```
# link: https://github.com/emiljdd/Tweepy-SparkTwitterI
# pip install textblob
import csv
from textblob import TextBlob
```

SENTIMENT ANALYSIS

- Sentiment polarity for an element defines the orientation of the expressed sentiment, i.e., it determines if the text expresses the positive, negative or neutral sentiment of the user about the entity in consideration.
- Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].
- https://planspace.org/20150607-textblob_sentiment/
- https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/#:~:text=Polarity%20is%20float%20which%20lies,of%20%5B0%2C1%5D.

```
import pandas as pd


# Function to tead txt file and convert to csv file


def read_and_pre_pro(file_in, file_out):
    sentences = []
    sentiment_polarity = []
    sentiment_subjectivity = []
    with open(file_in, 'r') as csvfile:
        rows = csv.reader(csvfile)
        for row in rows:
            sentence = row[0]
            blob = TextBlob(sentence)
            if ("Error on_data" not in sentence):
                #print (sentence)
                #print (blob.sentiment.polarity, blob.sentiment.subjectivity)
                sentences.append(sentence)
                sentiment_polarity.append(blob.sentiment.polarity)
                sentiment_subjectivity.append(blob.sentiment.subjectivity)
        data = pd.DataFrame({"sentence": sentences,
                     "sentiment_polarity":sentiment_polarity,
                     "sentiment_subjectivity":sentiment_subjectivity
                    })
        data.sentence = data.sentence.str.replace("b'", "")
        # con rat nhieu tien xu ly khac
```

```
        data.to_csv(file_out)


file_in = "tweets_covid_19.txt"
file_out = "tweets_covid_19.csv"
read_and_pre_pro(file_in, file_out)


df = pd.read_csv(file_out, index_col=0)


df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2221 entries, 0 to 2220
Data columns (total 3 columns):
sentence                 2221 non-null object
sentiment_polarity       2221 non-null float64
sentiment_subjectivity   2221 non-null float64
dtypes: float64(2), object(1)
memory usage: 69.4+ KB
```

```
df.head()
```

	sentence	sentiment_polarity	sentiment_subjectivity
0	Listening on port: 5555	0.00	0.0
1	Received request from: ('127.0.0.1'	-0.75	1.0
2	Helping Job-seekers in the Covid-19\xc2\xa0Eco...	0.00	0.0
3	RT @justinpalaces: COVID-19 unit at Doctors He...	0.00	0.0

```
indexNames = df[df['sentence'].str.contains("Listening on port")].index
# Delete these row indexes from dataFrame
df = df.drop(indexNames)


indexNames = df[df['sentence'].str.contains("Received request from")].index
# Delete these row indexes from dataFrame
df = df.drop(indexNames)


df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2219 entries, 2 to 2220
Data columns (total 3 columns):
```

```
df.head()
```

	sentence	sentiment_polarity	sentiment_subjectivity
2	Helping Job-seekers in the Covid-19\xc2\xa0Eco...	0.000000	0.000000
3	RT @justinpalaces: COVID-19 unit at Doctors He...	0.000000	0.000000
4	RT @Franklin_Graham: .@TheHermanCain	0.000000	0.000000
5	District Administration	0.000000	0.000000