

Crawling

Pada tugas ini data yang digunakan untuk dataset adalah tweet, tweet diambil dengan cara crawling menggunakan library python tweepy

terdapat dua library yang digunakan pada code ini yaitu tweepy untuk crawling data dan csv untuk eksport data menjadi data CSV.

untuk menggunakan tweepy kita membutuhkan beberapa key sebagai autentifikasi, key tersebut bisa diperoleh di <https://developer.twitter.com/en.html> (<https://developer.twitter.com/en.html>) berikut key yang dibutuhkan

- consumer_key
- consumer_secret
- access_token
- access_token_secret

untuk mendapatkan key atau yang biasa disebut token diatas kita diharuskan mendaftar sebagai developer di tweeter dan mengisi beberapa form untuk verifikasi akun, setelah mendapatkan akun , token dapat digunakan menggunakan kode

- self.auth = tweepy.OAuthHandler(self.consumer_key, self.consumer_secret)
- self.auth.set_access_token(self.access_token, self.access_token_secret)

data yang diambil pada tweet ini adalah tweet yang mengandung "**@prabowo**", berikut detail datanya:

- id tweet
- id user
- Screen name (username)
- tweet
- created at

```

In [ ]: import tweepy
import csv

class Services:

    def __init__(self):
        self.consumer_key = 'aqh26LmWvnCwEqoWnBkNDgq3X'
        self.consumer_secret = 'WLTGI57tcvei5AwTZNxiyrgDmTqhmVEaeQlXhknzIFZBym58I'
        self.access_token = '232106297-SLMzoxQqdr0boZl0s8VLSHayem1FUr99pGE8Y8Wi'
        self.access_token_secret = '8qlseHuNKDsnIVnxuDZcXfjDepCrWDcp5U1JtZ4PfOU9'
        self.auth = tweepy.OAuthHandler(self.consumer_key, self.consumer_secret)
        self.auth.set_access_token(self.access_token, self.access_token_secret)
        self.api = tweepy.API(self.auth, wait_on_rate_limit=True, wait_on_rate_lim

    def getData(self, parameter):
        print("starting")
        i = 0
        with open('data_crawling.csv', 'w', newline='') as csvFile: # membuat fi
            writer = csv.writer(csvFile) #membuat csv writer
            row = ["ID Tweet", "ID User", "Screen Name", "Tweet", "Timestamp"] #
            writer.writerow(row) # menulis data header
            for post in tweepy.Cursor(self.api.search, q=parameter, tweet_mode='e
                # tweepy menggunakan cursor untuk mencari sesuai parameter yaitu
                # tweet_mode="extended" digunakan untuk mendapatkan full text dan
                try:
                    if (post.retweet_count == 0): #menghindari tweet hasil retwee
                        if i <= 99 : #membatasi hanya 100 data saja yang akan di
                            row = [post.id_str.encode('utf-8'), post.user.id_str
                                post.full_text.encode('utf-8'), post.created_at
                                # data seperti id_tweet, id_user di encode karena data
                                # oleh csv sehingga harus d encode agar mendapatkan j
                                # untuk full_text (tweet) di encode untuk menghindar
                                writer.writerow(row) #perintah untuk menulis ke csv
                            else:
                                break
                            i = i + 1 # iterasi
                            print(i, " tweet downloaded ", post.created_at)
                        except tweepy.RateLimitError: # untuk handle apabila token t
                            print("sleep") # crawling akan di jeda sampai token mendapat
                        except tweepy.TweepError as e: # untuk handle apabila terjadi
                            print('tweepy error')
                            print(e.reason)
                            continue
                        except StopIteration: # stop iteration ketika tidak ada data ya
                            print('habis')
                            break
                    csvFile.close() #close file csv
                return True

```

```

In [ ]: search = "@prabowo"
if (Services().getData(search)):
    print("successfully")
else:
    print("failed")

```

Data hasil crawling disimpan dalam bentuk CSV, berikut data hasil crawling yang di peroleh:

```
In [1]: import pandas as pd
```

```
In [2]: pd.read_csv('data_crawling.csv')
```

```
Out[2]:
```

	ID Tweet	ID User	Screen Name	Tweet	Tir
0	b'1127125080545189888'	b'1115202813968015360'	SGanjen	b'@katakitaatweet @Aryprasetyo85 @emhaainunnadj...	20
1	b'1127125040342786050'	b'973292600017895425'	brothe28	b'@RachlanNashidik @prabowo @jokowi Biasa demo...	20
2	b'1127125027827044352'	b'1058630527119323136'	Toyez1	b'@DiniKurnia21 @MangiranNing @MahesaTiwi @kan...	20
3	b'1127125020407320576'	b'1119035387387600896'	Faqih07380284	b'@permadiaktivis @bawaslu_RI @prabowo @DivHum...	20
4	b'1127125017613881344'	b'1018171197350035456'	Jusca07538974	b'Makin muak dgn tingkah2 para Pengumpat dari ...	20
5	b'1127125008566837248'	b'887520780354953216'	leenahanwoo	b'@FerdinandHae2 @jokowi @KPU_ID @prabowo In...	20
6	b'1127124908838871040'	b'1100944001115406336'	sasauw_nelson	b'@FerdinandHae2 @jokowi @KPU_ID @prabowo HA...	20
7	b'1127124904405377024'	b'437379680'	Rien_Harbani	b'@RachlanNashidik @prabowo @jokowi Klo drmkokr...	20
8	b'1127124885279350784'	b'941513645925658624'	beritaemak	b'@FerdinandHae2 Pa @prabowo gausah turun bi...	20
9	b'1127124880598585345'	b'2321810466'	NiswariSejuk	b'@Demokrat_TV @renandabachtar Apa 02 gk suka ...	20
10	b'1127124750986121217'	b'1119072386970578944'	IQLee1	b'@FerdinandHae2 @jokowi @KPU_ID @prabowo Je...	20
11	b'1127124733848219649'	b'1122933644568317952'	winningbrutally	b"@francmohede @prabowo @sandiuno Jangan artik...	20
12	b'1127124614121811968'	b'1542495308'	bagussegara1	b'@prabowo \nJangan offline side\n\nDarah mera...	20
13	b'1127124564629082112'	b'778827993091158016'	adilmakmur02	b'Senin lalu Bapak @prabowo dan Bang @sandiuno...	20
14	b'1127124480797466624'	b'1116198612914937856'	Odhe231	b'@RachlanNashidik @prabowo @jokowi Jd apa gun...	20
15	b'1127124480612978688'	b'1093059512687194112'	yenyeni11	b'@RachlanNashidik @prabowo @jokowi Bentuklah ...	20

	ID Tweet	ID User	Screen Name	Tweet	Tir
16	b'1127124430684020736'	b'438784653'	oma_ranny	b'@mamat_sura @SBYudhoyono @DonAdam08 @zarazet...	20
17	b'1127124261317922816'	b'177473596'	djenggozainal	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo Po...	20
18	b'1127124231819448321'	b'852012329541566464'	pudjiDP26	b'Pokeke #SusiDuaPeriode inilah lawan @prabowo...	20
19	b'1127124172457398272'	b'1011633112915120128'	AlistyaF	b'@AD151MO @prihnalang @rockygerung @prabowo B...	20
20	b'1127124086503534592'	b'1120038964847497216'	gelastehkaca	b'@Bunda51650806 @KPU_ID @mas_abudiman @bawasl...	20
21	b'1127124064089194497'	b'4232545247'	MDeva909	b'@prabowo @fadlizon @PartaiSocmed @Herreza @...	20
22	b'1127123885789335552'	b'1542495308'	bagussegara1	b'@prabowo @sandiuno @Prabowo_Sandi_ \nMassa s...	20
23	b'1127123880219250689'	b'1108293658506194945'	yuyun_sumiarsih	b'@abiyu991223 @prabowo Alhamdulillah puisi N...	20
24	b'1127123876826107904'	b'1100944001115406336'	sasauw_nelson	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo MU...	20
25	b'1127123813290823681'	b'1050625405965615105'	duhmnas3	b'@khokomint @MGHI79182954 @ustadtengkuzul Men...	20
26	b'1127123758890639362'	b'3222496663'	edi_sadja	b'@RachlanNashidik @prabowo @jokowi Jangan sok...	20
27	b'1127123704033316864'	b'3012991285'	GedeRamaditya15	b'@habibthink @ethnicnjh @DivHumas_Polri Gila ...	20
28	b'1127123701235834881'	b'1113126434673795072'	jacobus_arnold	b'@FerdinandHaeen2 @jokowi @prabowo @AgusYudho...	20
29	b'1127123679706472448'	b'998001176271405056'	SenopatiMuhamad	b'@bantoro29_ @prabowo @BertemanM @Dahnilarzar...	20
...
70	b'1127121002964537344'	b'1018174948811001857'	Pultak3	b'@CNNIndonesia @prabowo Segera penjarakan.\nM...	20
71	b'1127120979988082688'	b'1100944001115406336'	sasauw_nelson	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo DA...	20

	ID Tweet	ID User	Screen Name	Tweet	Tir
72	b'1127120864720211968'	b'2377146110'	DidikGani	b'@RachlanNashidik @prabowo @jokowi Win win so...	20
73	b'1127120763587153920'	b'1075410898267779072'	AgungAbriansah	b'@Kunto_Drummer @prabowo @sandiuno Kerennnn O...	20
74	b'1127120742414336002'	b'1107834703384440832'	devra2066	b'@Dahnilaranza saya sudah baca dan menduga per...	20
75	b'1127120711976284161'	b'769109771915972609'	alberta_Ys	b'@FerdinandHaeen2 @jokowi @prabowo @AgusYudho...	20
76	b'1127120488038318080'	b'1102171318131187712'	JOKER05769925	b'@RachlanNashidik @prabowo @jokowi Ketika sla...	20
77	b'1127120485240561667'	b'232511930'	ifalcrew	b'@iwanfals @TsamaraDKI @paula_oetomo @SaprixS...	20
78	b'1127120458669645824'	b'115717990'	fergianto	b'@CakKhum @Alesha0707 Cc in ah ..@prabowo'	20
79	b'1127120344207093761'	b'1103847150557945856'	TheLilies1	b'@FerdinandHaeen2 @AbahRizqy @jokowi @prabowo...	20
80	b'1127120289131745280'	b'945899333928357888'	Rangga_Wijaya89	b'@oma_ranny @FerdinandHaeen2 @KPU_ID @sandiun...	20
81	b'1127120248254058501'	b'3326317159'	yudhinulis	b'@prabowo @sandiuno sudahlah kalian mengaku K...	20
82	b'1127120060852555776'	b'824052154360938496'	hafid20hamid	b'@RachlanNashidik @prabowo @jokowi Tk akan pe...	20
83	b'1127119969748054016'	b'951783594904641536'	teguh68841451	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo Bu...	20
84	b'1127119885442535426'	b'1100944001115406336'	sasauw_nelson	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo AP...	20
85	b'1127119780857516032'	b'984696309142581248'	anjaniriani28	b'@CintaNKRI12 @prabowo @sandiuno @KPU_ID @baw...	20
86	b'1127119763237265408'	b'982495653329694721'	PermanaBenn	b'@RachlanNashidik @prabowo @jokowi Pengkhiana...	20
87	b'1127119404599136257'	b'1084862187514941444'	Gede48634982	b'Indonesia ibu kandung kita.\nJika kita biark...	20
88	b'1127119225120743424'	b'994301594'	hikmatsholeh	b'@BamsBulaksumur @Johns_purba @jokowi @prabow...	20

	ID Tweet	ID User	Screen Name	Tweet	Tir
89	b'1127119197878640640'	b'1120703549455757312'	Arioseno20	b'@FerdinandHaeen2 @prabowo @sandiuno Saya sal...	20
90	b'1127118983977508871'	b'1061053936826634246'	AzrielFirdaus3	b'@LisaAmartaTara Alhamdulillah...\nBukannya #...	20
91	b'1127118926515597314'	b'223018351'	Qois_ahmed	b'@sy_haris @AndiArief_ @prabowo UKRI BOCORKA...	20
92	b'1127118879308767232'	b'1479572520'	cebretbagus	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo sa...	20
93	b'1127118862548328448'	b'1083684635471572992'	bundananik7	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo Un...	20
94	b'1127118779178074112'	b'1105142034388574211'	marlanramazan	b'@OrangKoslet @HarryRidwan_Ay @RachlanNashidi...	20
95	b'1127118727248367616'	b'802065987159760896'	haryono1178	b'@newfiali12341 @rian4PS @prabowo @sandiuno M...	20
96	b'1127118686580461569'	b'1100944001115406336'	sasauw_nelson	b'@FerdinandHaeen2 @jokowi @KPU_ID @prabowo DI...	20
97	b'1127118639092518912'	b'1053565825007767552'	Bengben79898180	b'@dariosmanues @FerdinandHaeen2 @RanupsProSan...	20
98	b'1127118474055020544'	b'1119224253021179905'	HirlanMh	b'@PNaposo @MSApunya @prabowo Alhamdulillah.....	20
99	b'1127118423626928128'	b'1907021312'	diazika81	b'@NurlelySiregar @DameriatiZein @prabowo @put...	20

100 rows × 5 columns



In []: