Twitter Data Sentiment Analysis (Python: NLP and Machine Learning)

- Retrieved 30K+ tweets from Twitter API using tweepy, parsed tweets using TextBlob, and created Word Cloud of repeated words
- Classified tweets for input a query as positive, negative, or neutral with 94% accuracy using sentiment classifier trained on Naïve Bayes

The 3 major steps in our program:

1. Authorize twitter API client.
2. Make a GET request to Twitter API to fetch tweets for a particular query.
3. Parse the tweets. Classify each tweet as positive, negative or neutral.

```
In [22]:  import re
          import tweepy
          from tweepy import OAuthHandler
          from textblob import TextBlob
```

1. First of all, we create a TwitterClient class. This class contains all the methods to interact with Twitter API and parsing tweets. We use **init** function to handle the authentication of API client.
2. In get_tweets functiIn get_tweet_sentiment we use textblob module.on, we use: fetched_tweets = self.api.search(q = query, count = count) to call the Twitter API to fetch tweets.
3. In get_tweet_sentiment we use textblob module. analysis = TextBlob(self.clean_tweet(tweet))

TextBlob:

A high level library built over top of NLTK library. First we call clean_tweet method to remove links, special characters, etc. from the tweet using some simple regex. Then, as we pass tweet to create a TextBlob object, following processing is done over text by textblob library:

- Tokenize the tweet ,i.e split words from body of text.
- Remove stopwords from the tokens.(stopwords are the commonly used words which are irrelevant in text analysis like I, am, you, are, etc.)
- Do POS(part of speech) tagging of the tokens and select only significant features/tokens like adjectives, adverbs, etc.
- Pass the tokens to a sentiment classifier which classifies the tweet sentiment as positive, negative or neutral by assigning it a polarity between -1.0 to 1.0 .
- Here is how sentiment classifier is created:

TextBlob uses a Movies Reviews dataset in which reviews have already been labelled as positive or negative. Positive and negative features are extracted from each positive and negative review respectively. Training data now consists of labelled positive and negative features. This data is trained on a Naive Bayes Classifier.

Then, we use sentiment.polarity method of TextBlob class to get the polarity of tweet between -1 to 1. Then, we classify polarity as: positive, negative and neutral tweets

```python
In [38]: class TwitterClient(object):
             '''
             Generic Twitter Class for sentiment analysis.
             '''
             def __init__(self):
                 '''
                 Class constructor or initialization method.
                 '''
                 # keys and tokens from the Twitter Dev Console
                 consumer_key = 'XXXXXXXXXXXXXXXXXXXXX'
                 consumer_secret = 'XXXXXXXXXXXXXXXXXXXXXXXXX'
                 access_token = 'XXXXXXXXXXXXXXXXXXXXXXXXX'
                 access_token_secret = 'XXXXXXXXXXXXXXXXXXXXXXXXX'

                 # attempt authentication
                 try:
                     # create OAuthHandler object
                     self.auth = OAuthHandler(consumer_key, consumer_secret)
                     # set access token and secret
                     self.auth.set_access_token(access_token, access_token_secret)
                     # create tweepy API object to fetch tweets
                     self.api = tweepy.API(self.auth)
                 except:
                     print("Error: Authentication Failed")

             def clean_tweet(self, tweet):
                 '''
                 Utility function to clean tweet text by removing links, special characters
                 using simple regex statements.
                 '''
                 return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) |(\w+:\/\/\S+)", " ", tweet).split())

             def get_tweet_sentiment(self, tweet):
                 '''
                 Utility function to classify sentiment of passed tweet
                 using textblob's sentiment method
                 '''
                 # create TextBlob object of passed tweet text
                 analysis = TextBlob(self.clean_tweet(tweet))
                 # set sentiment
                 if analysis.sentiment.polarity > 0:
                     return 'positive'
                 elif analysis.sentiment.polarity == 0:
                     return 'neutral'
                 else:
                     return 'negative'

             def get_tweets(self, query, count = 10):
                 '''
                 Main function to fetch tweets and parse them.
                 '''
                 # empty list to store parsed tweets
                 tweets = []

                 try:
                     # call twitter api to fetch tweets
                     fetched_tweets = self.api.search(q = query, count = count)

                     # parsing tweets one by one
                     for tweet in fetched_tweets:
```

```python
            # empty dictionary to store required params of a tweet
            parsed_tweet = {}

            # saving text of tweet
            parsed_tweet['text'] = tweet.text
            # saving sentiment of tweet
            parsed_tweet['sentiment'] = self.get_tweet_sentiment(tweet.text)

            # appending parsed tweet to tweets list
            if tweet.retweet_count > 0:
                # if tweet has retweets, ensure that it is appended only once
                if parsed_tweet not in tweets:
                    tweets.append(parsed_tweet)
            else:
                tweets.append(parsed_tweet)

        # return parsed tweets
        return tweets

    except tweepy.TweepError as e:
        # print error (if any)
        print("Error : " + str(e))
```

Finally, parsed tweets are returned. Then, we can do various type of statistical analysis on the tweets. For example, in above program, we tried to find the percentage of positive, negative and neutral tweets about a query.

```python
def main():
    # creating object of TwitterClient Class
    api = TwitterClient()
    # calling function to get tweets
    searchTerm = input("Enter Keyword/Tag to search about: ")
    NoOfTerms = int(input("Enter how many tweets to search: "))
    tweets = api.get_tweets(query = searchTerm, count = NoOfTerms)

    # picking positive tweets from tweets
    ptweets = [tweet for tweet in tweets if tweet['sentiment'] == 'positive']
    # percentage of positive tweets
    print("Positive tweets percentage: {} %".format(100*len(ptweets)/len(tweets)))
    # picking negative tweets from tweets
    ntweets = [tweet for tweet in tweets if tweet['sentiment'] == 'negative']
    # percentage of negative tweets
    print("Negative tweets percentage: {} %".format(100*len(ntweets)/len(tweets)))
    # percentage of neutral tweets
    print("Neutral tweets percentage: {} % ".format(100*(len(tweets) -(len( ntweets )+len( ptweets)))/len(tweets)))

    # printing first 5 positive tweets
    print("\n\nPositive tweets:")
    for tweet in ptweets[:20]:
        print(tweet['text'])

    # printing first 5 negative tweets
    print("\n\nNegative tweets:")
    for tweet in ntweets[:20]:
        print(tweet['text'])

if __name__ == "__main__":
    # calling main function
    main()
```

```
Enter Keyword/Tag to search about: donald trump
Enter how many tweets to search: 30000
Positive tweets percentage: 30.64516129032258 %
Negative tweets percentage: 19.35483870967742 %
Neutral tweets percentage: 50.0 %


Positive tweets:
RT @mchooyah: Mystery British businessman bets $5million on Donald Trump winning the US presidential election in 'largest political wager e…
RT @DianeLong22: 🦋🦋🦋MUST SEE🦋🦋🦋🦋: "The Vietnamese Soul Choir" Release Wonderful New Video Singing Their Support for Donald Trump via @gatew…
RT @JoeBiden: Donald Trump is the most corrupt president in modern history.

Donald Trump is the most racist president in modern history.…
RT @KamalaHarris: Donald Trump wants you to deny reality. To pretend that, during a global pandemic, you don't see the bills piling up or t…
@didikins4life @leahkrevit Have to post this again!!!

Wishing that every republican getting ready to head to the p… https://t.co/SJtNxjUD7n (https://t.co/SJtNxjUD7n)
RT @LVIS_AGVIRRE: Donald Trump making Nov 1st (cultural holidy of Day of the Dead/Dia de Los Muertos for many latinos) a national remembran…
RT @PeteButtigieg: Like @JoeBiden, and most Americans, I trust women to make their own health care decisions.

Donald Trump doesn't. Do you?
RT @AndyPuzder: Bottom line: despite what you may hear in the media, it's OK to vote for Donald Trump — he's not only an extremely effectiv…
RT @NaomiBiden: "Are you ready to tell Donald Trump: You're fired?" - @drbiden https://t.co/rar1XExgQ0 (https://t.co/rar1XExgQ0)
RT @politico: New Jersey officials might press charges related to Sunday's caravan of President Donald Trump supporters who blocked traffic…
@ShadyPlanes "an adjustment". PA declares Joe Biden the winner. "We're asking for PA to adjust that to ... Donald Trump". LOL
RT @ProjectLincoln: Donald Trump, the American people are speaking directly to you with a simple message: It's time to start packing your b…
```

RT @MaryLTrump: "Traitor" by @djrothkopf is a brilliant, concise, and damning indictment of the current administration and the failures of…
RT @Tom___Scott: If Joe Biden wins, taco trucks will erase our saviour Donald J. Trump.

Can't come soon enough. https://t.co/AkGmG8jjyJ (https://t.co/AkGmG8jjyJ)
RT @BillKristol: Robert Gates balks at attempt by @WSJopinion to enlist his good name in rationalizing support for Donald Trump https://t.c… (https://t.c…)
RT @JoeBiden: When America votes, America will be heard. And when America is heard, I believe the message is going to be loud and clear: It…
RT @TVNewsHQ: Watch: Tucker Carlson explains why "millions of Americans sincerely love Donald Trump." https://t.co/n4GFrUP9pw (https://t.co/n4GFrUP9pw)
RT @PeterAlexander: Biden, now: "Last night, Trump said he was going to fire Dr. Fauci. Isn't that wonderful? I have a better idea... elect…
RT @RonaldKlain: I'm that guy. And if Donald Trump really cares about what I think, here you go: no President has bungled anything, ever,…

Negative tweets:
RT @charliekirk11: Donald Trump was up past midnight campaigning last night. What do you think Joe Biden was doing past midnight?

If he do…
RT @michaelianblack: MY OFFICIAL PRESIDENTIAL ENDORSEMENT THREAD!

As some of you know it's been VERY difficult for me to make up my mind…
RT @DailyCaller: TUCKER: "Donald Trump, in other words, is and has always been a living indictment of the people who run this country. That…
RT @ProjectLincoln: Reminder: late last night, Donald Trump said he'll fire Dr. Fauci after the election.

 https://t.co/48MlqaJdxA (https://t.co/48MlqaJdxA)
RT @Reuters: How the election is dividing American families: When lifelong Democrat Mayra Gomez told her 21-year-old son five months ago th…
RT @VanBalls: Fuck Donald Trump
RT @mmpadellan: To #GullibleWhiteMaleTrumpVoters:

Donald trump is not smart.
Donald trump is not a billionaire.
Donald trump is not good a…
RT @DailyCaller: Hong Kong Democracy Protesters Support Trump, View Biden As Weak https://t.co/8ZUp0I1e1K (https://t.co/8ZUp0I1e1K)
RT @NBCNews: Dr. Fauci says the country needs to make an "abrupt change" in its public health practices and behaviors as the holiday season…
RT @cacouteau: @MeidasTouch Donald Trump's record for Ohio sucks! https://t.co/09JC8k8uBM (https://t.co/09JC8k8uBM)
FUCK DONALD TRUMP AND ALL ITS SUPPORTERS. If I know you DELETE ME AND A BIG F U https://t.co/y6mfUjPjUB (https://t.co/y6mfUjPjUB)
RT @AugustTakala: TUCKER: Millions of Americans love Donald Trump. He doesn't hate them for their weakness. Trump is a living indictment of…

In [ ]: