

Praktikum Pemodelan Twitter Crawling



Wahyu Ikbal Maulana

3323600056

D4 SDT B

Politeknik Elektronika Negeri Surabaya

```
In [ ]: #@title Twitter Auth Token
```

```
twitter_auth_token = '8684a9a776eccf3ae4422b7cf3bfb5e0d3e9b42a'
```

```
In [ ]: # 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
```

```
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://
```

```
!sudo apt-get update
```

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

```
!node -v
```

0% [Waiting for headers]

Hit:8 <https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu> jammy InRelease
Hit:9 <https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu> jammy InRelease
Hit:10 <https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu> jammy InRelease
Hit:11 <https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu> jammy InRelease
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
ca-certificates is already the newest version (20230311ubuntu0.22.04.1).
curl is already the newest version (7.81.0-1ubuntu1.16).
gnupg is already the newest version (2.2.27-3ubuntu2.1).
0 upgraded, 0 newly installed, 0 to remove and 45 not upgraded.
gpg: cannot open '/dev/tty': No such device or address
curl: (23) Failed writing body
deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_20.x nodistro main
Hit:1 <https://cloud.r-project.org/bin/linux/ubuntu> jammy-cran40/ InRelease
Hit:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64 InRelease
Hit:3 https://deb.nodesource.com/node_20.x nodistro InRelease
Hit:4 <http://archive.ubuntu.com/ubuntu> jammy InRelease
Hit:5 <http://archive.ubuntu.com/ubuntu> jammy-updates InRelease
Hit:6 <http://archive.ubuntu.com/ubuntu> jammy-backports InRelease
Hit:7 <http://security.ubuntu.com/ubuntu> jammy-security InRelease
Hit:8 <https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu> jammy InRelease
Hit:9 <https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu> jammy InRelease
Hit:10 <https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu> jammy InRelease
Hit:11 <https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu> jammy InRelease
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
nodejs is already the newest version (20.12.2-1nodesource1).
0 upgraded, 0 newly installed, 0 to remove and 45 not upgraded.
v20.12.2
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (2.0.3)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Hit:1 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64 InRelease
Hit:2 <https://cloud.r-project.org/bin/linux/ubuntu> jammy-cran40/ InRelease
Hit:3 https://deb.nodesource.com/node_20.x nodistro InRelease
Hit:4 <http://archive.ubuntu.com/ubuntu> jammy InRelease
Hit:5 <http://security.ubuntu.com/ubuntu> jammy-security InRelease
Hit:6 <http://archive.ubuntu.com/ubuntu> jammy-updates InRelease

```

Hit:7 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:8 https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:10 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy InRelease
Hit:11 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
ca-certificates is already the newest version (20230311ubuntu0.22.04.1).
curl is already the newest version (7.81.0-1ubuntu1.16).
gnupg is already the newest version (2.2.27-3ubuntu2.1).
0 upgraded, 0 newly installed, 0 to remove and 45 not upgraded.
gpg: cannot open '/dev/tty': No such device or address
curl: (23) Failed writing body
deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_
20.x nodistro main
Hit:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease
Hit:3 https://deb.nodesource.com/node_20.x nodistro InRelease
Hit:4 http://archive.ubuntu.com/ubuntu jammy InRelease
Hit:5 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:6 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:7 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:8 https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:10 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy InRelease
Hit:11 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
nodejs is already the newest version (20.12.2-1nodesource1).
0 upgraded, 0 newly installed, 0 to remove and 45 not upgraded.
v20.12.2

```

```

In [ ]: # Crawl Data

filename = 'prabowo.csv'
search_keyword = 'prabowo lang:id'
limit = 100

!npx -y tweet-harvest@2.6.0 -o "{filename}" -s "{search_keyword}" --tab "LATEST"

```

Tweet Harvest [v2.6.0]

Research by **Helmi Satria**

Use it for Educational Purposes only!

This script uses Chromium Browser to crawl data from Twitter with **your Twitter auth token**.

Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.

Note: This script only runs on your local device.

Opening twitter search page...

Found existing file ./tweets-data/prabowo.csv, renaming to ./tweets-data/prabowo.old.csv

Filling in keywords: prabowo lang:id

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/prabowo.csv

Total tweets saved: 19

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/prabowo.csv

Total tweets saved: 38

-- Scrolling... (1) (2)

Your tweets saved to: /content/tweets-data/prabowo.csv

Total tweets saved: 56

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/prabowo.csv

Total tweets saved: 73

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/prabowo.csv

Total tweets saved: 93

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/prabowo.csv

Total tweets saved: 113

--Taking a break, waiting for 10 seconds...

Got 113 tweets, done scrolling...

```
In [ ]: 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)
```

| | conversation_id_str | created_at | favorite_count | full_text | |
|-----|---------------------|--------------------------------|----------------|---|-------------|
| 0 | 1785554466877653096 | Wed May 01 15:40:00 +0000 2024 | 0 | @kompascom Partai yg membesarkan n kasi fasili... | 17856956759 |
| 1 | 1785695644633715089 | Wed May 01 15:39:52 +0000 2024 | 0 | Kemarin mimpi prabowo sama jokowi datang kerum... | 17856956446 |
| 2 | 1785651560690884956 | Wed May 01 15:39:42 +0000 2024 | 0 | @yaniarsim @prabowo @gibran_tweet Sy benci ke ... | 17856955997 |
| 3 | 1785695566019854572 | Wed May 01 15:39:34 +0000 2024 | 0 | Jerry Massie: Jokowi dan SBY Jadi Mentor Andal... | 17856955660 |
| 4 | 1785695383198298381 | Wed May 01 15:38:50 +0000 2024 | 0 | @PDI_Perjuangan https://t.co/lalSy804wX | 17856953831 |
| ... | ... | ... | ... | ... | ... |
| 108 | 1785585127881654283 | Wed May 01 15:06:22 +0000 2024 | 0 | @Araghutama Kalo kamu mah pasti dibales terus ... | 17856872137 |
| 109 | 1784744191127367754 | Wed May 01 15:06:16 +0000 2024 | 0 | @H1m4_W4r1 @BuayaUbanan @snjayajonggrang @_sek... | 17856871889 |
| 110 | 1784744191127367754 | Wed May 01 15:05:28 +0000 2024 | 0 | @H1m4_W4r1 @BuayaUbanan @snjayajonggrang @_sek... | 17856869869 |
| 111 | 1785678044038049809 | Wed May 01 15:04:36 +0000 2024 | 0 | @BadmintonTalk @BonCabe Gregoria Mariska Tunju... | 17856867672 |
| 112 | 1784525877789028424 | Wed May 01 15:04:34 | 0 | @sholxs @CAMERALIVE6849 Kenapa Allah pilih Jok... | 17856867602 |

| conversation_id_str | created_at | favorite_count | full_text |
|---------------------|------------|----------------|-----------|
| | +0000 | | |
| | 2024 | | |

113 rows × 15 columns

```
In [ ]: # Cek jumlah data yang didapatkan

num_tweets = len(df)
print(f"Jumlah tweet dalam dataframe adalah {num_tweets}.")
```

Jumlah tweet dalam dataframe adalah 113.