

## **Projek Akhir Teknologi Big Data**

**“Analisis Big Data Pada Media Sosial (Twitter) Untuk Analisis Gempa  
Bumi Di Dunia”**



*Risma Dayanti*

*1901020030*

**Teknologi Informasi**

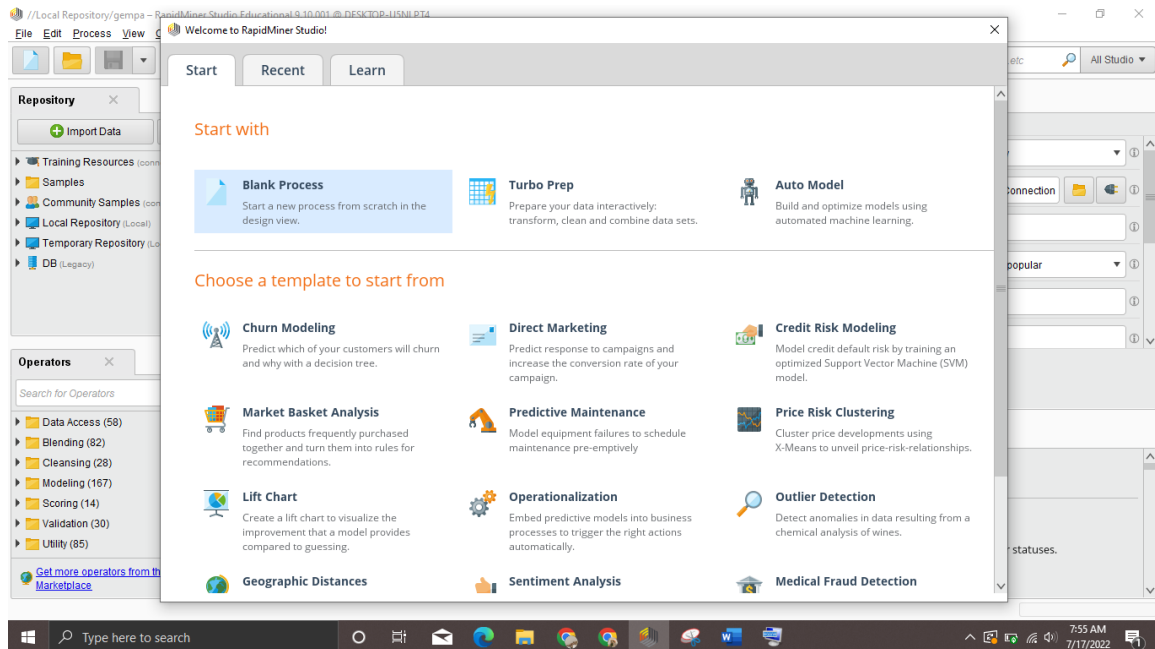
**Fakultas Teknik**

**Universitas Bumigora**

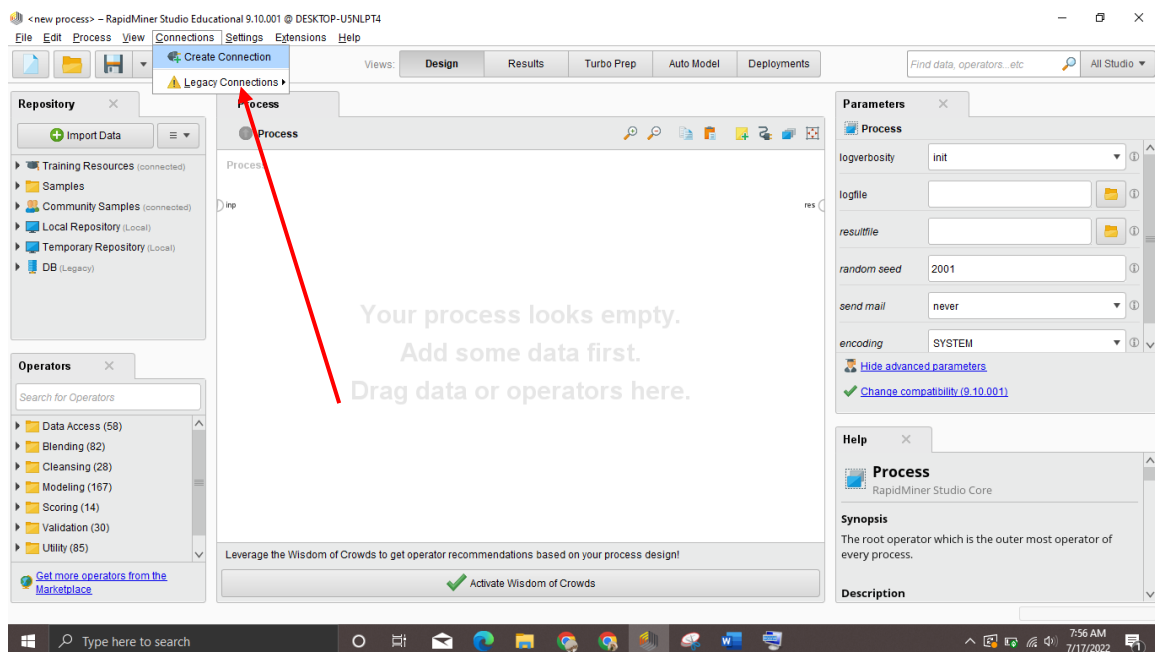
**2022**

# Proses Crawling Data Dari Twitter Menggunakan Tools Rapidminer

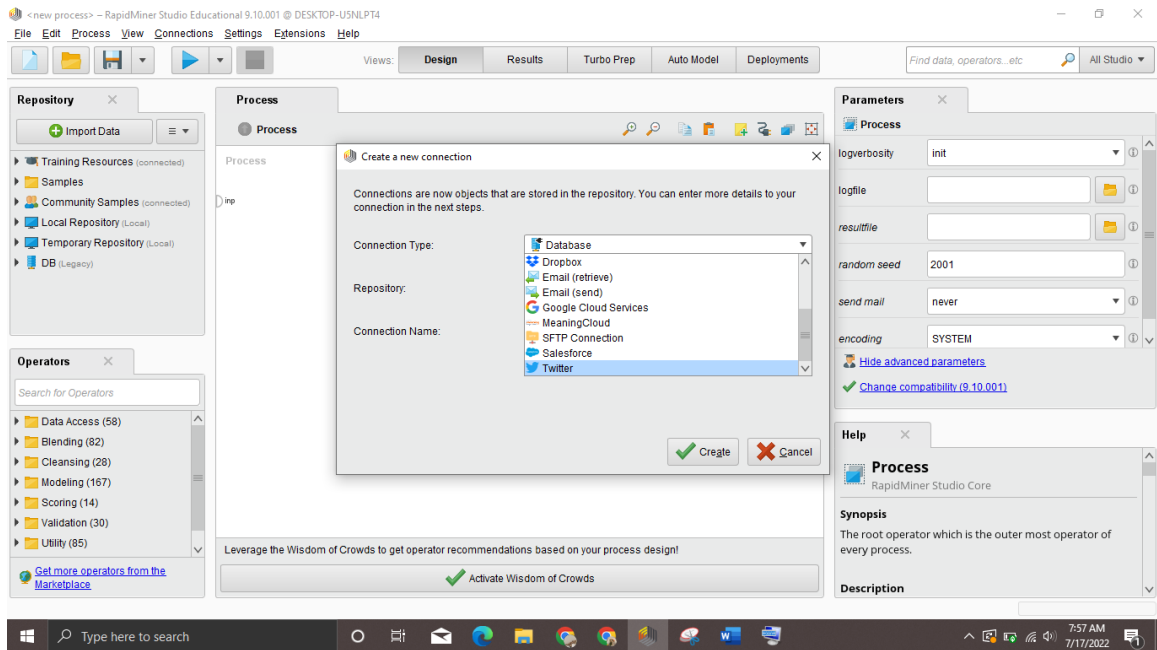
1. Membuat proses baru dengan mengklik blank proses.



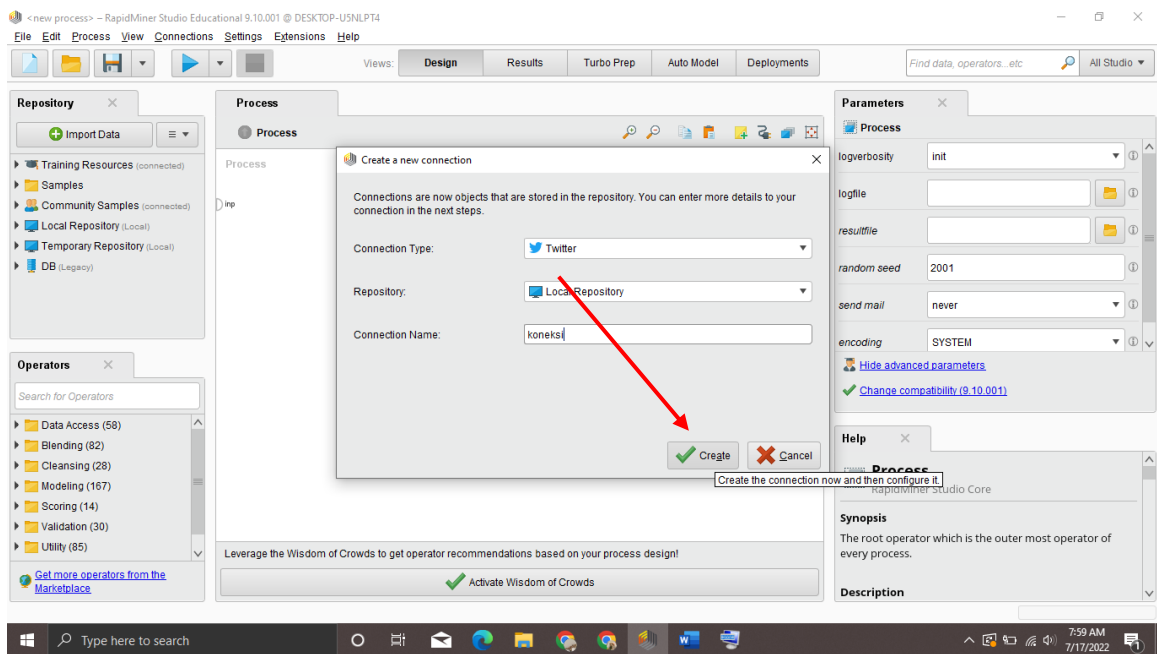
2. Membuat koneksi untuk menghubungkan akun Twitter dengan RapidMiner. Pilih “Connection” lalu klik “Create Connection”.



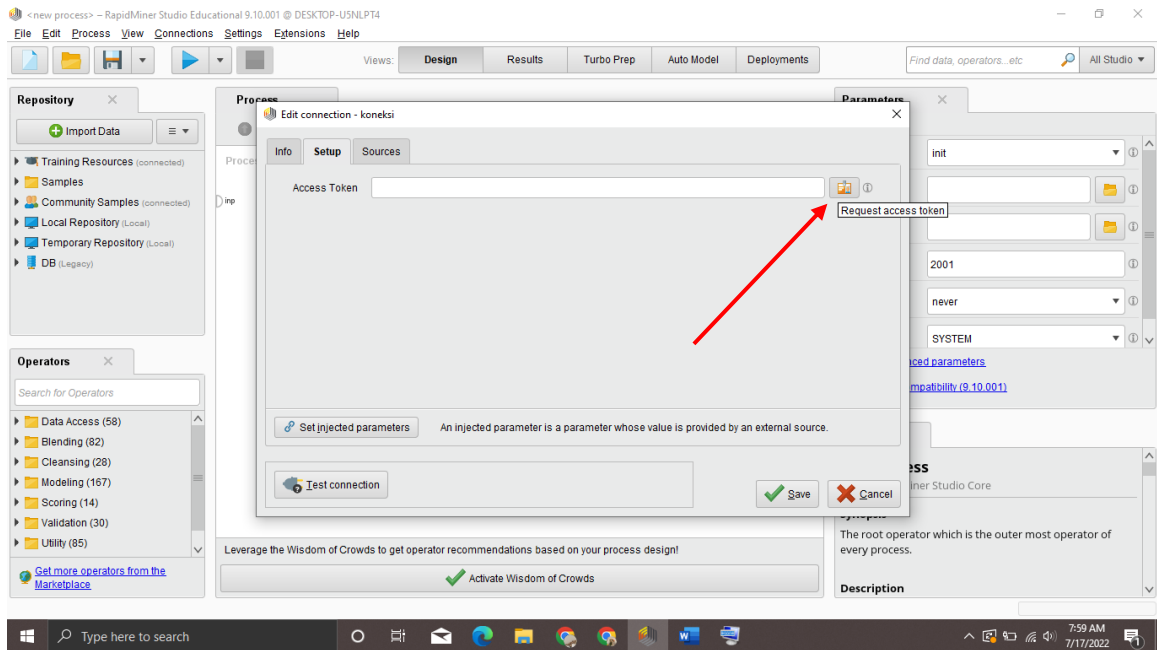
### 3. Pada database pilih Twitter.



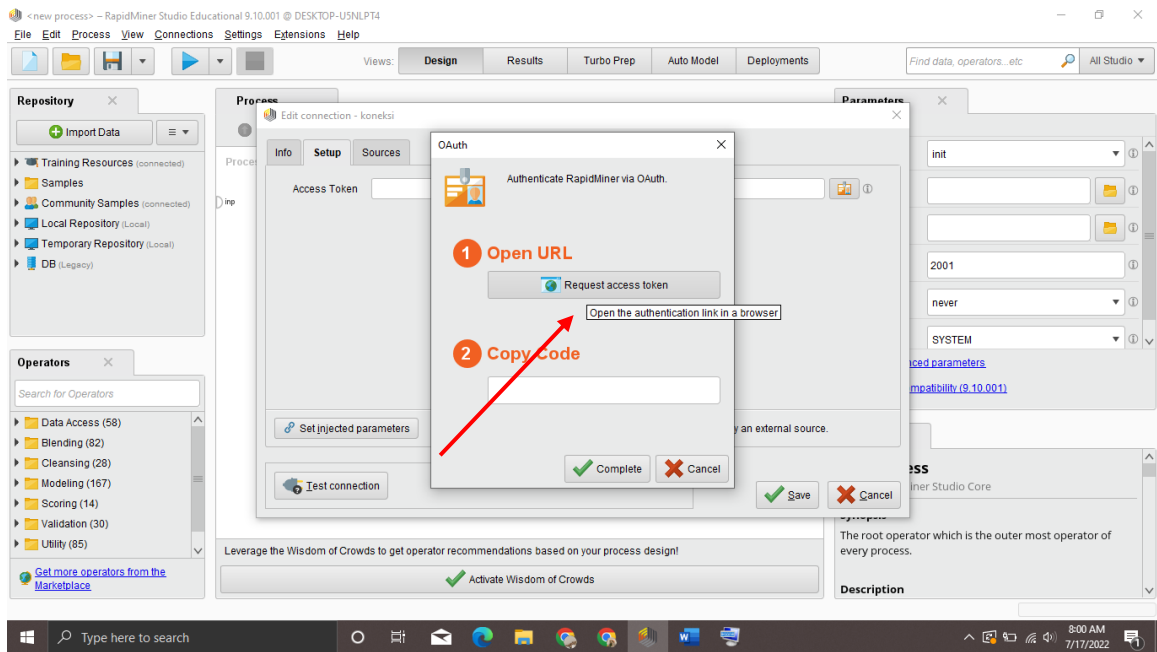
### 4. Kemudian pada Connection Name diisi sesuai dengan nama koneksi yang diinginkan, misalnya disini “koneksi”. Lalu pilih “Create”.



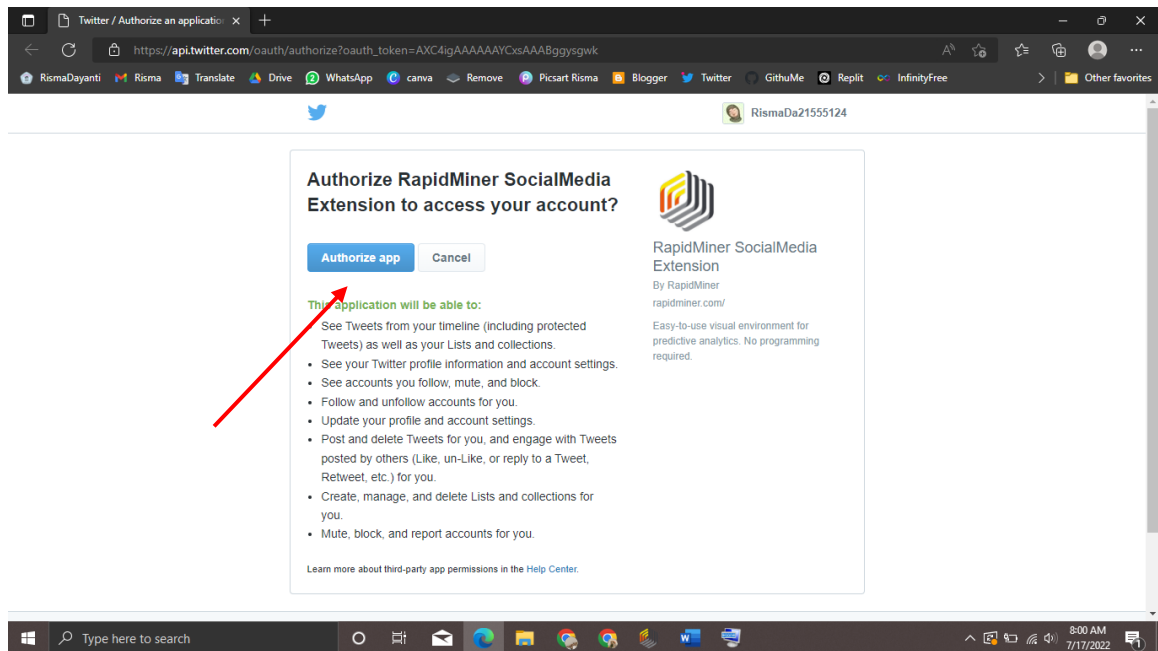
5. Setelah itu pada Access Token pilih kotak kuning.



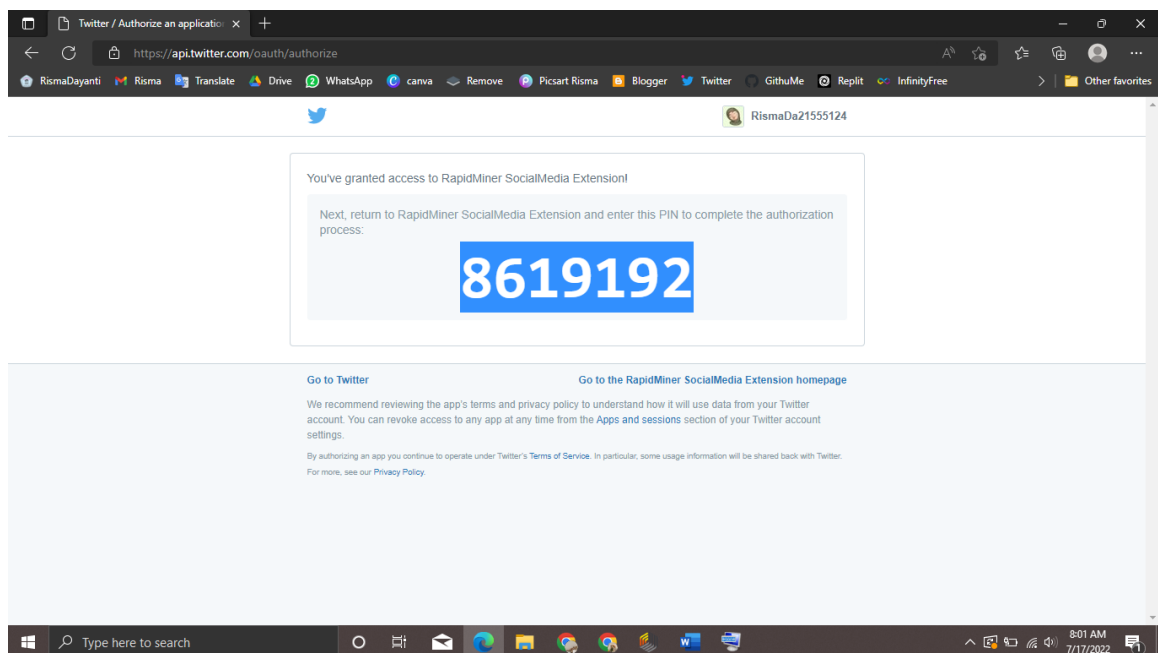
6. Kemudian akan muncul kotak Auth. Pada Open URL klik Request access token.



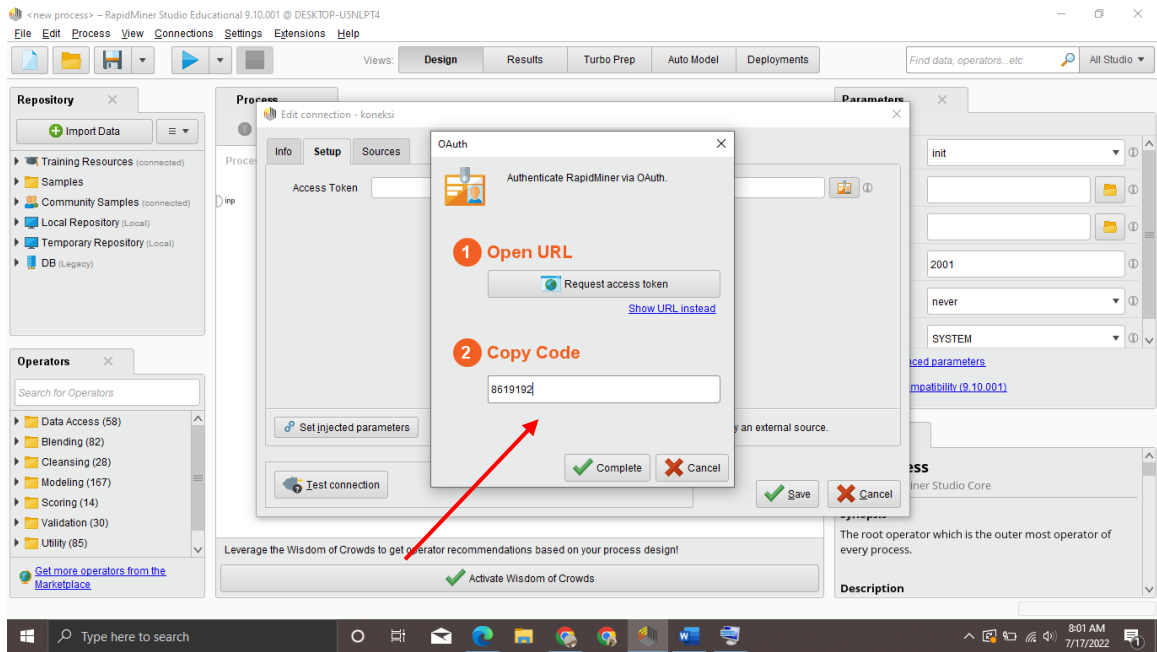
7. Lalu akan diarahkan ke halaman web berikut. Pilih Authomize app untuk mendapatkan kode.



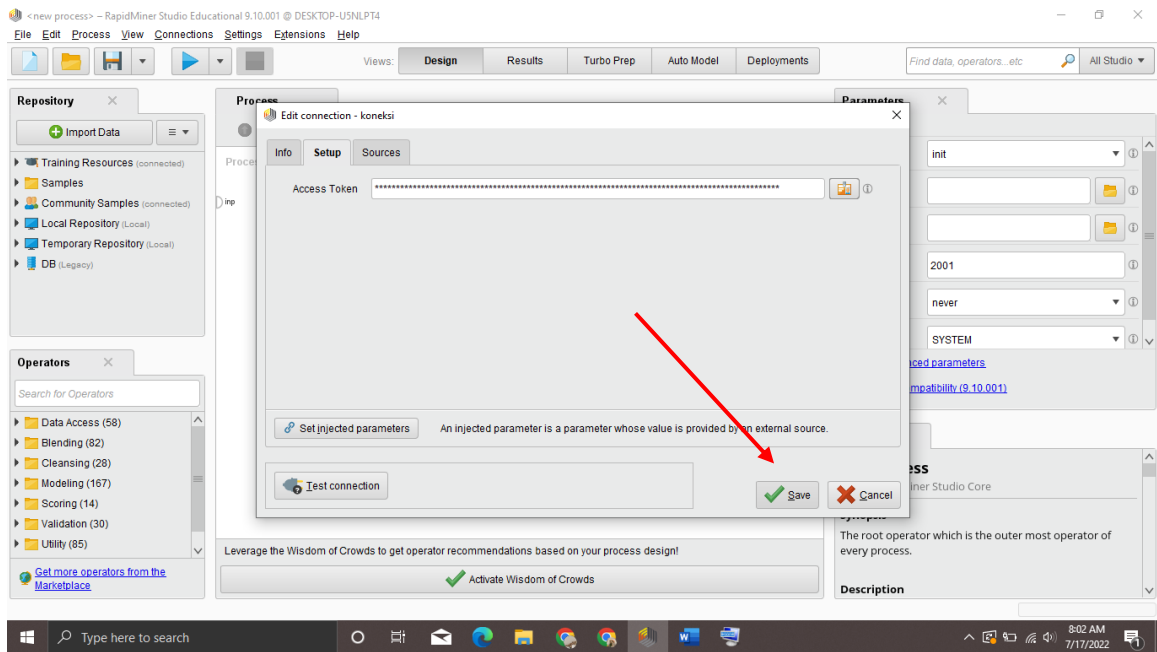
8. Setelah mendapatkan kode authomize, copy code tersebut.



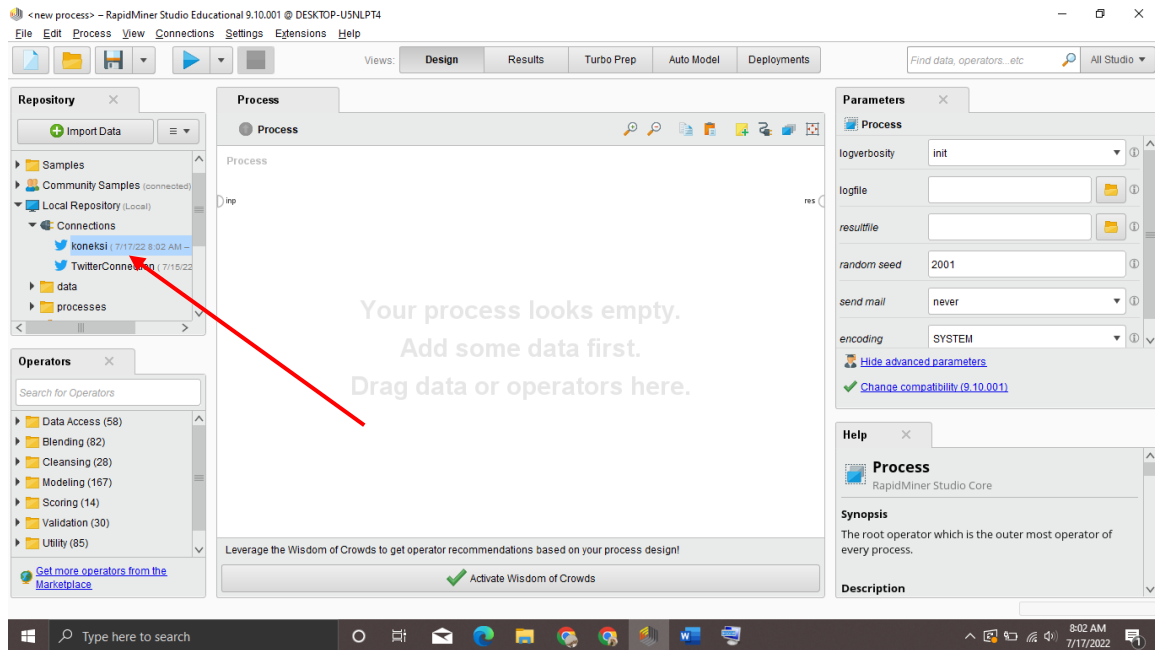
9. Kemudian paste pada box “Copy Code”. Lalu pilih Complete



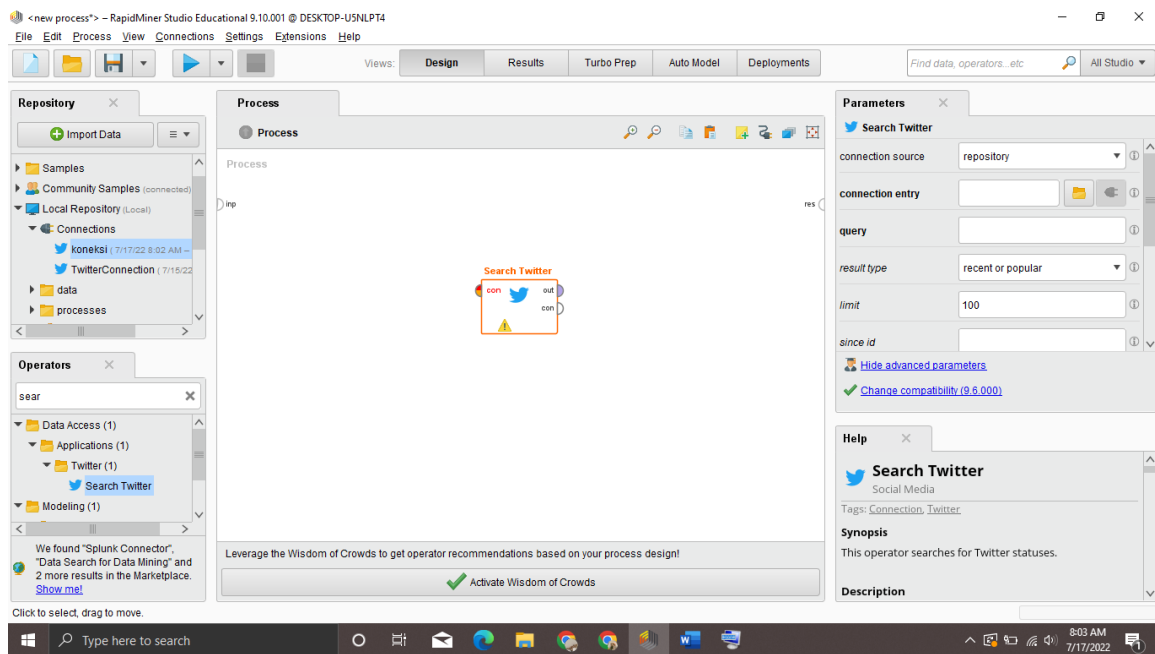
10. Setelah itu, secara otomatis access token akan terisi seperti berikut. Kemudian pilih “Save”.



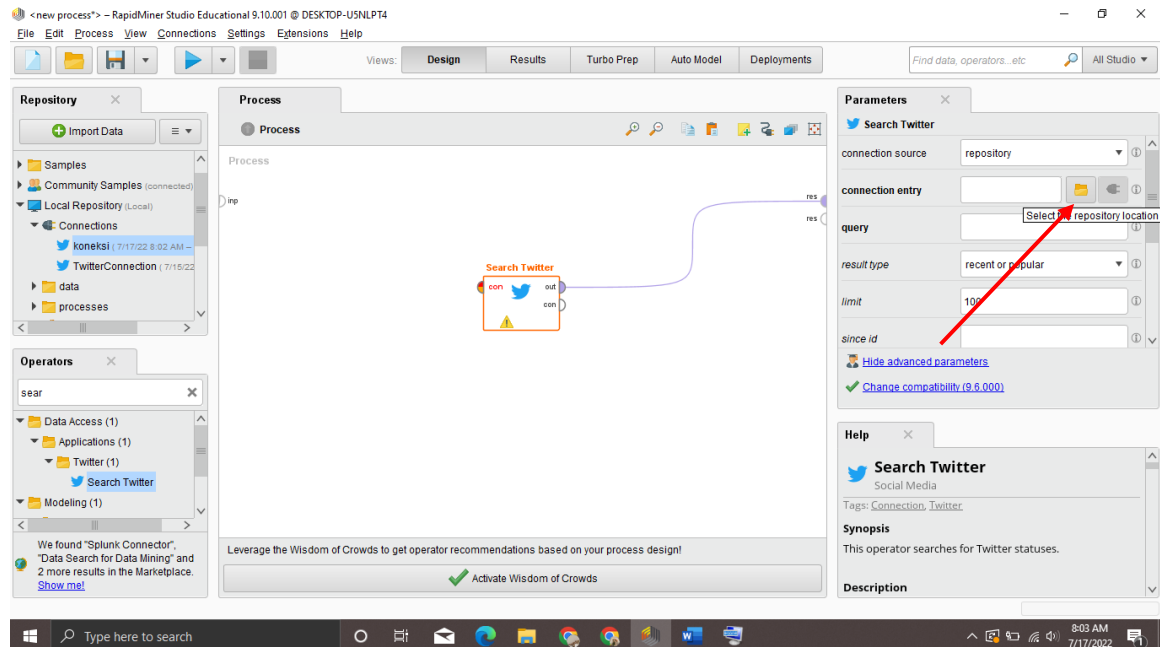
11. Koneksi yang sudah dibuat dapat dilihat pada Repository.



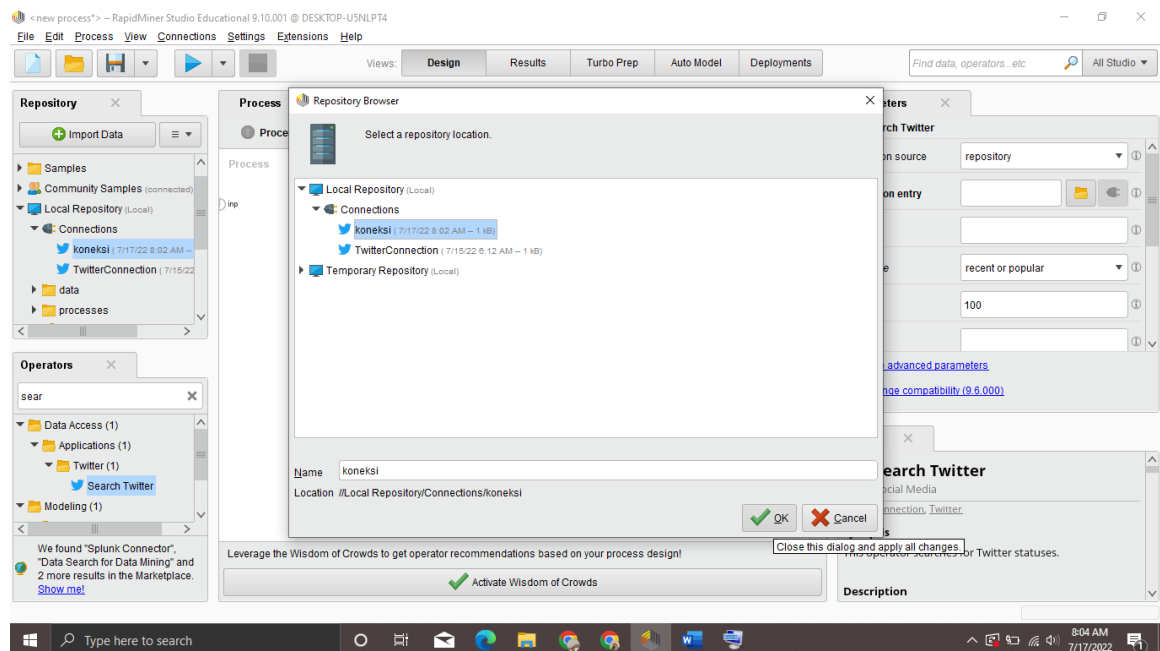
12. Selanjutnya pada bagian Operator ketik “Search Twitter”, kemudian drag and drop pada kotak Process seperti gambar berikut.



13. Pilih atribut Search Twitter, lalu pada “Parameters” pilih kotak kuning untuk menambahkan Twitter.

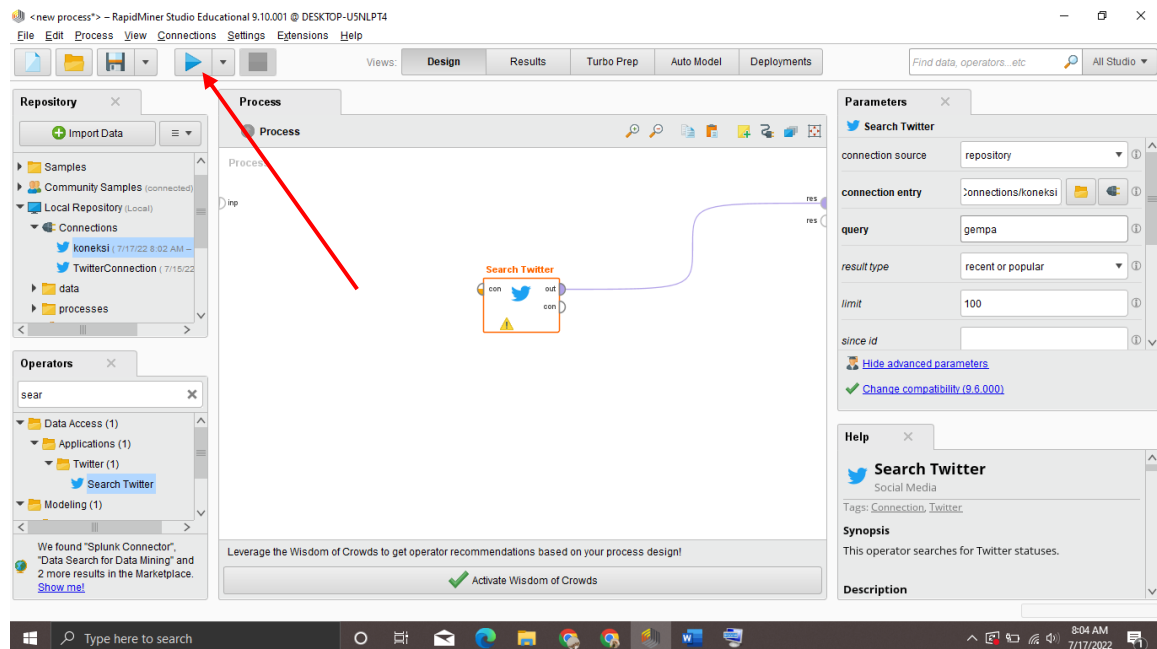


14. Masukkan koneksi yang sudah dibuat sebelumnya. Lalu klik “Ok”.

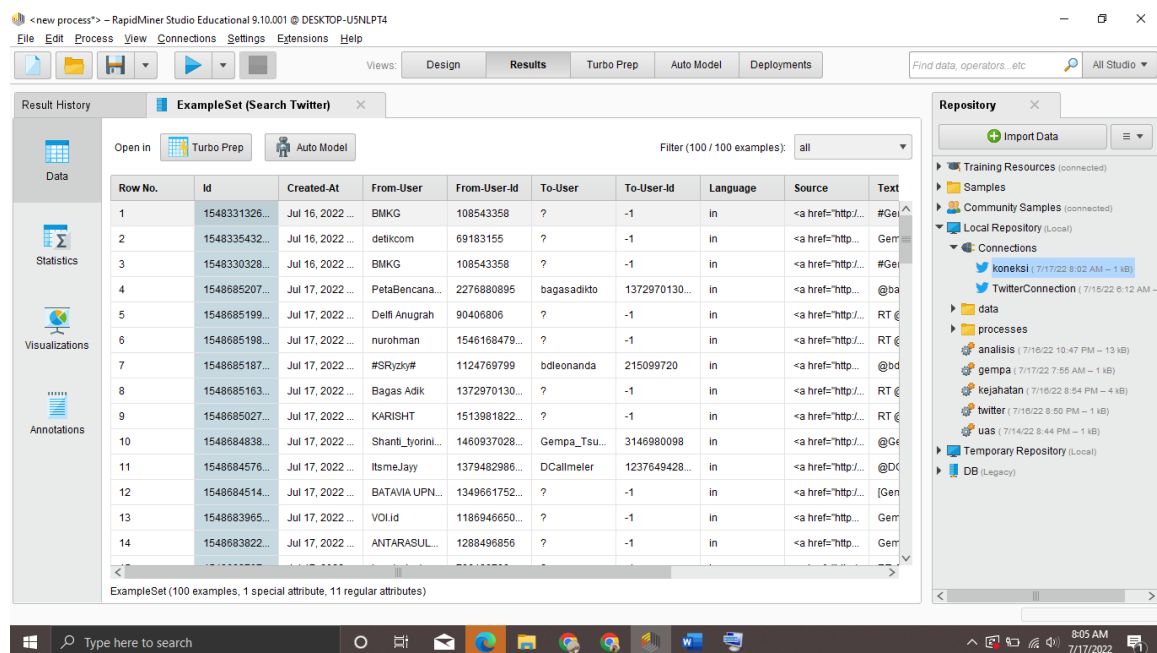




15. Setelah itu, masukkan kata kunci pada “query”, misalkan query yang ingin di ambil ialah gempa. Lalu jalankan prosesnya.



16. Berikut tampilan jika sudah dijalankan.



17. Selanjutnya untuk mengekspor data menjadi csv ialah dengan memilih “Turbo Prep” lalu pada titik tiga disisi kiri pilih “Export”.

The screenshot shows the RapidMiner Studio Turbo Prep interface. The 'Data from Process' section displays a table of data. The 'EXPORT' button is highlighted with a red arrow. The 'Data from Process' section also shows a table of data.

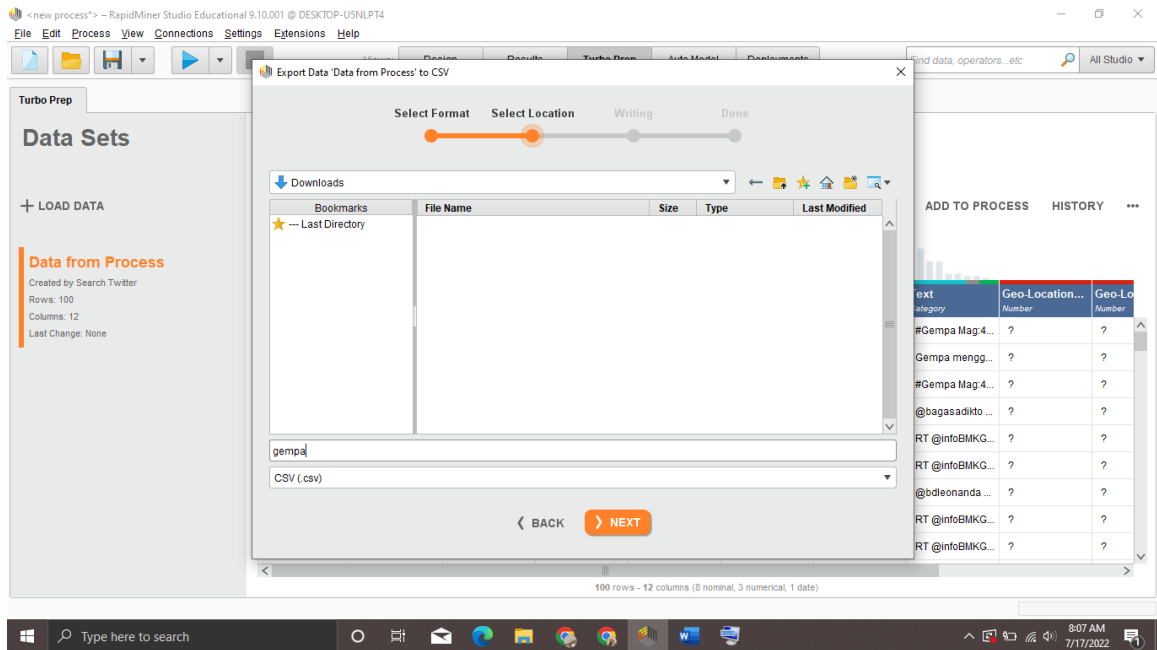
Created At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text	Geo-Location...
Date / Time	Category	Category	Category	Category	Category	Category	Category	Number
Jul 16, 2022 8:...	BMKG	108543358	?	-1	in	<a href="http://i...	#Gempa Mag 4...	?
Jul 16, 2022 8:...	detikcom	69183155	?	-1	in	<a href="https://...	Gempa meng...	?
Jul 16, 2022 8:...	BMKG	108543358	?	-1	in	<a href="http://i...	#Gempa Mag 4...	?
Jul 17, 2022 8:...	PetaBencana.id	2276880895	bagasadikto	137297013024...	in	<a href="https://...	@bagasadikto ...	?
Jul 17, 2022 8:...	Detik Anugrah	90406806	?	-1	in	<a href="http://i...	RT @infoBMKG...	?
Jul 17, 2022 8:...	nurohman	154616847948...	?	-1	in	<a href="http://i...	RT @infoBMKG...	?
Jul 17, 2022 8:...	#SRjzky#	1124769799	bdleonanda	215099720	in	<a href="https://...	@bdleonanda ...	?
Jul 17, 2022 8:...	Bagas Adik	137297013024...	?	-1	in	<a href="http://i...	RT @infoBMKG...	?
Jul 17, 2022 8:...	KARISHT	151398182252...	?	-1	in	<a href="http://i...	RT @infoBMKG...	?

18. Selanjutnya pilih format yang diinginkan. Disini kita pilih csv, kemudian next.

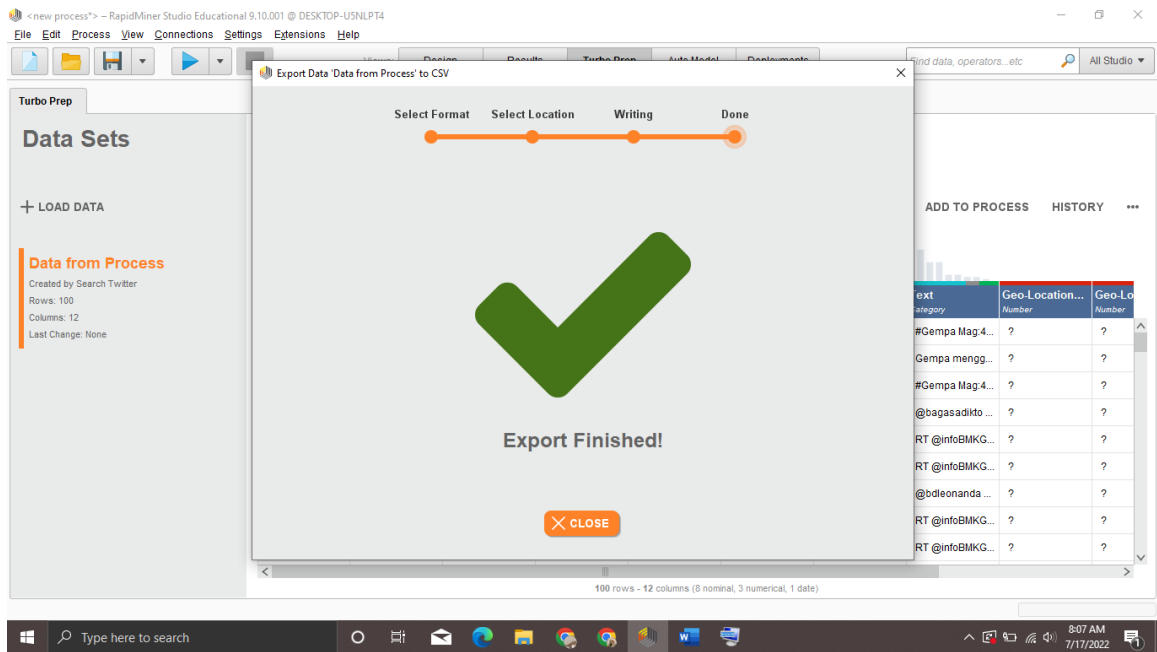
The screenshot shows the RapidMiner Studio Turbo Prep interface with the 'Export Data' dialog box open. The dialog box has four options: Repository, Qlik, Excel, and CSV. The CSV option is selected, and the 'NEXT' button is highlighted.

Text	Geo-Location...	Geo-Lo
Category	Number	Number
#Gempa Mag 4...	?	?
Gempa meng...	?	?
#Gempa Mag 4...	?	?
@bagasadikto ...	?	?
RT @infoBMKG...	?	?
RT @infoBMKG...	?	?
@bdleonanda ...	?	?
RT @infoBMKG...	?	?
RT @infoBMKG...	?	?

19. Pilih lokasi penyimpanan, kemudian create nama filenya. Pilih next.

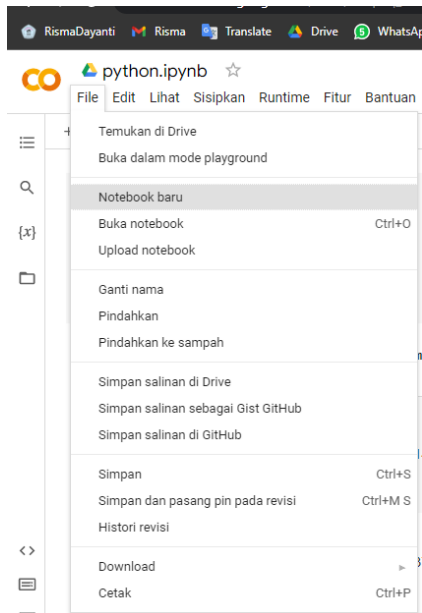


20. Terakhir pilih close. Periksa file csv yang disimpan sebelumnya.

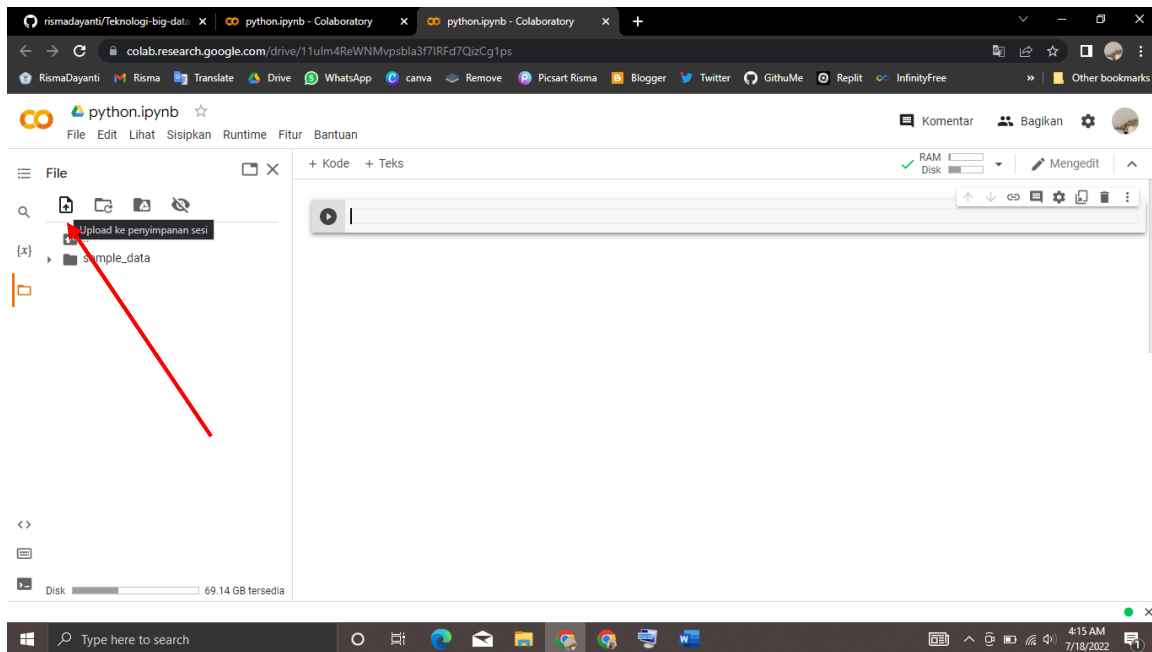


# Proses Import dan Analisis Data Menggunakan Python

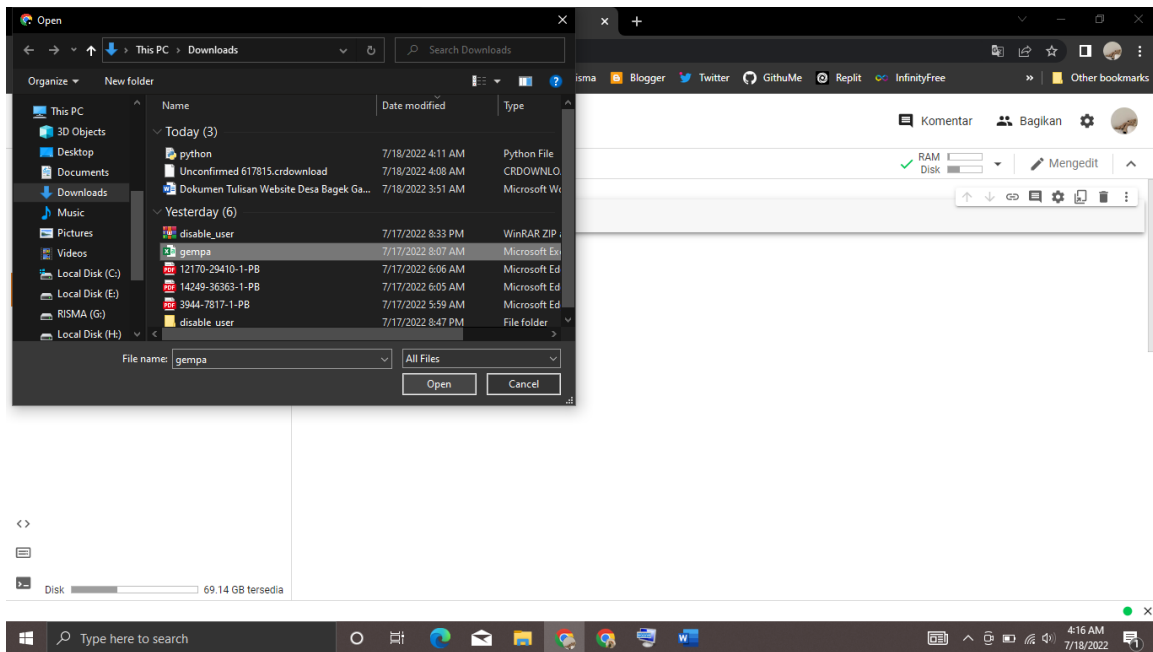
## 1. Membuat Netbook baru.



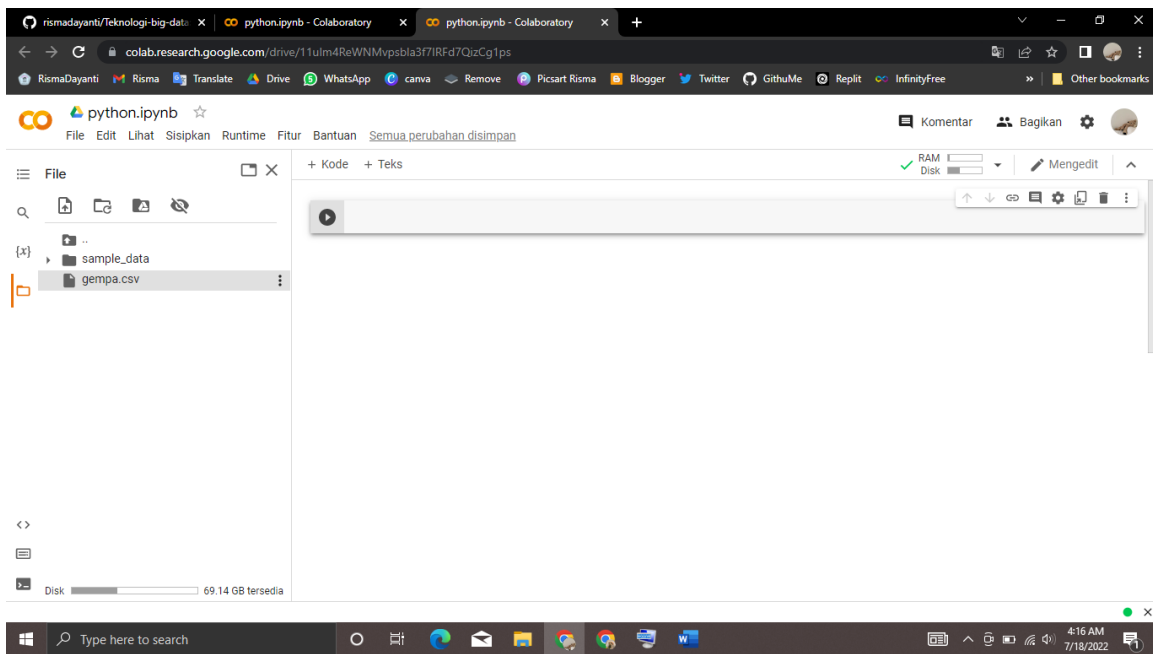
## 2. Mengupload file.csv.



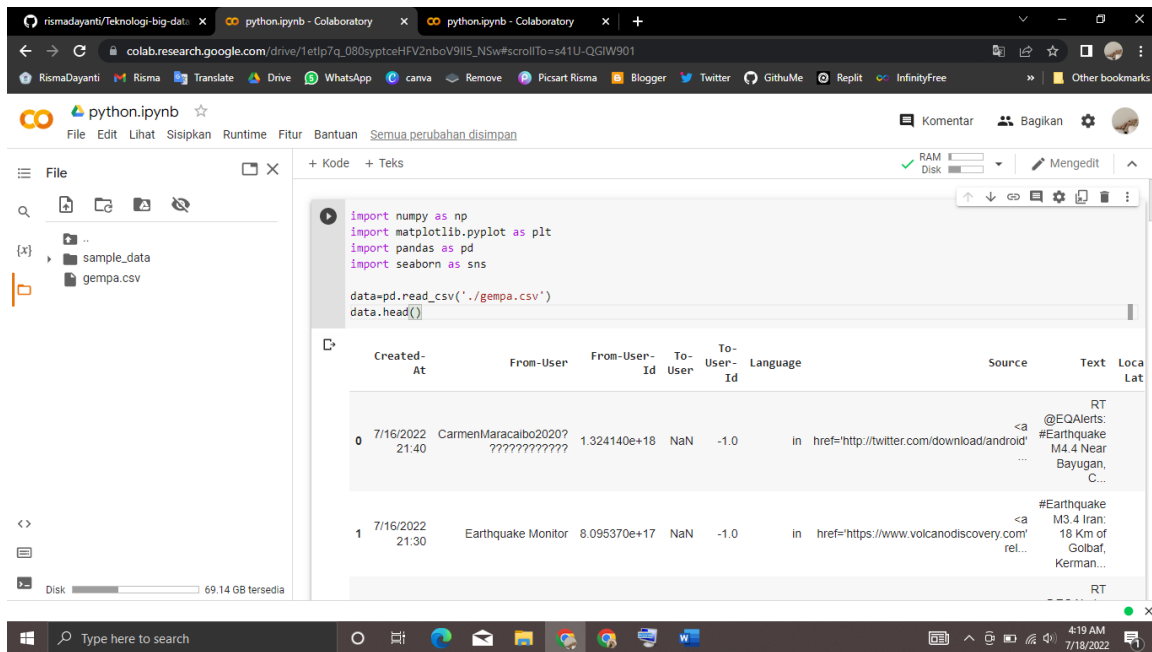
3. Pilih file.csv yang sudah diambil sebelumnya. Kemudian pilih open.



4. Tunggu hingga proses penguploadan selesai.



## 5. Selanjutnya import library dan file.csv nya.



The screenshot shows a Google Colab notebook with the following code in the first cell:

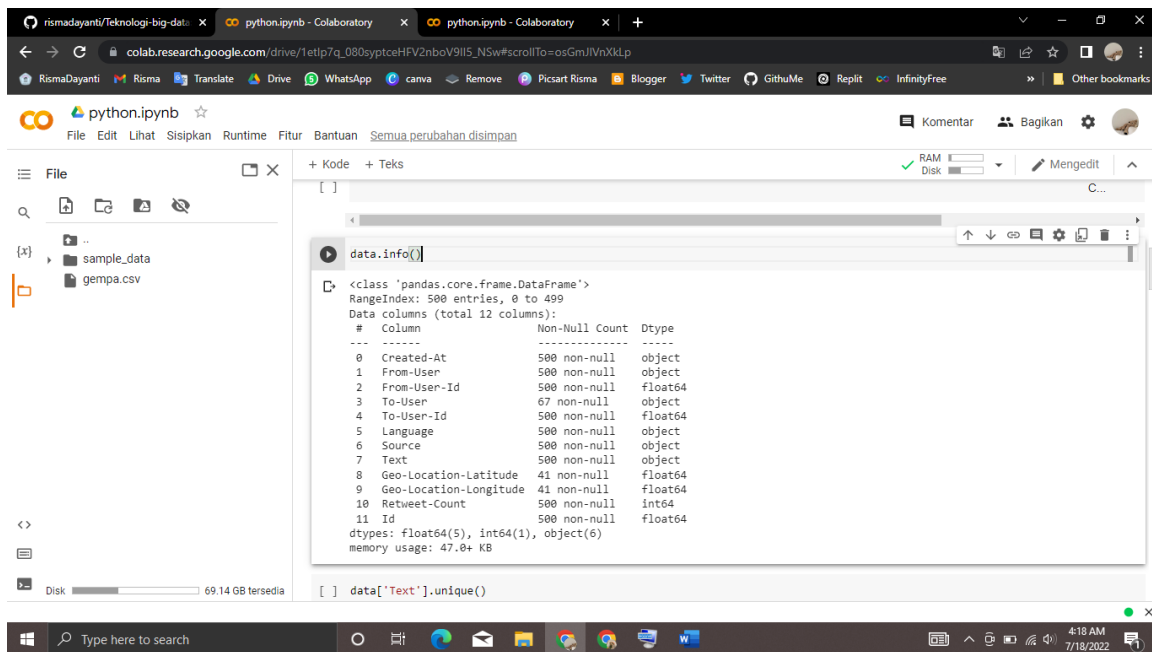
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

data=pd.read_csv('./gempa.csv')
data.head()
```

The output of the code is a preview of the first two rows of the 'gempa.csv' file:

	Created-At	From-User	From-User-Id	To-User-Id	Language	Source	Text	Loca	Lat
0	7/16/2022 21:40	CarmenMaracalbo2020?	1.324140e+18	NaN	-1.0	in href='http://twitter.com/download/android'	<a @EQAlerts: #Earthquake M4.4 Near Bayugan, C...		
1	7/16/2022 21:30	Earthquake Monitor	8.095370e+17	NaN	-1.0	in href='https://www.volcanodiscovery.com/rel...	<a #Earthquake M3.4 Iran: 18 Km of Golbaf, Kerman...		

## 6. Kemudian untuk melihat isi file gunakan script “data.info()”.



The screenshot shows the same Google Colab notebook with the following code in the second cell:

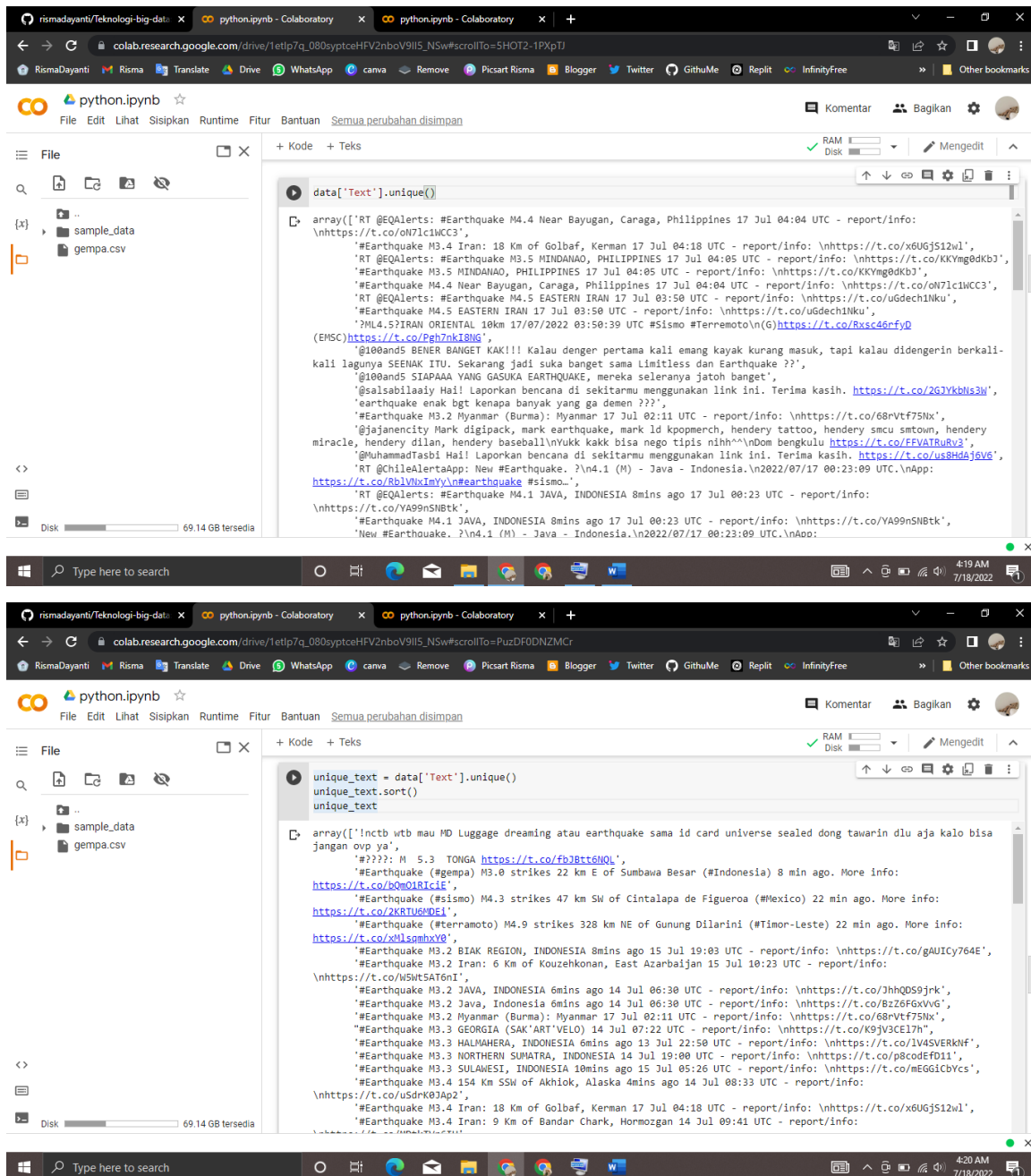
```
data.info()
```

The output of the code is the following information about the DataFrame:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Created-At            500 non-null   object
1   From-User             500 non-null   object
2   From-User-Id          500 non-null   float64
3   To-User               67 non-null    object
4   To-User-Id            500 non-null   float64
5   Language              500 non-null   object
6   Source                500 non-null   object
7   Text                  500 non-null   object
8   Geo-Location-Latitude 41 non-null    float64
9   Geo-Location-Longitude 41 non-null    float64
10  Retweet-Count          500 non-null   int64
11  Id                    500 non-null   float64
dtypes: float64(5), int64(1), object(6)
memory usage: 47.0+ KB
```

Below the output, the code `data['Text'].unique()` is visible in the editor.

## 7. Lalu cari teks unik dengan script “data[‘Text’].unique()”.



The image displays two screenshots of a Jupyter Notebook interface, likely Google Colab, showing the process of finding unique text in a dataset.

**Top Screenshot:** The notebook shows a file named `gempa.csv` loaded into the `sample_data` directory. The code cell contains the following Python code:

```
data['Text'].unique()
```

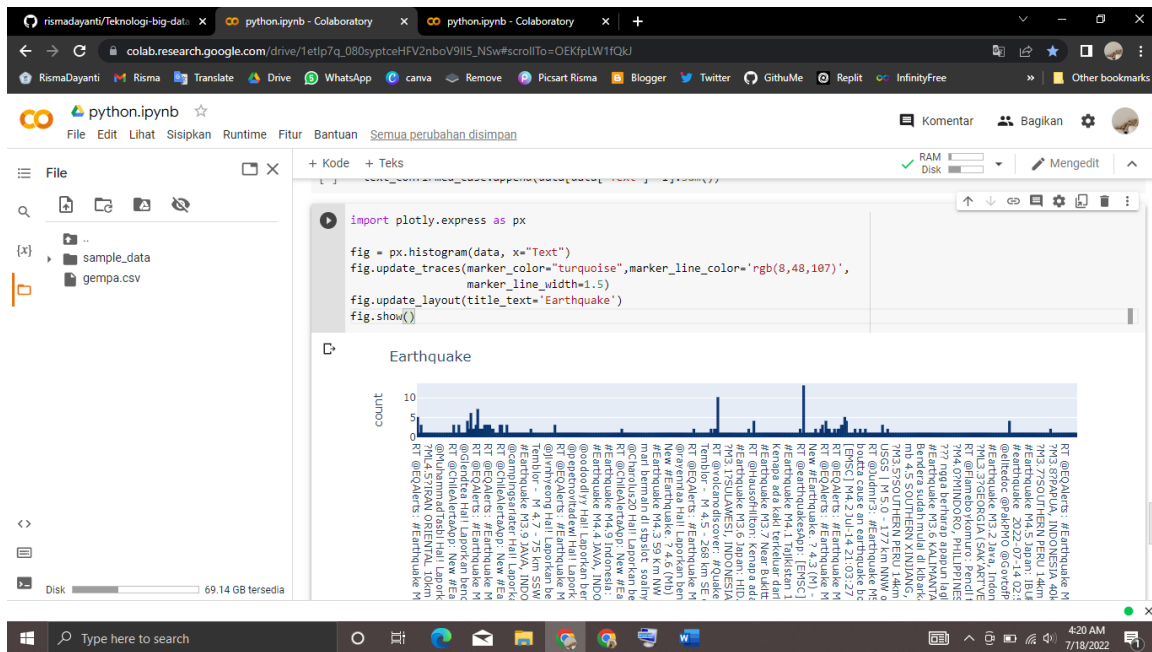
The output shows a large array of unique text entries, including earthquake reports and social media posts. The text is truncated in the image.

**Bottom Screenshot:** The notebook shows the same code cell with the following Python code:

```
unique_text = data['Text'].unique()
unique_text.sort()
unique_text
```

The output shows a sorted array of unique text entries, including earthquake reports and social media posts. The text is truncated in the image.

8. Langkah terakhir untuk menampilkan data analisis menggunakan script seperti di gambar.



Terimakasih, mohon maaf jika ada kekurangan 😊