



PREPROCESSING DATA DARI TWITTER

Pemrosesan Teks



Nim : 5200411488

Nama : Arieska Restu Harpian Dwika

Mengambil Data dari Twitter

```
9
10 ###input key dan token
11 consumer_key='ly0MjTrxKVG3CIbYwxSfGnYZq'
12 consumer_secret='PRTGKnmI1kKFFgWUS7jYv5p3UpgAh7afHEU7hf6hgwbTuFB57'
13
14 access_token='1493952640577044486-prDDvGVRingcTy3jT6KDMtFUBYOdBvg'
15 access_token_secret='070Hm6qiWltP9ZUOpB2Q8xcyRdZrFR8UQKb4FtPvz3'
16
17 auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
18 auth.set_access_token(access_token, access_token_secret)
19 api = tweepy.API(auth,wait_on_rate_limit=True)
20
21 #####MotoGP Mandalika
22 # Open/Create a file to append data
23 csvFile = open('3rd Meet (Thursday, February, 24th 2022)\Practice\Data\Rusia.csv', 'a')
24 #Use csv Writer
25 csvWriter = csv.writer(csvFile)
26
27 for tweet in tweepy.Cursor(api.search_tweets,q="#rusia",count=2,lang="id").items():
28     csvWriter.writerow([tweet.text.encode('utf8')])
29
30 # =====
31
32 #read data
33 def load_data():
34     data = pd.read_csv('3rd Meet (Thursday, February, 24th 2022)\Practice\Data\Rusia.csv')
35     return data
36
37 #Pembuatan dataframe
38 teks = load_data()
39 teks = '''b'Harga minyak dunia melonjak pada Selasa (1/3), dengan minyak mentah AS mencapai level tertinggi sejak Juni 2014 kar\xe2\x80\xa6 https://t.co/e3nHzH8Rto'''''
40
```

Melakukan Uppercase

```
42
43 # Uppercase
44 print(teks.upper() + '\n')
45
```

Melakukan Lowercase

```
46 # Lowercase
47 print(teks.lower() + '\n')
48
```

Melakukan Split Text

```
48  
49 # Split teks  
50 for i in re.split(' ', teks):  
51     print(i)  
52 print()
```

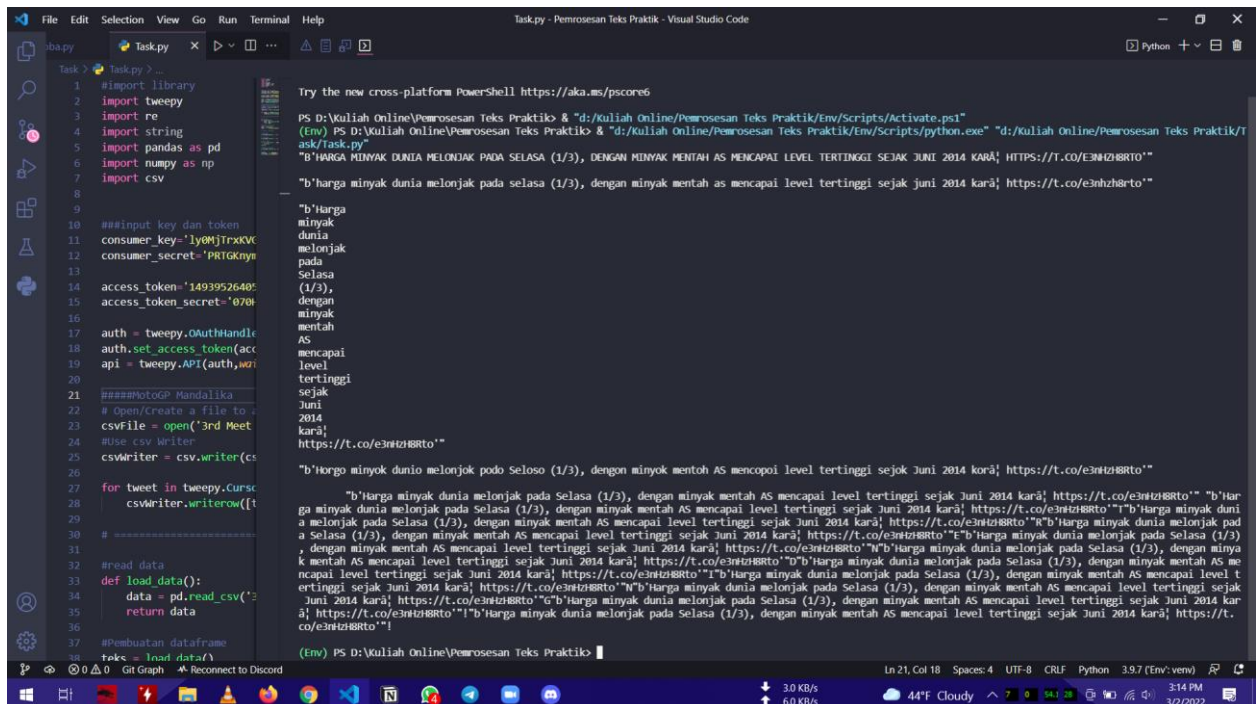
Melakukan Replace Karakter

```
54 # Replace karakter  
55 print(teks.replace('a', 'o') + '\n')  
56
```

Melakukan Join

```
57 # Join  
58 kataJoin = "\t TRENDING!!"  
59 print(teks.join(kataJoin) + '\n')
```

Hasil



```
Task.py
1 #import library
2 import tweepy
3 import re
4 import string
5 import pandas as pd
6 import numpy as np
7 import csv
8
9
10 ###input key dan token
11 consumer_key='ly0MjTrxKVG3CIbYwxSfGnYZq'
12 consumer_secret='PRTGKnymI1kKFFgNUS7jYv5p3UpgAh7afHEU7hF6hgwbTuFB57'
13
14 access_token='1493952640577044486-prDDvGVRngcTyJjT6KDMtFUBYODBvg'
15 access_token_secret='070Hm6qiWltP9ZU0pB2QBxcyRdZrFR8UQKb4FtPvzJ'
16
17 auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
18 auth.set_access_token(access_token, access_token_secret)
19 api = tweepy.API(auth,wait_on_rate_limit=True)
20
21 #####MotoGP Mandalika
22 # Open/Create a file to append data
23 csvFile = open("3rd Meet")
24 #use csv writer
25 csvwriter = csv.writer(csvFile)
26
27 for tweet in tweepy.Cursor(api.search_tweets, "b'harga minyak dunia melonjak pada Selasa (1/3), dengan minyak mentah AS mencapai level tertinggi sejak Juni 2014 karã', https://t.co/e3nhtz8Rto",
28                             tweet_mode='extended').items():
29     # =====
30     #read data
31     def load_data():
32         data = pd.read_csv('data.csv')
33         return data
34
35     #membuat dataframe
36     data = load_data()
37
38     #menambahkan data ke dataframe
39     data = data.append(tweet._json, ignore_index=True)
```

Source Code

```
#import library
import tweepy
import re
import string
import pandas as pd
import numpy as np
import csv

###input key dan token
consumer_key='ly0MjTrxKVG3CIbYwxSfGnYZq'
consumer_secret='PRTGKnymI1kKFFgNUS7jYv5p3UpgAh7afHEU7hF6hgwbTuFB57'

access_token='1493952640577044486-prDDvGVRngcTyJjT6KDMtFUBYODBvg'
access_token_secret='070Hm6qiWltP9ZU0pB2QBxcyRdZrFR8UQKb4FtPvzJ'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth,wait_on_rate_limit=True)

#####MotoGP Mandalika
# Open/Create a file to append data
```

```

csvFile = open('3rd Meet (Thursday, February, 24th
2022)\Practice\Data\Rusia.csv', 'a')
#Use csv Writer
csvWriter = csv.writer(csvFile)

for tweet in
tweepy.Cursor(api.search_tweets,q="#rusia",count=2,Lang="id").items():
    csvWriter.writerow([tweet.text.encode('utf8')])

# =====

#read data
def load_data():
    data = pd.read_csv('3rd Meet (Thursday, February, 24th
2022)\Practice\Data\Rusia.csv')
    return data

#Pembuatan dataframe
teks = load_data()
teks = '''b'Harga minyak dunia melonjak pada Selasa (1/3), dengan minyak
mentah AS mencapai level tertinggi sejak Juni 2014 kar\xe2\x80\xa6
https://t.co/e3nHzH8Rto'''

# # teks = texts[0]

# Uppercase
print(teks.upper() + '\n')

# Lowercase
print(teks.lower() + '\n')

# Split teks
for i in re.split(' ', teks):
    print(i)
print()

# Replace karakter
print(teks.replace('a', 'o') + '\n')

# Join
kataJoin = "\t TRENDING!!"
print(teks.join(kataJoin) + '\n')

```