

▼ Crawling Data

```
#install twint
!pip3 install twint
```

```
Collecting twint
  Downloading twint-2.1.20.tar.gz (31 kB)
Collecting aiohttp
  Downloading aiohttp-3.7.4.post0-cp37-cp37m-manylinux2014_x86_64.whl (1.3 MB)
|████████████████████████████████████████| 1.3 MB 36.5 MB/s
Collecting aiodns
  Downloading aiodns-3.0.0-py3-none-any.whl (5.0 kB)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.7/dist-packages
Collecting cchardet
  Downloading cchardet-2.1.7-cp37-cp37m-manylinux2010_x86_64.whl (263 kB)
|████████████████████████████████████████| 263 kB 71.7 MB/s
Collecting elasticsearch
  Downloading elasticsearch-7.15.1-py2.py3-none-any.whl (378 kB)
|████████████████████████████████████████| 378 kB 48.7 MB/s
Requirement already satisfied: pysocks in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from
Collecting aiohttp_socks
  Downloading aiohttp_socks-0.6.0-py3-none-any.whl (9.2 kB)
Collecting schedule
  Downloading schedule-1.1.0-py2.py3-none-any.whl (10 kB)
Requirement already satisfied: geopy in /usr/local/lib/python3.7/dist-packages (from
Collecting fake-useragent
  Downloading fake-useragent-0.1.11.tar.gz (13 kB)
Collecting googletransx
  Downloading googletransx-2.4.2.tar.gz (13 kB)
Collecting pycares>=4.0.0
  Downloading pycares-4.1.2-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylin
|████████████████████████████████████████| 291 kB 65.6 MB/s
Requirement already satisfied: cffi>=1.5.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (f
Collecting yarll<2.0,>=1.0
  Downloading yarll-1.7.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_
|████████████████████████████████████████| 271 kB 72.4 MB/s
Collecting async-timeout<4.0,>=3.0
  Downloading async_timeout-3.0.1-py3-none-any.whl (8.2 kB)
Collecting multidict<7.0,>=4.5
  Downloading multidict-5.2.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manyl
|████████████████████████████████████████| 160 kB 60.7 MB/s
Requirement already satisfied: typing-extensions>=3.6.5 in /usr/local/lib/python3.7/d
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: chardet<5.0,>=2.0 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: idna>=2.0 in /usr/local/lib/python3.7/dist-packages (f
Collecting python-socks[asyncio]>=1.2.2
  Downloading python_socks-1.2.4-py3-none-any.whl (35 kB)
Requirement already satisfied: urllib3<2,>=1.21.1 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from
```

```
Requirement already satisfied: geographiclib<2,>=1.49 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Building wheels for collected packages: twint, fake-useragent, googletransx
  Building wheel for twint (setup.py) ... done
  Created wheel for twint: filename=twint-2.1.20-py3-none-any.whl size=33930 sha256=0
  Stored in directory: /root/.cache/pip/wheels/44/fc/77/99887a36b5c265a87516158858697
```

```
!git clone --depth=1 https://github.com/twintproject/twint.git
%cd twint
!pip3 install . -r requirements.txt
```

```
Cloning into 'twint'...
remote: Enumerating objects: 47, done.
remote: Counting objects: 100% (47/47), done.
remote: Compressing objects: 100% (44/44), done.
remote: Total 47 (delta 3), reused 14 (delta 0), pack-reused 0
Unpacking objects: 100% (47/47), done.
/content/twint
Processing /content/twint
  DEPRECATION: A future pip version will change local packages to be built in-place w
  pip 21.3 will remove support for this functionality. You can find discussion regar
Requirement already satisfied: aiohttp in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: aiodns in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: cchardet in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Collecting dataclasses
  Downloading dataclasses-0.6-py3-none-any.whl (14 kB)
Requirement already satisfied: elasticsearch in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: pysocks in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: pandas>=0.23.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Collecting aiohttp_socks<=0.4.1
  Downloading aiohttp_socks-0.4.1-py3-none-any.whl (17 kB)
Requirement already satisfied: schedule in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: geopy in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: fake-useragent in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: googletransx in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: attrs>=19.2.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: typing-extensions>=3.6.5 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: chardet<5.0,>=2.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: async-timeout<4.0,>=3.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: idna>=2.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: pycares>=4.0.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: cffi>=1.5.0 in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (from twint==2.1.20)
```

```

Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: urllib3<2,>=1.21.1 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: geographiclib<2,>=1.49 in /usr/local/lib/python3.7/dis
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (fro
Building wheels for collected packages: twint
  Building wheel for twint (setup.py) ... done
  Created wheel for twint: filename=twint-2.1.21-py3-none-any.whl size=38870 sha256=1
  Stored in directory: /tmp/pip-ephem-wheel-cache-t9givcxe/wheels/f7/3e/11/2803f3c6890
Successfully built twint
Installing collected packages: dataclasses, aiohttp-socks, twint
  Attempting uninstall: aiohttp-socks
    Found existing installation: aiohttp-socks 0.6.0
    Uninstalling aiohttp-socks-0.6.0:
      Successfully uninstalled aiohttp-socks-0.6.0
  Attempting uninstall: twint
    Found existing installation: twint 2.1.20
    Uninstalling twint-2.1.20:
      Successfullv uninstalled twint-2.1.20

```

```

import nest_asyncio
nest_asyncio.apply()# used once to enable concurrent actions within a Jupyter notebook.

```

```
import twint
```

```

#mengambil data tweet
c = twint.Config()
c.Search = '#PercumaLaporPolisi'
c.Pandas = True

```

```
twint.run.Search(c)
```

```

1453721693147394060 2021-10-28 13:53:55 +0000 <Asroel_Boz> #PercumaLaporPolisi
1453706666688405504 2021-10-28 12:54:12 +0000 <nalarTV> Jelas2 kejadian kekerasannya
1453703773344264196 2021-10-28 12:42:42 +0000 <bangded82122889> @txttdarisange @txttd
1453700525455659015 2021-10-28 12:29:48 +0000 <SugiKawuloAlit> Apabila pelanggaran hu
1453691026355286016 2021-10-28 11:52:03 +0000 <singgih_berlan> tapi ttp aja #PercumaL
1453688639079141379 2021-10-28 11:42:34 +0000 <DianaHeartily> @geloraco #percumalapor
1453676582309007364 2021-10-28 10:54:39 +0000 <Iwan_SR> @detikcom #percumalaporpolisi
1453664023904350209 2021-10-28 10:04:45 +0000 <_pamflet> Di Hari Sumpah Pemuda ini, k
1453660794835460098 2021-10-28 09:51:55 +0000 <KitaNgememe> @AmsyarSyafiie2 @zoelfick
1453650976053805061 2021-10-28 09:12:54 +0000 <bocahmalang77> @3Rauuul @FajarMu81 @ke
1453650159833792520 2021-10-28 09:09:40 +0000 <teataehosea> Gua pernah ngurus buat la
1453649938248785923 2021-10-28 09:08:47 +0000 <bocahmalang77> #percumalaporpolisi
1453648561904648205 2021-10-28 09:03:19 +0000 <indotanahbeta> @kompascom #percumalapo
1453647750998872064 2021-10-28 09:00:05 +0000 <rumahnenekkk> @AmsyarSyafiie2 @zoelfic
1453644988831596545 2021-10-28 08:49:07 +0000 <TirtoID> #PercumaLaporPolisi, lalu apa
1453642897224503300 2021-10-28 08:40:48 +0000 <bocahmalang77> @tubirfess Si SYAIFUL A
1453639157553582084 2021-10-28 08:25:57 +0000 <altauf2004> @MCAOps polisi pengecut... p
1453635779268284416 2021-10-28 08:12:31 +0000 <thefineryreport> Sejak #PercumaLaporPo
1453633669978021891 2021-10-28 08:04:08 +0000 <bocahmalang77> @vichan0717 @kwanieonyo
1453624841089667072 2021-10-28 07:29:03 +0000 <Devamoyas> @tubirfess #percumalaporpol
1453619193228644356 2021-10-28 07:06:37 +0000 <amorakook> aowkwowk #percumalaporpolis

```

1453612563619405827	2021-10-28	06:40:16	+0000	<doydoydoyyyy> @LotusWist @bertanyarl #
1453610330454564865	2021-10-28	06:31:24	+0000	<Ezharhap> @detikcom Orang kehilangan b
1453604344981442564	2021-10-28	06:07:37	+0000	<SugiKawuloAlit> @tvOneNews Wajarlah ka
1453601334221963266	2021-10-28	05:55:39	+0000	<suluhpergerakan> #percumalaporpolisi
1453600846722191366	2021-10-28	05:53:43	+0000	<oedaykita> Mayan lah, biar keliatan te
1453597370738876417	2021-10-28	05:39:54	+0000	<blackwafferroll> @lustervantt @AREAJUL
1453596806684708865	2021-10-28	05:37:39	+0000	<blackwafferroll> @AREAJULID Apa boleh
1453587086750015488	2021-10-28	04:59:02	+0000	<kamisanbdg> [Aksi Kamisan Bandung Ke-3
1453586960631558144	2021-10-28	04:58:32	+0000	<aliyahS88887990> #percumalaporpolisi
1453586284706877447	2021-10-28	04:55:51	+0000	<BuruhKasar88> @Irwan2yah @Rm_lndr4one4
1453578897673490435	2021-10-28	04:26:30	+0000	<Ilhamnfadilah> @SeputarTetangga #percum
1453570302189789190	2021-10-28	03:52:20	+0000	<rafy_hananta> @AREAJULID mereka penggi
1453569624428011520	2021-10-28	03:49:39	+0000	<beritaKBR> Kepercayaan publik ke Polri
1453562228595052545	2021-10-28	03:20:15	+0000	<oedaykita> Nyatanya banyak kepala sepe
1453561584274403331	2021-10-28	03:17:42	+0000	<rivaanggara> Indomaret selaku pemilik
1453559307342016514	2021-10-28	03:08:39	+0000	<Merimer50024844> @AREAJULID Ini salah
1453556366333730824	2021-10-28	02:56:58	+0000	<fixindonesiacom> Belakang media sosial
1453551908619051010	2021-10-28	02:39:15	+0000	<killingsinside00> @detikcom #percumalapo
1453548992709623812	2021-10-28	02:27:40	+0000	<JSusiono> #percumalaporpolisi https://
1453535231659569153	2021-10-28	01:32:59	+0000	<DDRezhaa> @Ameeranti #PercumaLaporPoli
1453535169420292104	2021-10-28	01:32:44	+0000	<RamadhanRenwrin> Mereka ini lagi apes,
1453534628317315073	2021-10-28	01:30:35	+0000	<JamalBoegis> @Ameeranti Kemarin ada ta
1453528784213209090	2021-10-28	01:07:22	+0000	<buhori93> @OposisiCerdas #PercumaLapor
1453528736935022593	2021-10-28	01:07:10	+0000	<MakaniNetizen> @kumparan Wah ngak jela
1453495947015557120	2021-10-27	22:56:53	+0000	<bodumeken> Ormas ini masih paham tidak
1453488830246436864	2021-10-27	22:28:36	+0000	<Agahogok> Jadi sbnrnya Soekarno tahu a
1453483460320919555	2021-10-27	22:07:16	+0000	<MisUlum> Merasa punya kuasa jadi peneg
1453418316685447168	2021-10-27	17:48:24	+0000	<masobihil> #PercumaLaporPolisi ketika i
1453410322681720837	2021-10-27	17:16:38	+0000	<masobihil> Akankah #PercumaLaporPolisi
1453393929093648391	2021-10-27	16:11:30	+0000	<SahrulFalah17> @OposisiCerdas #Percuma
1453388577291657218	2021-10-27	15:50:14	+0000	<xtrsyz> klo udah kaya begini #PercumaL
1453388294801162240	2021-10-27	15:49:06	+0000	<adv_supyadi> @DDRezhaa Hidup lagi taga
1453386841298587653	2021-10-27	15:43:20	+0000	<DDRezhaa> @adv_supyadi Semoga tidak #P
1453374228988792832	2021-10-27	14:53:13	+0000	<linsay1008> Mencoba untuk tetap percay
14533718444468215813	2021-10-27	14:43:44	+0000	<tahucabegarem> Apakah aku akan menjadi

```
#convert hasil crawling menjadi dataset
Tweets_df = twint.storage.panda.Tweets_df
Tweets_df.head()
```

	id	conversation_id	created_at	date	timezone	place
0	1453721693147394060	1453721693147394060	1.635429e+12	2021-10-28 13:53:55	+0000	#F
1	1453706666688405504	1453706666688405504	1.635426e+12	2021-10-28 13:53:55	+0000	

```
#menyimpan dataset sebagai csv
Tweets_df.to_csv("HasilCrawling.csv")
```

12:42:42

▼ Data Preprocessing

Prepare dataset

11:52:03

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('HasilCrawling.csv')
```

```
df.head()
```

	id	conversation_id	created_at	date	timezone	place
0	1.448260e+18	1.448230e+18	1.634130e+12	10/13/2021 19:19	700	NaN @BossTemle Belanda be
1	1.448260e+18	1.448250e+18	1.634130e+12	10/13/2021 19:18	700	NaN @Komp Belanda berke
2	1.448260e+18	1.448230e+18	1.634130e+12	10/13/2021 19:18	700	NaN @BossTemle #PercumaL
3	1.448260e+18	1.448260e+18	1.634130e+12	10/13/2021 19:18	700	NaN #PercumaL #PolisiSesuai
4	1.448260e+18	1.448190e+18	1.634130e+12	10/13/2021 19:18	700	NaN @CNNIndor Belanda t

```
df.shape
```

```
(5011, 38)
```

```
#menghapus kolom yang tidak dibutuhkan
```

```
df = df.drop(df.columns[[0,1,2,3,4,5,7,8,9,10,11,12,13,14,15,16,17,18,
                        19,20,21,22,23,24,25,26,27,28,29,30,
                        31,32,33,34,35,36,37]],axis=1)
```

```
df.head()
```

	tweet
0	@BossTemlen Opas Belanda bertindak berlebihan ...
1	@KompasTV Opas Belanda bertindak berlebihan dl...
2	@BossTemlen Naikkan lagi #PercumaLaporPolisi
3	#PercumaLaporPolisi kalo #PolisiSesuaiProsedur...
4	@CNNIndonesia Opas Belanda bertindak berlebiha...

Cleaning Dataset (Menghilangkan Tanda Baca dan Karakter Non-alfabet)

```
import re
```

```
#fungsi untuk menghapus semua karakter non-alfabet
```

```
def clean(text):
    text = re.sub('[^A-Za-z]+', ' ', text)
    return text
```

```
df['tweet'] = df['tweet'].apply(clean)
```

```
df.head()
```

	tweet
0	BossTemlen Opas Belanda bertindak berlebihan ...
1	KompasTV Opas Belanda bertindak berlebihan dl...
2	BossTemlen Naikkan lagi PercumaLaporPolisi
3	PercumaLaporPolisi kalo PolisiSesuaiProsedur ...
4	CNNIndonesia Opas Belanda bertindak berlebiha...

Case Folding (Merubah Seluruh Huruf Menjadi Huruf Kecil)

```
import re

def casefolding(tweet):
    tweet = tweet.lower()
    tweet = tweet.strip(" ")
    tweet = re.sub(r'[?|$.|!^_:")(-+.]', '', tweet)
    return tweet

df['tweet']=df['tweet'].apply(casefolding)

df.head()
```

	tweet
0	bosstemlen opas belanda bertindak berlebihan d...
1	kompastv opas belanda bertindak berlebihan dlm...
2	bosstemlen naikkan lagi percumalaporpolisi
3	percumalaporpolisi kalo polisisesuaiprosedur a...
4	cnnindonesia opas belanda bertindak berlebihan...

Tokenizing (Memecah Kata Menjadi Bagian Kata)

```
def token(tweet):
    nstr = tweet.split(' ')
    dat = []
    a = -1
    for hu in nstr:
        a = a+1
        if hu=='':
            dat.append(a)
    p = 0
    b = 0
    for q in dat :
        b = q-p
        del nstr[b]
        p = p+1
    return nstr

df['tweet']=df['tweet'].apply(token)

df.head()
```

	tweet
0	[bosstemlen, opas, belanda, bertindak, berlebi...
1	[kompastv, opas, belanda, bertindak, berlebiha...
2	[bosstemlen, naikkan, lagi, percumalaporpolisi]
3	[percumalaporpolisi, kalo, polisisesuaiprosedu...

Filtering (Stopword Removal)

```
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords

def stopword_removal(tweet):
    filtering = stopwords.words('indonesian', 'english')
    x = []
    df = []
    def myFunc(x):
        if x in filtering :
            return False
        else :
            return True
    fit = filter(myFunc, tweet)
    for x in fit :
        df.append(x)
    return df

df['tweet']=df['tweet'].apply(stopword_removal)

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

df.head()
```

	tweet
0	[bosstemlen, opas, belanda, bertindak, dlm, na...
1	[kompastv, opas, belanda, bertindak, dlm, naga...
2	[bosstemlen, naikkan, percumalaporpolisi]
3	[percumalaporpolisi, kalo, polisisesuaiprosedu...
4	[cnnindonesia, opas, belanda, bertindak, dlm, ...

Stemming (Tranformasi Kata ke Kata Dasarnya)


```
!pip install Sastrawi
```

```
Collecting Sastrawi
```

```
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
```

```
    |████████████████████████████████████████| 209 kB 5.4 MB/s
```

```
Installing collected packages: Sastrawi
```

```
Successfully installed Sastrawi-1.0.1
```

```
from sklearn.pipeline import Pipeline
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

```
def stemming(tweet):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in tweet:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean
```

```
df['stemming'] = df['tweet'].apply(stemming)
```

Streaming output truncated to the last 5000 lines.

```
republikaonline opas belanda tindak dlm naganin demo jk mahasiswa angkat senjata miya
geloraco percumalaporpolisi
tempodotco opas belanda tindak dlm naganin demo jk mahasiswa angkat senjata miyanmar
demo hut tangerang ricuh percumalaporpolisi https t co yrat uagvj
percumalaporpolisi
geloraco percumalaporpolisi
demo hut tangerang rusuh demo jatuh percumalaporpolisi https t co v icxm drp
buzzer polisi naik tagar percumalaporpolisi polisikriminal
percumaadapolisi percumalaporpolisi
maarifwizdanul republikaonline percumalaporpolisi
selasapagi edan percumalaporpolisi
biadap seragam percumalaporpolisi
percumalaporpolisi smackdown smackdownlive
tempodotco ga percaya percumalaporpolisi polisisesuaiprosedur polisihumanis
aksilangsung polritegashumanis percumalaporpolisi
polisi bantah pakai buzzer lawan tagar percumalaporpolisi https t co ivsfwtwdye
percumalaporpolisi
cita cita ken main mma paksa ortunya suruh polisi makan smackdown percumalaporpolisi
kerja buruk polisi lahir tagar percumalaporpolisi evaluasi sambut dg tagar tanding po
percumalaporpolisi acab
trending percumalaporpolisi eh video polisi ngebanting mahasiswa ujukrasa mantap oknu
republikaonline percumaadapolisi percumalaporpolisi
percumalaporpolisi
wih karang musim tandingannya kemaren percumalaporpolisi karang tagar polisisesuaipro
this is both so sad amp sickening it is maddeningly funny percumalaporpolisi in every
pas banget berita pas sudut perkara percumalaporpolisi iya kerja polrisesuaiprosedur
percumalaporpolisi
```

selasapagi percumalaporpholisi ga laku polisi
 udah lapor polisi demo banting sampe kejang kejang polisi lindung rakyat ngerti ga ra
 areajulid aja bully pake hastag percumalaporpholisi eh bikin ulah g breafing dlu ama k
 tubirfess harap polkis percumalaporpholisi
 percumalaporpholisi
 percumalaporpholisi
 geloraco percumalaporpholisi gilir korban penjara serba salah emang hidup wkwwkwlnd
 beliau nggak ngecek kali ya tagar percumalaporpholisi mah beneran fakta resah masyarakat
 percumalaporpholisi tau nggak sih besar gaji polisi indonesia https t co plb ujavb
 percumalaporpholisi
 ya bener emang percumalaporpholisi kdrt dijadiin sangka tolong
 glrhn hubung percumalaporpholisi https t co plb ujavb
 ylbhi percumalaporpholisi
 polri perang tagar media sosial percumalaporpholisi polisisesuaiprosedur ranah polisi
 percumalaporpholisi divhumas polri polhukamri kontras lokataru id tvonetv dpr ri
 geloraco percumalaporpholisi polisisesuaisopngawur polisipenyiksarakyat
 selasapagi percumalaporpholisi
 atie isk gibraltarns sitalopa lapor mana percumalaporpholisi percumalaporpholisi
 sih polisi smackdown polisi ramah polisi australia percumalaporpholisi
 aksilangsung percumalaporpholisi
 heran rakyat bilang percumalaporpholisi
 an kiim sarap kejang dokter kmana percumalaporpholisi percumalaporpholisi
 percumalaporpholisi it s just a localized acab how did we end up here
 vivacoid syamsud naik percumalaporpholisi percumalaporpholisi
 percumalaporpholisi
 txtdaritng percumalaporpholisi
 ampun si jago smackdown percumalaporpholisi
 detikcom rakyat susah diam rakyat bicara bungkam rakyat hantam percumalaporpholisi
 gantieol percumalaporpholisi tukang smackdown gitu sesuai prosedur divhumas polri

```
df.head()
```

	tweet	stemming
0	[bosstemlen, opas, belanda, bertindak, dlm, na...	bosstemlen opas belanda tindak dlm naganin dem...
1	[kompastv, opas, belanda, bertindak, dlm, naga...	kompastv opas belanda tindak dlm naganin demo ...
2	[bosstemlen, naikkan, percumalaporpholisi]	bosstemlen naik percumalaporpholisi
3	[percumalaporpholisi, kalo, polisisesuaiprosedu...	percumalaporpholisi kalo polisisesuaiprosedur a...

```
df.to_csv('DataCleaned.csv', index=False)
```

▼ Sentiment Analysis Using TextBlob

```
fin_data = pd.read_csv('DataCleaned.csv')
```

```
fin_data.head()
```

	tweet	stemming
0	['bosstemlen', 'opas', 'belanda', 'bertindak', ...	bosstemlen opas belanda tindak dlm naganin dem...
1	['kompastv', 'opas', 'belanda', 'bertindak', '...	kompastv opas belanda tindak dlm naganin demo ...
2	['bosstemlen', 'naikkan', 'percumalaporpholisi']	bosstemlen naik percumalaporpholisi
3	['percumalaporpholisi', 'kalo', 'polisisesuaipr...	percumalaporpholisi kalo polisisesuaiprosedur a...
4	['cnnindonesia', 'opas', 'belanda', 'bertindak...	cnnindonesia opas belanda tindak dlm naganin d...

```
!pip install textblob
```

```
Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages (0.15
Requirement already satisfied: nltk>=3.1 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from nltk:
```



```
from textblob import TextBlob
```

```
#fungsi untuk menghitung subjektivitas
```

```
def getSubjectivity(review):
```

```
    return TextBlob(review).sentiment.subjectivity
```

```
#fungsi untuk menghitung polarity
```

```
def getPolarity(review):
```

```
    return TextBlob(review).sentiment.polarity
```

```
#fungsi untuk menganalisis tweet
```

```
def analysis(score):
```

```
    if score < 0:
```

```
        return 'Negative'
```

```
    elif score == 0:
```

```
        return 'Neutral'
```

```
    else:
```

```
        return 'Positive'
```

```
# fin_data['Subjectivity'] = fin_data['Lemma'].apply(getSubjectivity)
```

```
fin_data['Polarity'] = fin_data['stemming'].apply(getPolarity)
```

```
fin_data['Analysis'] = fin_data['Polarity'].apply(analysis)
```

```
fin_data.head()
```

	tweet	stemming	Polarity	Analysis
0	['bosstemlen', 'opas', 'belanda', 'bertindak', ...]	bosstemlen opas belanda tindak dlm naganin dem...	0.0	Neutral
1	['kompastv', 'opas', 'belanda', 'bertindak', ...]	kompastv opas belanda tindak dlm naganin demo ...	0.0	Neutral

```
tb_counts = fin_data.Analysis.value_counts()
```

```
tb_counts
```

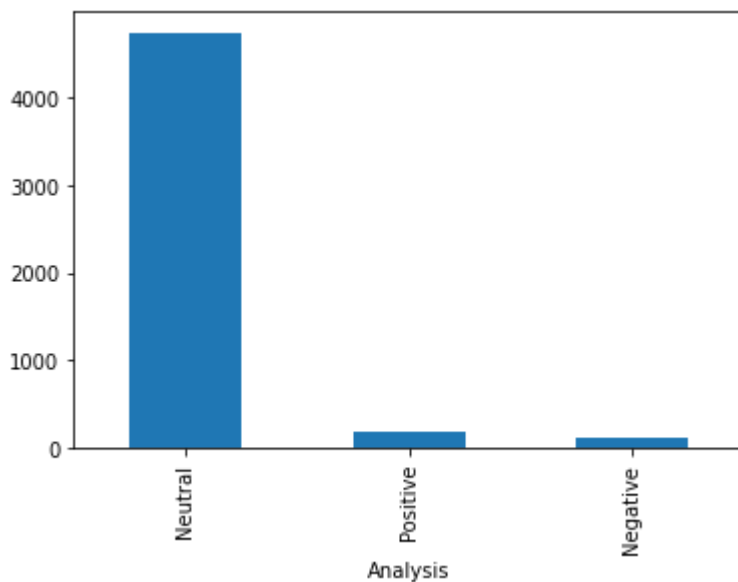
```
Neutral      4738
Positive      170
Negative      103
Name: Analysis, dtype: int64
```

```
import matplotlib.pyplot as plt
```

```
%matplotlib inline
```

```
fin_data.value_counts('Analysis').plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc1926f2d10>
```



Wordcloud

```
#memisahkan positif,negatif,dan netral tweet
```

```
negative=fin_data[fin_data['Analysis']=='Negative'].stemming
```

```
neutral=fin_data[fin_data['Analysis']=='Neutral'].stemming
```

```
positive=fin_data[fin_data['Analysis']=='Positive'].stemming
```

```
#menggabungkan tweet
```

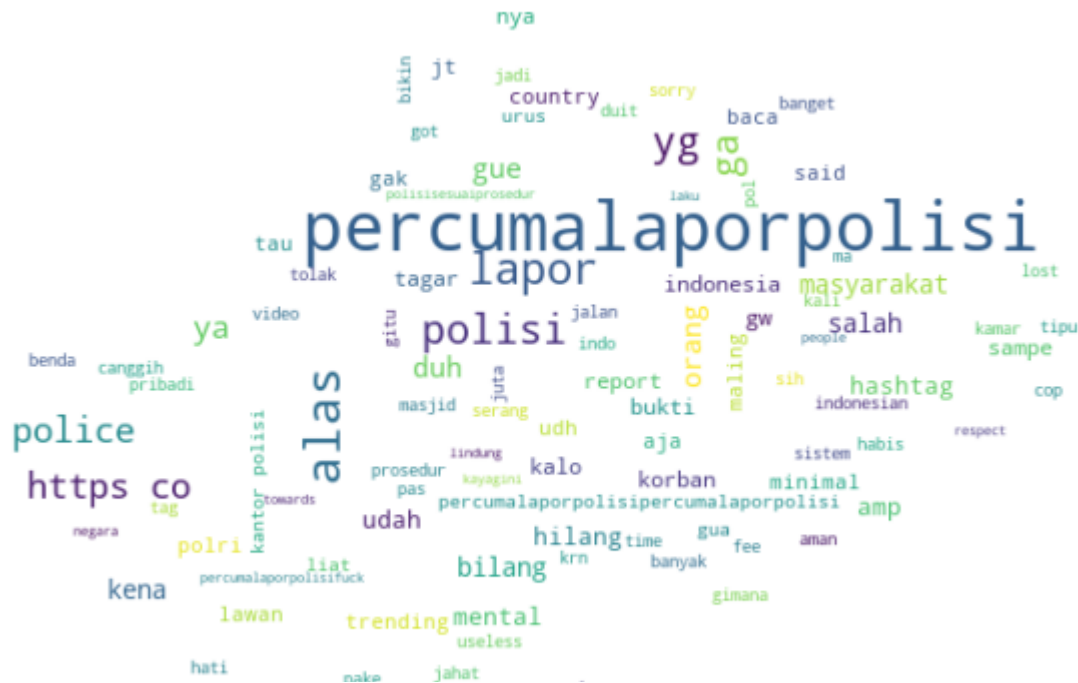
```
word_negative=''.join(negative)
```

```
word_neutral=''.join(neutral)
```

```
word_positive=''.join(positive)
```

```
plt.show(makingWordCloud(word_negative))
```

Word Cloud of Negative Tweet



```
plt.figure(figsize=(10,10))
plt.title('Word Cloud of Neutral Tweet')
plt.show(makingWordCloud(word_neutral))
```

Word Cloud of Neutral Tweet



TF-IDF (Menghitung Bobot Setiap Kata di Dokumen)

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf=TfidfVectorizer()
```

```
tfidf.fit(fin_data['stemming'])
```

```
TfidfVectorizer(analyzer='word', binary=False, decode_error='strict',
                dtype=<class 'numpy.float64'>, encoding='utf-8',
                input='content', lowercase=True, max_df=1.0, max_features=None,
                min_df=1, ngram_range=(1, 1), norm='l2', preprocessor=None,
                smooth_idf=True, stop_words=None, strip_accents=None,
                sublinear_tf=False, token_pattern='(?u)\\b\\w\\w+\\b',
                tokenizer=None, use_idf=True, vocabulary=None)
```

```
tfidf.vocabulary_
```

```
{'bosstemlen': 1012,
 'opas': 5731,
 'belanda': 736,
 'tindak': 8000,
 'dlm': 1860,
 'naganin': 5157,
 'demo': 1617,
 'jk': 3494,
 'mahasiswa': 4625,
 'angkat': 297,
 'senjata': 7147,
 'miyanmar': 4971,
 'mikir': 4930,
 'cc': 1232,
 'mohmahfudmd': 5024,
 'kompolnas': 4075,
 'ri': 6758,
 'percumalaporpholisi': 6049,
 'https': 2978,
 'co': 1368,
 'ej': 2037,
 'nf': 5294,
 'kompastv': 4069,
 'naik': 5162,
 'kalo': 3652,
 'polisisesuaiprosedur': 6255,
 'anjaayyy': 307,
 'cnnindonesia': 1365,
 'aksilangsung': 161,
 'henrysubiakto': 2873,
 'erenwe': 2140,
 'percumaadapolisi': 6032,
 'republikaonline': 6706,
 'geloraco': 2527,
 'tempodotco': 7861,
 'hut': 3014,
 'tangerang': 7745,
```

```
'ricuh': 6767,
'yrat': 8915,
'uagvj': 8268,
'rusuh': 6872,
'jatuh': 3402,
'icxm': 3052,
'drp': 1944,
'buzzer': 1159,
'polisi': 6229,
'tagar': 7696,
'polisikriminal': 6244,
'maarifwizdanul': 4598,
'selasapagi': 7100,
'edan': 2004,
'biadap': 852,
'seragam': 7164,
'smackdown': 7401,
'smackdownlive': 7402,
'ga': 2398,
'percaya': 6028,
'polisihumanis': 6220
```

▼ Handling Imbalanced Labeling in Dataset with SMOTE

Karena peng-kategorian yang didapatkan tidak seimbang antara kategori positif, negatif, dan netral maka dibutuhkan proses ini untuk meningkatkan kualitas dari train dan test nantinya. Karena jika peng-kategorian yang didapatkan hanya dominan pada salah satu kategori, maka train yang digunakan tidak mendukung untuk mendapatkan classification yang bagus.

Label Encoding of Analysis

```
fin_data['Analysis']

0      Neutral
1      Neutral
2      Neutral
3      Neutral
4      Neutral
...
5006   Neutral
5007   Neutral
5008   Neutral
5009   Neutral
5010   Neutral
Name: Analysis, Length: 5011, dtype: object
```

```
from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()
```



```
result=lb.fit_transform(fin_data['Analysis'])
```

```
result
```

```
array([1, 1, 1, ..., 1, 1, 1])
```

```
df2 = pd.DataFrame(result)
```

```
df2
```

	0
0	1
1	1
2	1
3	1
4	1
...	...
5006	1
5007	1
5008	1
5009	1
5010	1

```
5011 rows × 1 columns
```

```
# menggabungkan dataframe fin_data dan df2
```

```
df3=pd.concat([fin_data,df2],axis=1)
```

```
df3
```

	tweet	stemming	Polarity	Analysis	0
0	['bosstemlen', 'opas', 'belanda', 'bertindak', ...]	bosstemlen opas belanda tindak dlm naganin dem...	0.0	Neutral	1
1	['kompastv', 'opas', 'belanda', 'bertindak', ...]	kompastv opas belanda tindak dlm naganin demo ...	0.0	Neutral	1

```
df4=df3.drop(['Analysis'],axis=1)
df4
```

	tweet	stemming	Polarity	0
0	['bosstemlen', 'opas', 'belanda', 'bertindak', ...]	bosstemlen opas belanda tindak dlm naganin dem...	0.0	1
1	['kompastv', 'opas', 'belanda', 'bertindak', ...]	kompastv opas belanda tindak dlm naganin demo ...	0.0	1
2	['bosstemlen', 'naikkan', 'percumalaporpolisi']	bosstemlen naik percumalaporpolisi	0.0	1
3	['percumalaporpolisi', 'kalo', 'polisisesuaipr...]	percumalaporpolisi kalo polisisesuaiprosedur a...	0.0	1
4	['cnnindonesia', 'opas', 'belanda', 'bertindak', ...]	cnnindonesia opas belanda tindak dlm naganin d...	0.0	1
...
5006	['tvonenews', 'anaku', 'motornya', 'begal', 't...]	tvonenews ana motor begal trs lapor jg tanggap...	0.0	1
5007	['hukum', 'indonesia', 'tegak', 'percumalaporp...]	hukum indonesia tegak percumalaporpolisi krist...	0.0	1
...

```
fin_data=df4.rename(columns={0:'Analysis'})
fin_data
```

	tweet	stemming	Polarity	Analysis
0	['bosstemlen', 'opas', 'belanda', 'bertindak',...]	bosstemlen opas belanda tindak dlm naganin dem...	0.0	1
1	['kompastv', 'opas', 'belanda', 'bertindak',...]	kompastv opas belanda tindak dlm naganin dem...	0.0	1

Transform TF-IDF to Array

```
tfidf_dtm =tfidf.transform(fin_data['stemming'])
tfidf_dtm

<5011x9009 sparse matrix of type '<class 'numpy.float64'>'
  with 46191 stored elements in Compressed Sparse Row format>
```

```
df5=tfidf_dtm.toarray()
df5
```

```
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

```
x = df5
y = fin_data['Analysis']
```

Transformasi Imbalance Dataset ke Balance Dataset

```
from imblearn.over_sampling import SMOTE
```

```
smt=SMOTE()
x_sm,y_sm=smt.fit_resample(x,y)
df6=pd.DataFrame(y_sm)
df6.value_counts()
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: F
  warnings.warn(msg, category=FutureWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: F
  warnings.warn(msg, category=FutureWarning)
2    4738
1    4738
0    4738
dtype: int64
```

▼ Classification

Train and Test Split (7:3)

```
from sklearn.model_selection import train_test_split
```

```
x1_train,x1_test,y1_train,y1_test=train_test_split(x_sm,y_sm,test_size=0.3,random_state=42)
```

```
x1_train
```

```
array([[0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       ...,  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.]])
```

```
y1_train
```

```
array([2, 0, 0, ..., 0, 1, 0])
```

```
x1_test
```

```
array([[0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       ...,  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.]])
```

```
y1_test
```

```
array([0, 2, 2, ..., 0, 0, 0])
```

Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier  
rf=RandomForestClassifier()  
rf.fit(x1_train,y1_train)  
predicted2=rf.predict(x1_test)  
predicted2
```

```
array([0, 2, 2, ..., 0, 0, 0])
```

Classification Report and F1-score

```
from sklearn.metrics import classification_report,f1_score,accuracy_score
```

```
#classification report dan f1-score  
cr=classification_report(y1_test,predicted2)  
fscore=f1_score(y1_test,predicted2,average=None)  
accuracy=accuracy_score(y1_test,predicted2)  
print(cr)  
print(fscore)  
print(accuracy)
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1433
1	0.99	1.00	1.00	1443
2	1.00	1.00	1.00	1389
accuracy			1.00	4265
macro avg	1.00	1.00	1.00	4265
weighted avg	1.00	1.00	1.00	4265

```
[0.99860238 0.99620035 0.99747566]  
0.9974208675263775
```