LAPORAN PRAKTIKUM

BIG DATA ANALYTIC

PERTEMUAN KE-4

****

**Disusun oleh :**

**NAMA : FERDI DIRGANTARA**

**NIM : 175410039**

**JURUSAN : TEKNIK INFORMATIKA**

**JENJANG : S1**

LABORATORIUM TERPADU

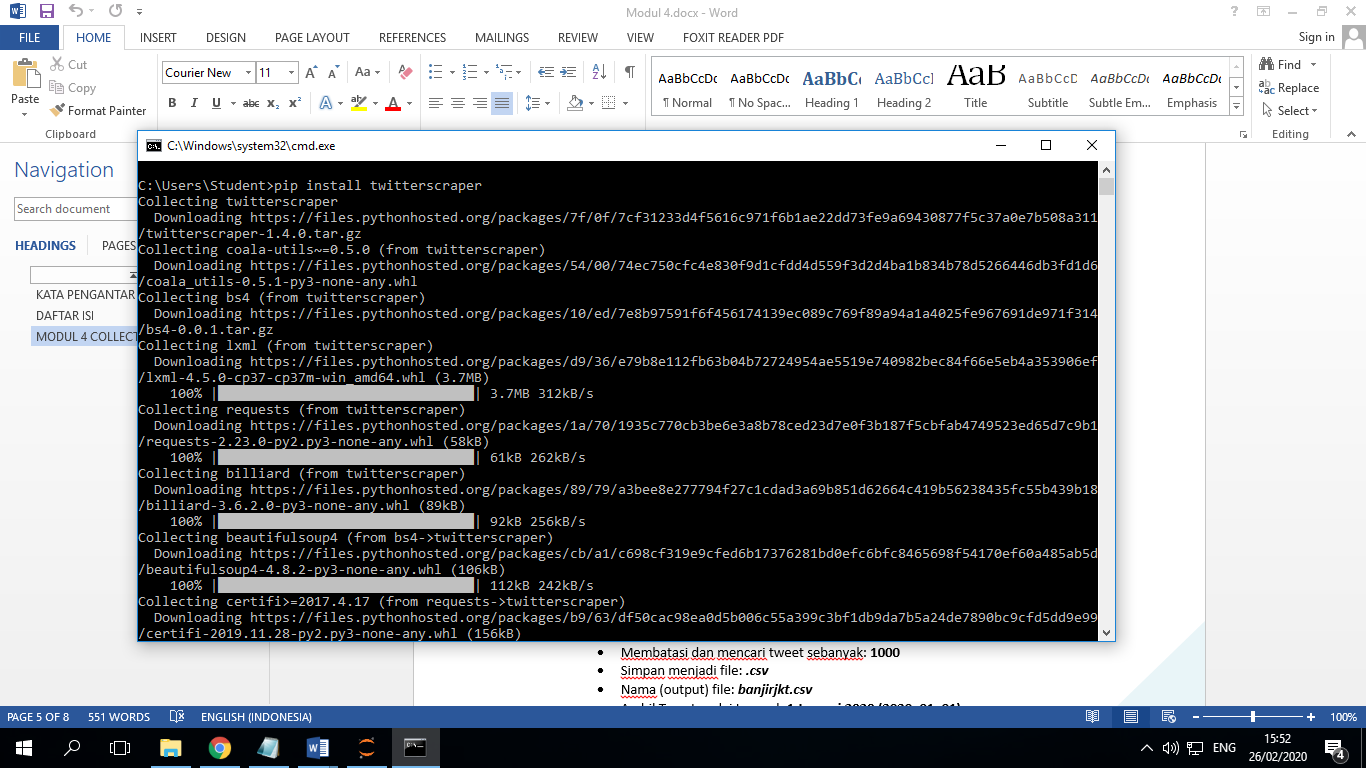
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTERAKAKOMYOGYAKARTA

2020

**LISTING**

**COLLECTING DATA**

1. Instal Twitterscraper



Proses diatas menjalankan perintah **pip install twitterscraper** yang digunakan untuk melakukan scraping dari twitter

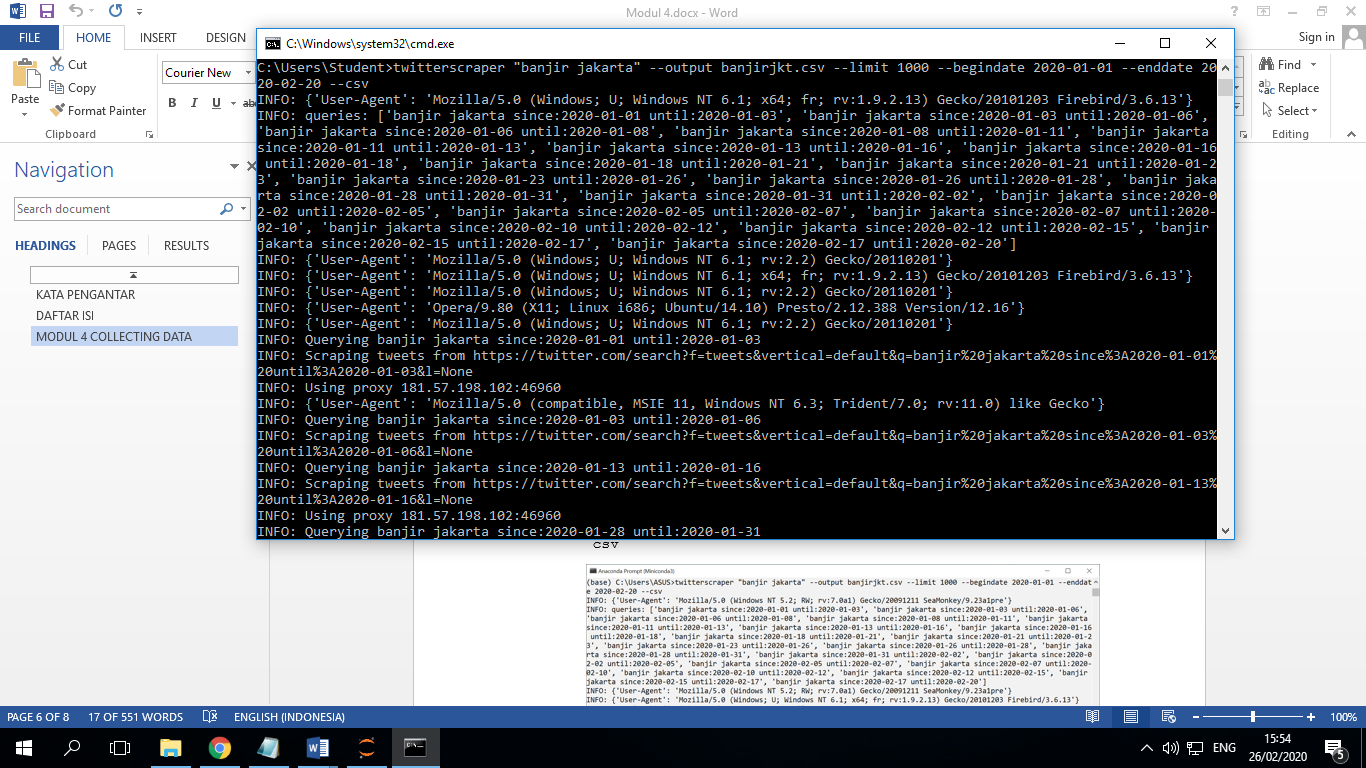
1. Misal ingin mengambil data tweet tentang banjir Jakarta

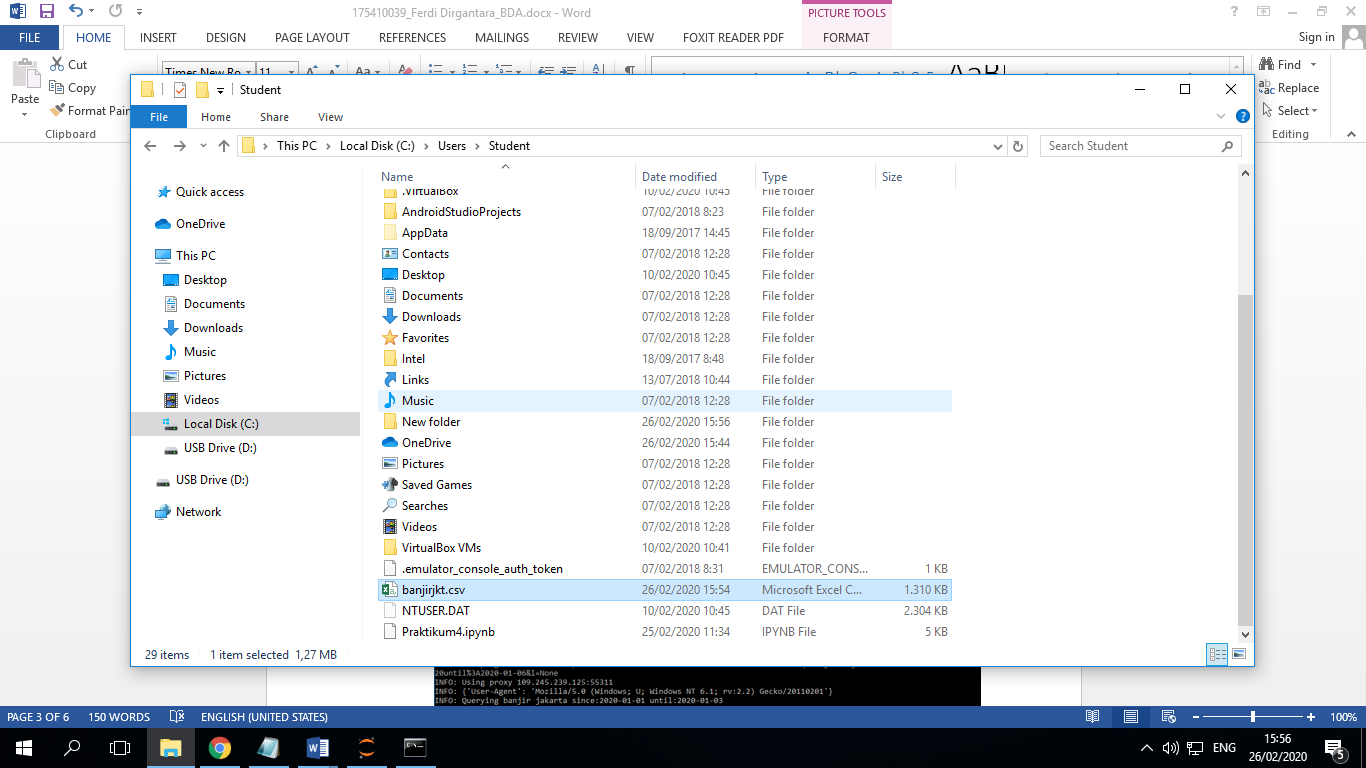
Dengan ketentuan, range data 1000 dengan formati file .csv dengan range waktu dari 1 Jaurari 2020 – 20 Februari 2020.

1. Mengakses dan Menyimpan data.
2. Penyimpanan format .csv

Perintah yang digunakan adalah

**twitterscraper "banjir jakarta" --output banjirjkt.csv --limit 1000 --begindate 2020-01-01 --enddate 2020-02-20 –csv**



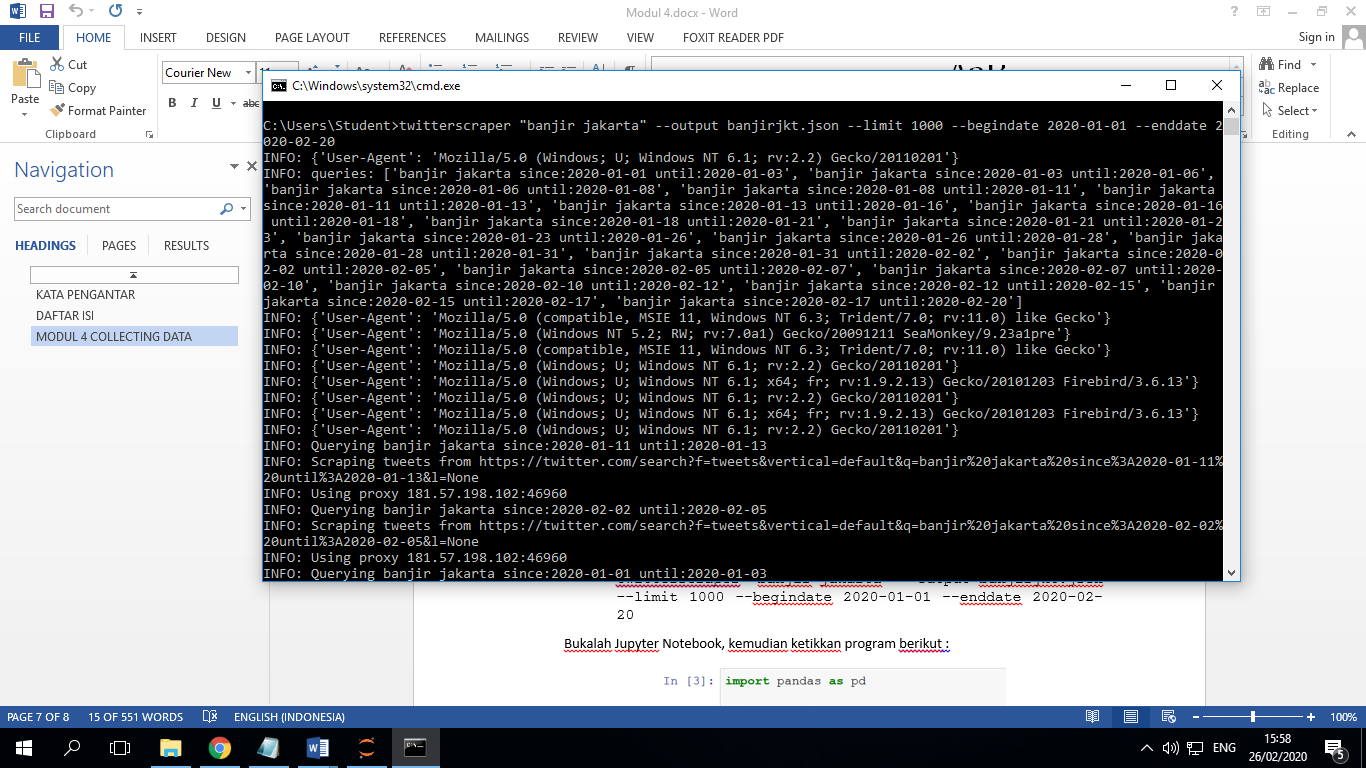


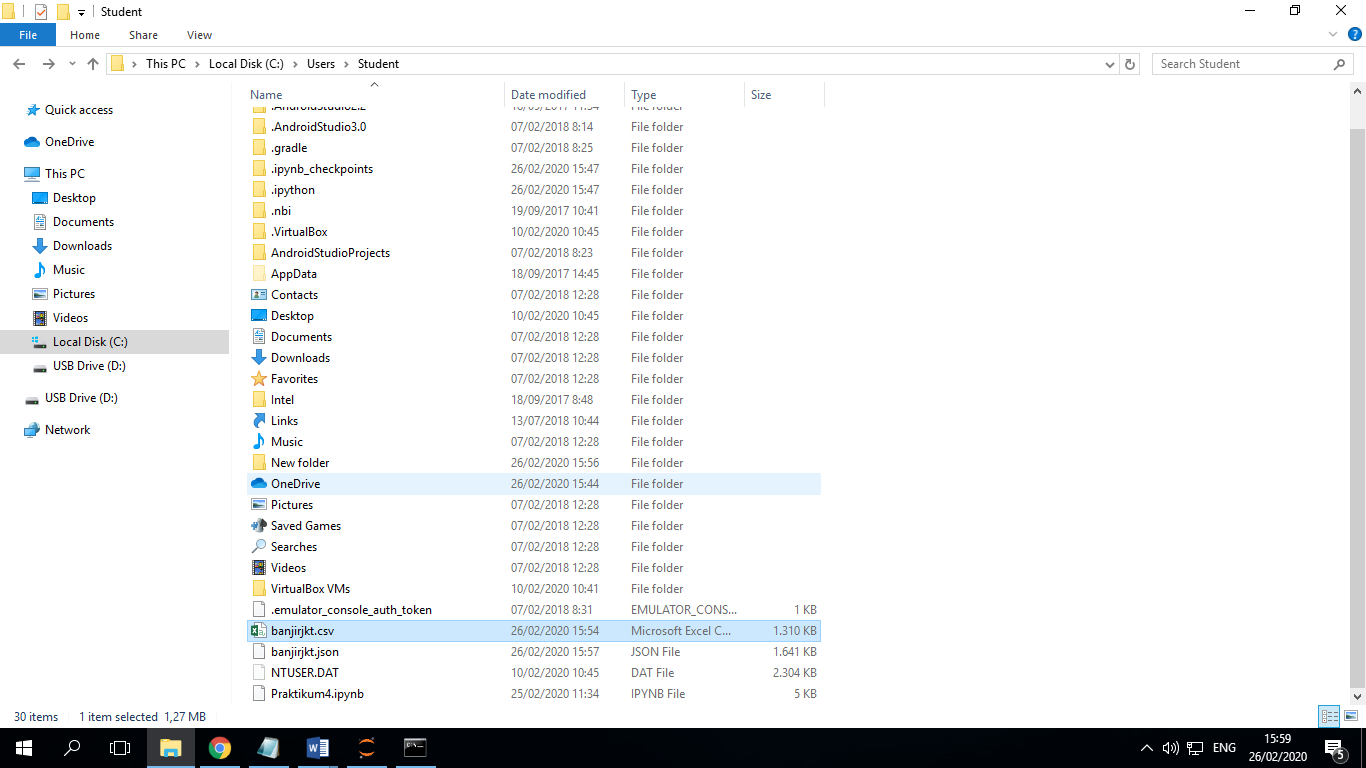
Melakukan scraping dari twitter untuk mengambil data atau posting yang terkait dengan banjir yang terjadi di Jakarta. Data yang di scrap kemudian disimpan kedalam bentuk file csv dengan nama banjirjkt.

1. Penyimpanan format .json

Perintah yang digunakan adalah

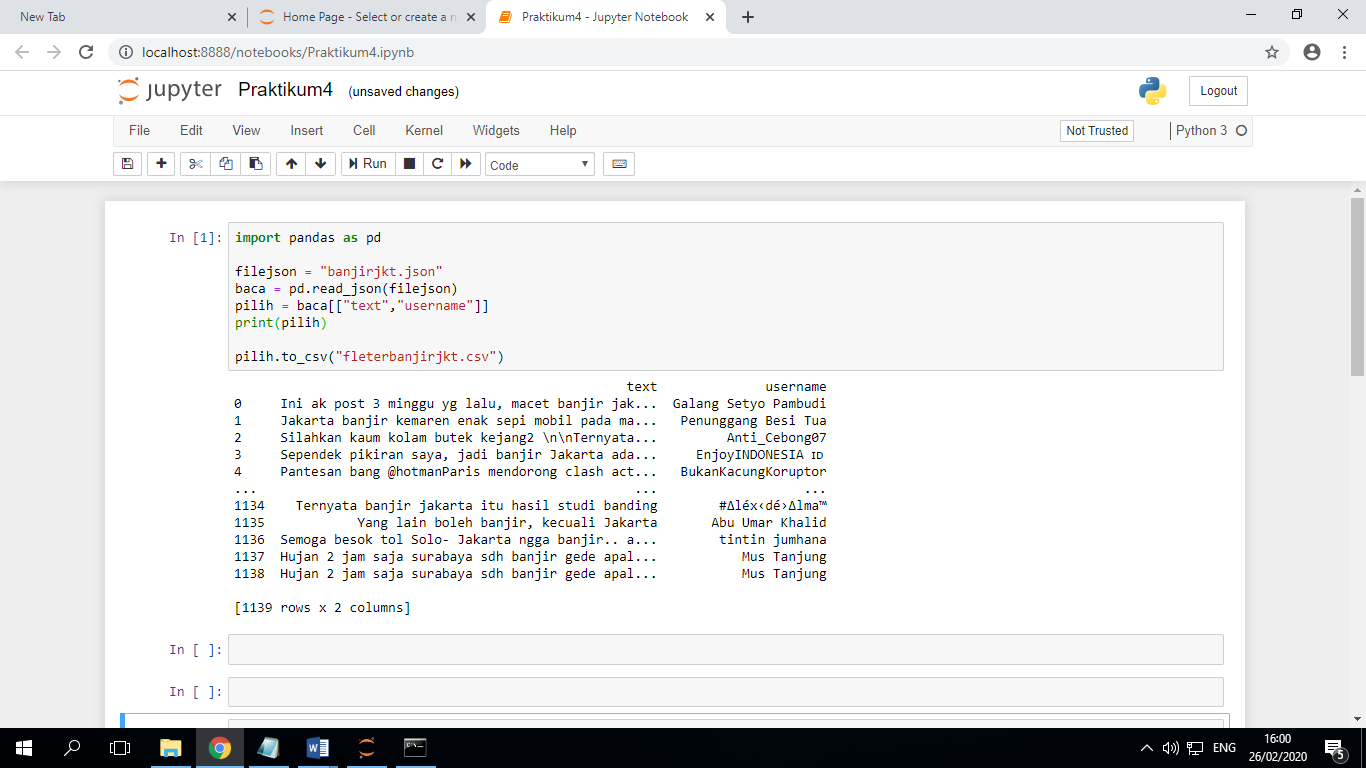
**twitterscraper "banjir jakarta" --output banjirjkt.json --limit 1000 --begindate 2020-01-01 --enddate 2020-02-20 –csv**





Sama seperti sebelumnya yaitu untuk mengambil data terkait banjir jakarata namun dengan output file json.

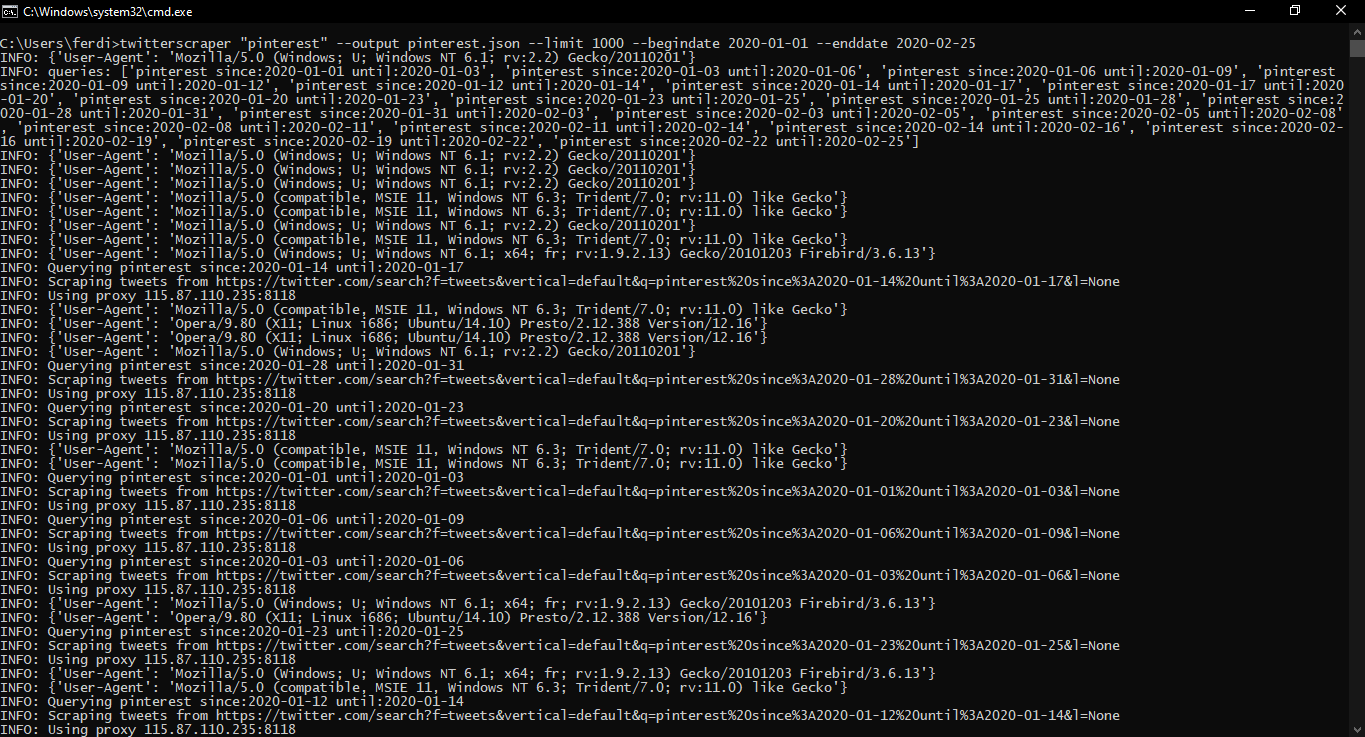
**Memfilter Data**

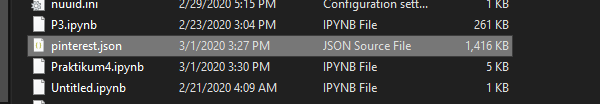


Data yang di scrap tadi kemudian di filter dengan menggunakan package pandas berupa text dan username dari pemilik postingan. Outputnya kemudian disimpan kedalam file fleterbanjirjkt.csv.

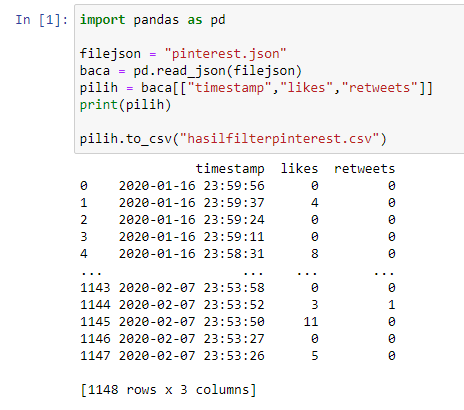
**LATIHAN**

Lakukanlah scraping dengan menggunakan kata kunci yang berbeda. Kemudian lakukan filter untuk menyimpan hanya data “timestamp”, “likes”, dan “retweets”.

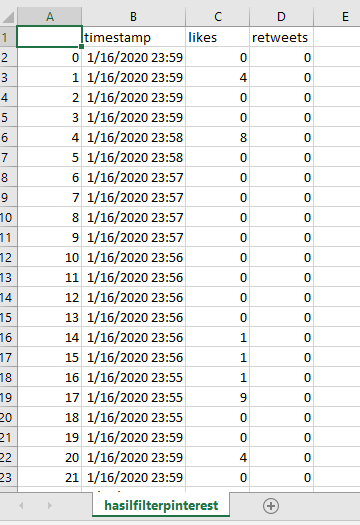




Memfilter berdasarkan Timestamp, Likes dan Retweets







Pada latihan diatas dilakukan scraping posting pada twitter yang terkait dengan pinterest pada range tanggal yang telah ditentukan. Data kemudian disimpan dalam bentuk file json dan dilakukan filtering berdasarkan timestamp, likes, dan retweets ke dalam file hasilfilterpinterest.csv.