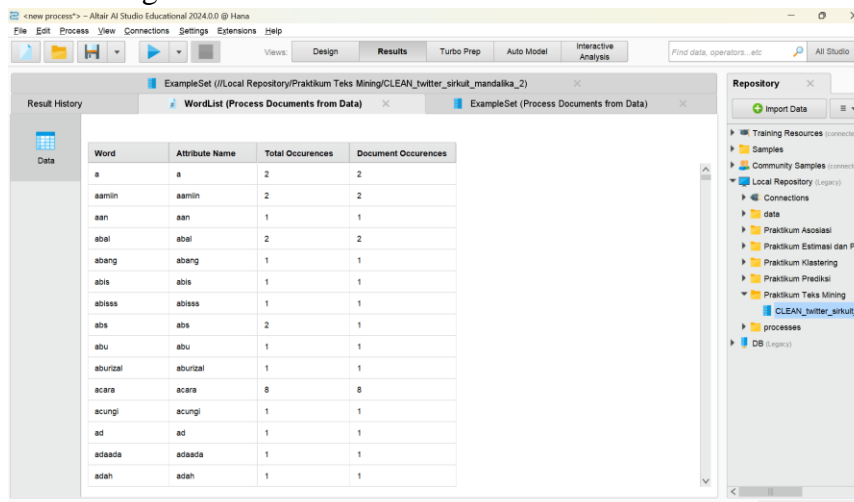




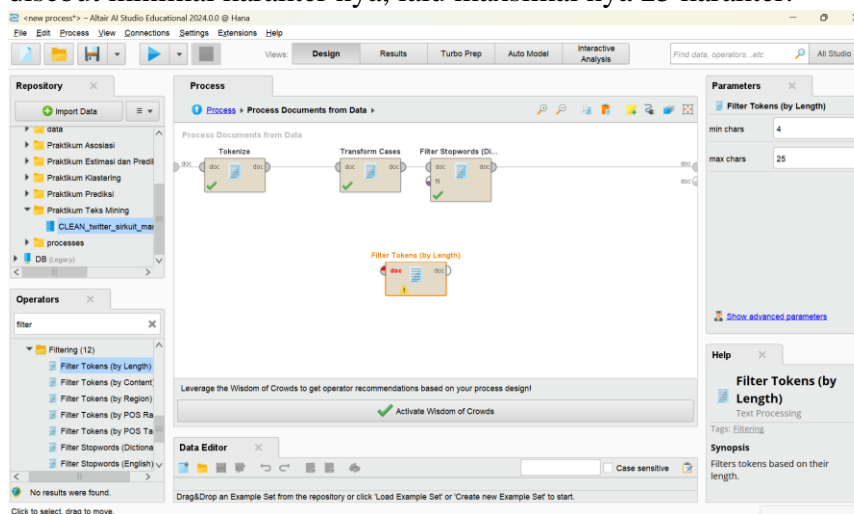
34. Lalu disini kita pilih file stopwordsbahasa.csv untuk menampilkan data yang tidak di singkat.



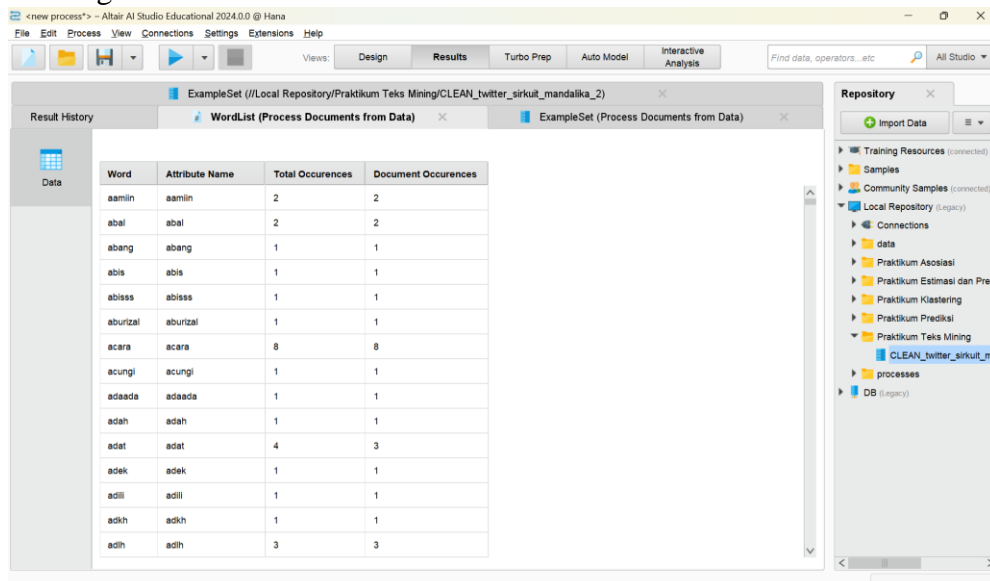
35. Lalu ini sudah muncul disini kita bisa lihat bahwa data yang muncul itu data yang sudah tidak disingkat.



36. Tahap Pre Processing untuk nge filter. Dimana kita memfilter min chars nya 4 atau bisa disebut minimal karakter nya, lalu maksimal nya 25 karakter.



37. Maka hasilnya akan menampilkan data yang minimal nya 4 maximal nya 25. Jika data melebihi 25 dan kurang dari 4 karakter maka data tidak akan dimunculkan atau dihilangkan oleh sistem.



The screenshot shows the Altair AI Studio interface. The main window displays the 'Results' tab for a process named 'WordList (Process Documents from Data)'. The results are shown in a table with the following data:

Word	Attribute Name	Total Occurences	Document Occurences
aamin	aamin	2	2
abal	abal	2	2
abang	abang	1	1
abis	abis	1	1
abiss	abiss	1	1
aburizal	aburizal	1	1
acara	acara	8	8
acungi	acungi	1	1
adaada	adaada	1	1
adah	adah	1	1
adat	adat	4	3
adek	adek	1	1
adili	adili	1	1
adikh	adikh	1	1
adilh	adilh	3	3

The interface also shows a 'Repository' panel on the right with various data sources and a 'Result History' panel on the left.

## Kesimpulan

Kesimpulan yang saya dapat dari praktikum ini adalah bahwa proses teks mining berhasil dilakukan dengan memanfaatkan oleh software Altair AI Studio dimana kita di dalam nya meng install Ekstensi Teks Processing untuk mengolah dataset dari Kaggle. Melalui tahapan pre-processing seperti Tokenize, Transform Cases, Lalu Filter Stopwords, serta pembatasan jumlah karakter. Dimana data mentah yang tadinya tidak terstruktur menjadi sebuah data yang terstruktur, bersih, rapi, dan siap di analisis. Hasil akhir dari praktikum ini adalah menunjukan bahwa penerapan teknik Teks Processing sangat penting dalam menyiapkan data teks agar dapat digunakan untuk analisis lanjutan, misal dalam menemukan sebuah kata yang sering muncul atau melakukan analisis sentiment dari suatu informasi atau data.