

# TWITTER NLP

## Twitter Natural Language Processing Stock Forecaster Documentation

Psreid

September 2021

### Purpose

This software package is capable of querying twitter via [twitter API](#) and conducting sentiment analysis with the textblob NLP package. The purpose of this sentiment analysis is to acquire secondary information for a multivariate forecaster.

This forecaster will combine stock data and twitter sentiment to better predict the outcome of future stock prices (best performance one day ahead). The advantage of this method is to quickly derive a public consensus on the perceived value of a stock, before it is represented in the price within the market itself. This is particularly advantageous when dealing with volatile short term stocks that are heavily influenced by social media (Meme Stocks)

### Workflow

Prior to startup you must register for twitter app privileges via [twitter API](#). You will then be granted access tokens which can be placed within the forecaster base package

An example jupyter notebook is included on the git page [https://github.com/psreid/Twitter\\_NLP\\_Stock\\_Forecaster](https://github.com/psreid/Twitter_NLP_Stock_Forecaster). This package functions in 4 (5) steps

1. Initialise a forecaster object from forecaster base with a specific stock ticker
2. Query both twitter and Google stocks for a week long timespan of ticker
3. Generate twitter sentiment scoring metrics with textblob, and combine with stock data to create a single dataframe

4. (Optional) append dataframe to previous queries and stored data, store combined data data
5. perform multivariate step-wise predictions with `statsmodels` `vector autoregression` and plot results

Twitter NLP Forecaster has the potential to identify sharp trends within a social media space before substantial action is taken in the open market. For example, in figure 1 we see a sharp increase in twitter controversy just before the a significant drop in market price early the next day (figure 3).

The current functionality of Twitter NLP forecaster is limited to the query ability of the current twitter API, where in general, it is best to automate a twitter query each day, and subsequently store the data with Twitter NLP's built in functionality. The metrics used to score sentiment are "aggregate sentiment", which is tweet volume (query volume) dependant, and controversy score, which is tweet volume independent.

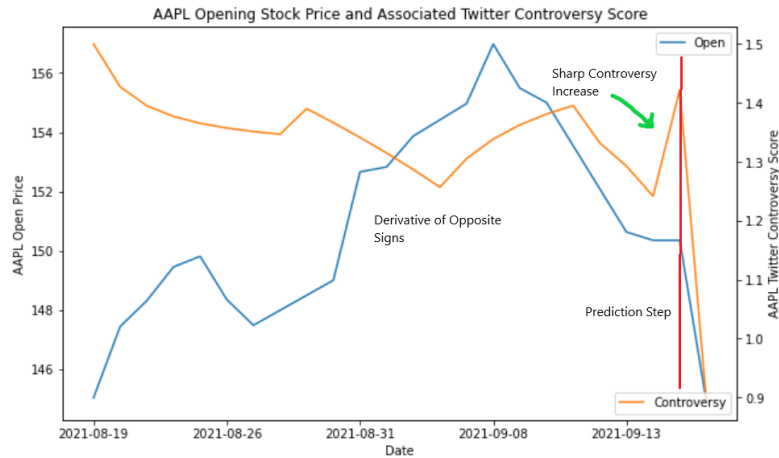


Figure 1: Apple opening stock price and associated twitter sentiment controversy score from 2021-09-19 to 2021-09-14 with a stock prediction at 2021-09-15

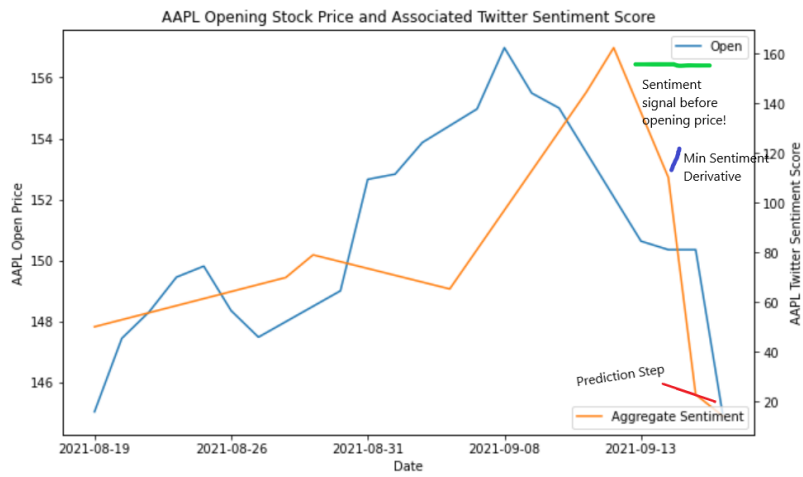


Figure 2: Apple opening stock price and aggregate twitter sentiment score from 2021-09-19 to 2021-09-14 with a stock prediction at 2021-09-15

#### Market Summary > Apple Inc

**149.03** USD

+0.91 (0.61%) ↑ today

Closed: Sep. 15, 7:59 p.m. EDT · Disclaimer

After hours 148.97 -0.060 (0.040%)

NASDAQ: AAPL

+ Follow

1D | 5D | 1M | 6M | YTD | 1Y | 5Y | Max



Open	148.56	Mkt cap	2.46T	CDP score	A
High	149.44	P/E ratio	29.19	52-wk high	157.26
Low	146.37	Div yield	0.59%	52-wk low	103.10

Figure 3: Stock price of Apple on the predicted 2021-09-15