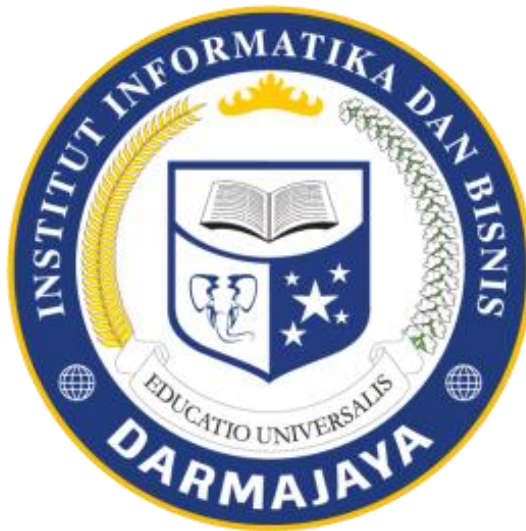


**ANALISIS BIG DATA
KONEKSI GOOGEL DRIVE, GITHUB DAN
CRAWLING**



**DISUSUN OLEH:
FIQQI AHLUDZIKRI (2111010034)**

**DOSEN:
Dr. M. SAID HASIBUAN**

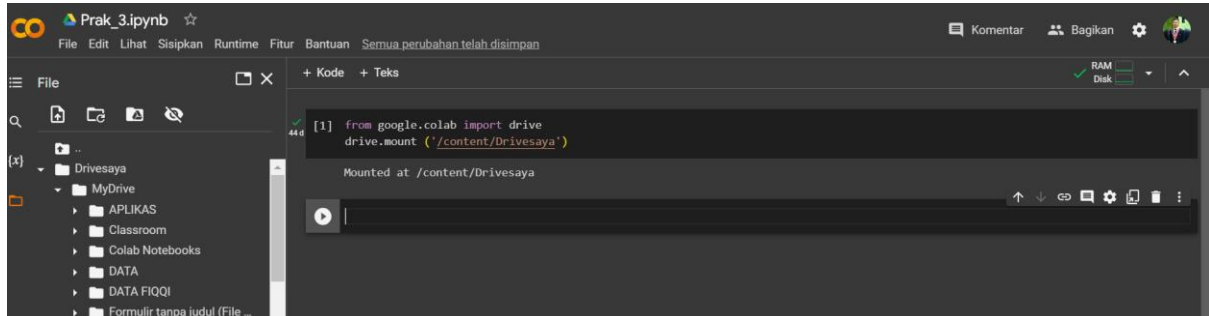
**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA
BANDAR LAMPUNG
TAHUN AJARAN 2023/2024**

Menghubungkan Google Colab dengan Drive

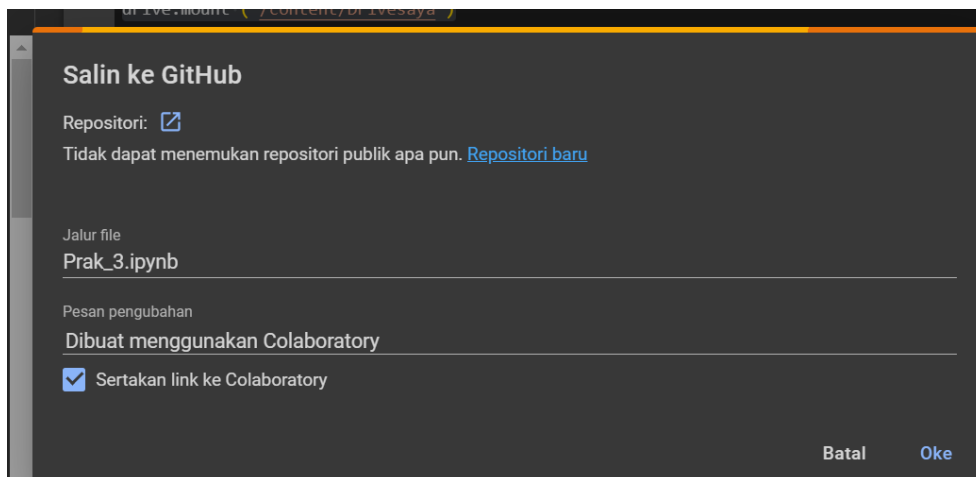
Code:

```
from google.colab import drive
drive.mount ('/content/Drivesaya')
```

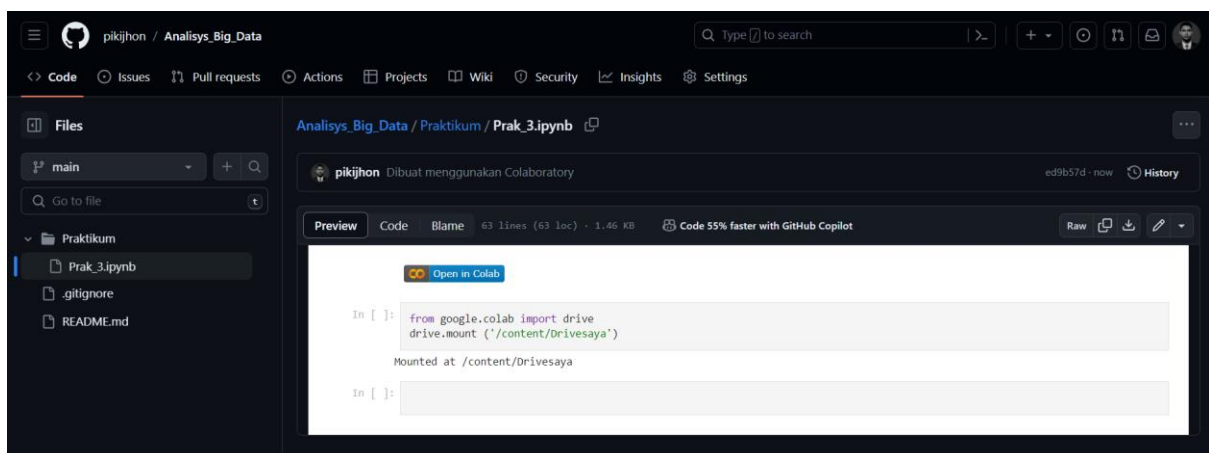
Hasil:



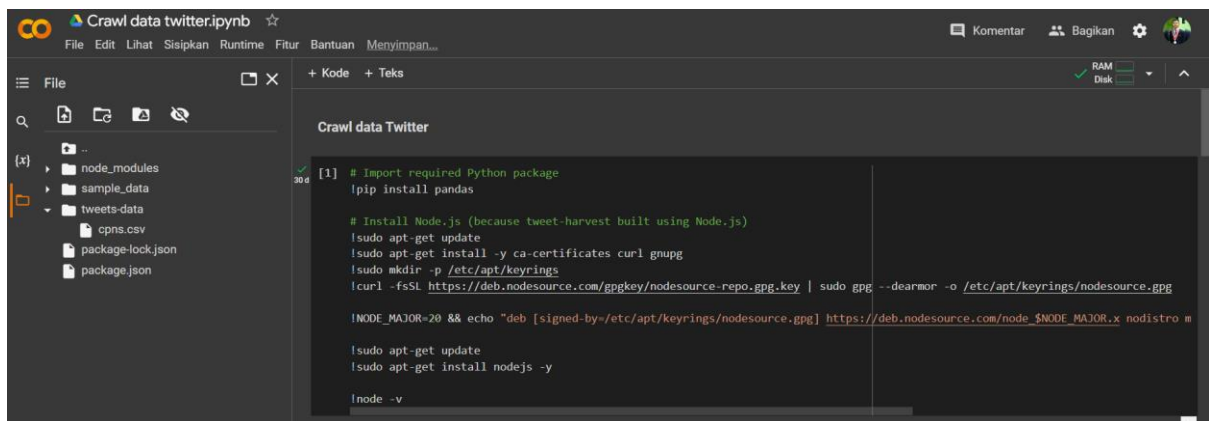
Menghubungkan Google Colab dengan Github



Simpan Salinan ke Github di repository



CRAWLING DATA TWITTER



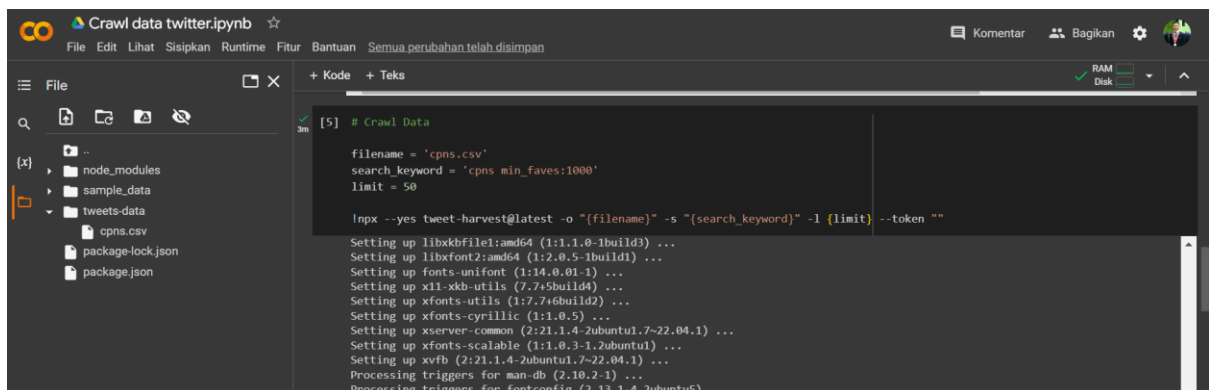
```
[1] # Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg

INODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro m

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
```



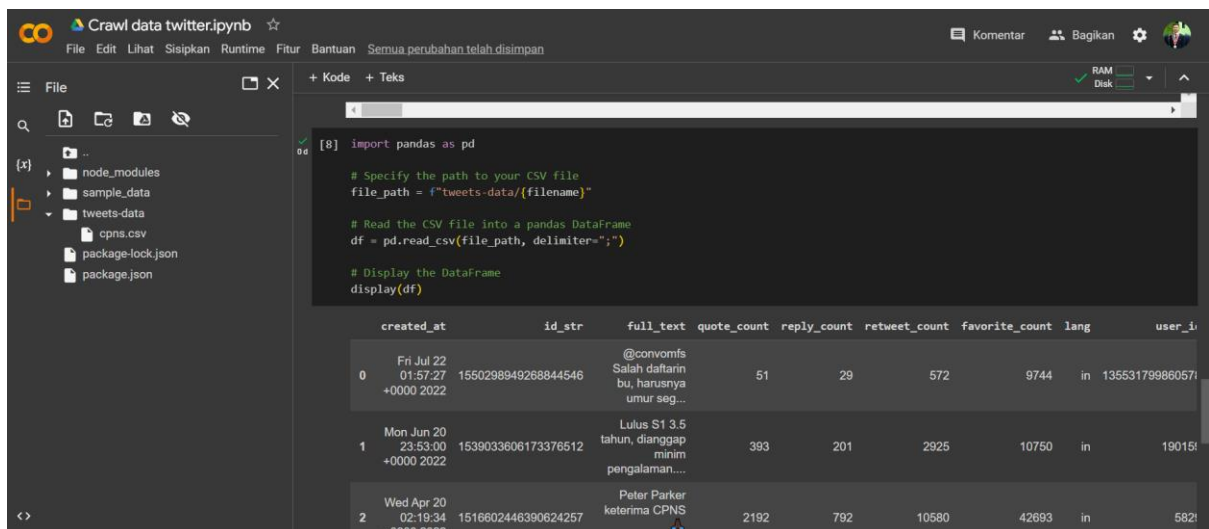
```
[5] # Crawl Data

filename = 'cpns.csv'
search_keyword = 'cpns min_faves:1000'
limit = 50

!lnpx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token ""

Setting up libxkbfile1:amd64 (1:1.1.0-1build3) ...
Setting up libxfont2:amd64 (1:2.0.5-1build1) ...
Setting up fonts-unifont (1:14.0.01-1) ...
Setting up x11-xkb-utils (7:7.4+5build4) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up xfonts-cyrillic (1:1.0.5) ...
Setting up xserver-common (2:21.1.4-2ubuntu1.7~22.04.1) ...
Setting up xfonts-scalable (1:1.0.3-1.2ubuntu1) ...
Setting up xvfb (2:21.1.4-2ubuntu1.7~22.04.1) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
```

Crawling data cpns dengan minimal like 1000.



```
[8] import pandas as pd

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=";")

# Display the DataFrame
display(df)
```

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_i
0	Fri Jul 22 01:57:27 +0000 2022	1550298949268844546	@convomfs Salah daftarin bu, harusnya umur seg...	51	29	572	9744	in	13553179886057
1	Mon Jun 20 23:53:00 +0000 2022	1539033606173376512	Lulus S1 3.5 tahun, dianggap minim pengalaman....	393	201	2925	10750	in	19015
2	Wed Apr 20 02:19:34 +0000 2022	1516602446390624257	Peter Parker katerima CPNS	2192	792	10580	42693	in	582

Menampilkan file cpns..csv



```
[8] 64 rows x 12 columns

# Cek jumlah data yang didapatkan
num_tweets = len(df)
print(f"Jumlah tweet dalam dataframe adalah {num_tweets}.")

Jumlah tweet dalam dataframe adalah 64.
```

