Ax

Jan 23 · 6 min read · Listen

Apple Inc. from Stock Market and Twitter Posts Perspectives

Authors: Junting Diao, Tommy Li, Melody Xu

> ***Takeaways***
>
> *Topics related to tweets about AAPL from 2018 to 2019 involve : Tim Cook and China, Patent,Trading, Top giants, Oil battle and … Robinhood; Sentiment does affect AAPL stock price considering lags, although perhaps not account for most variation.*
>
> *Github: https://github.com/tommyli01/sentiment-anlasysis-with-AAPL-stock-price*

We are sure that you are already very familiar with Apple Inc. Tim Cook, Steve Jobs, iPhone, Mac……Therefore, we can save some time not repeating the backgrounds of the company. Let's cut to the chase!

Curious about the impact of public opinion on Apple's stock price (AAPL), we plan to use Twitter comments as our dataset for analysis to search for answers.

Data Preparation

- Date is transformed to standard format from seconds since epoch.
- Dataset is filtered for AAPL ticker only, filtered for the time period between 2018/1/1 to 2019/12/31.
- Tickers and tweet bodies are merged by date through Tableau Prep.

We will perform the following steps:

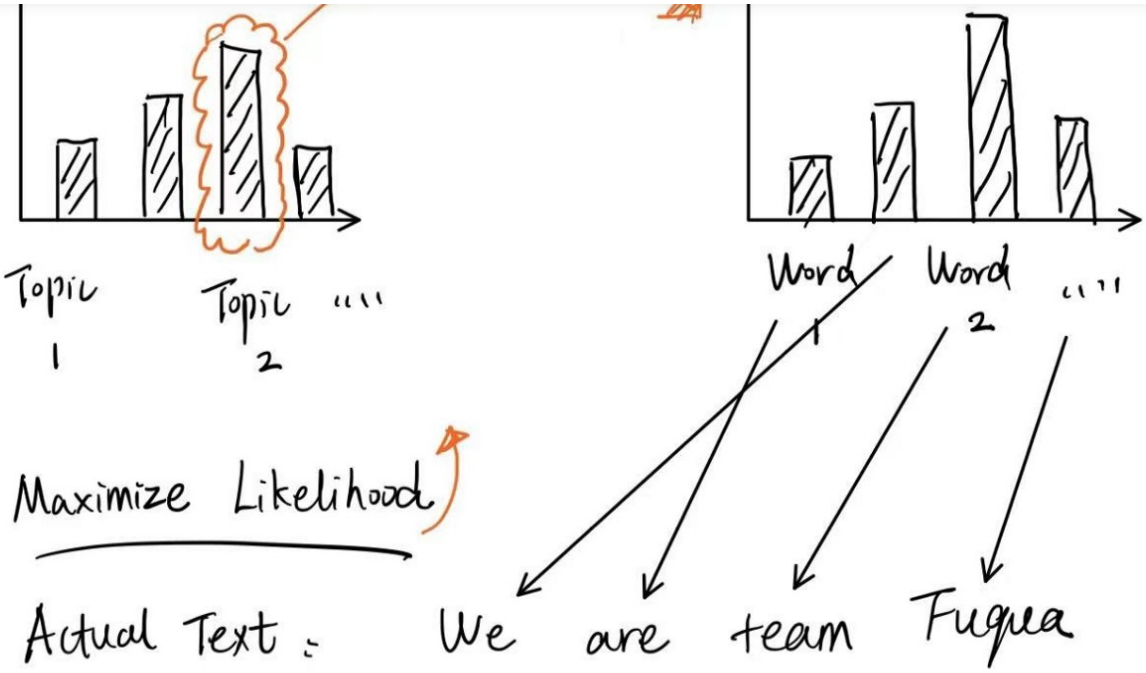
- Lowercase the words and remove punctuation.
- Tokenize the text and split the sentences into words.
- Remove stop words and digits.
- Lemmatize the words. Change the third person to first person, and past and future tense to present tense.

TF-IDF

After data preprocessing, we use Term Frequency–Inverse Document Frequency (TF-IDF) to create features and enable words to be expressed mathematically. Instead of word count, the value corresponding to the word is assigned a TF-IDF value, which increases as the number of times a word appears in the text increases, and decreases as the number of text that contains the word increases, so that common words, such as “is”, would not be assigned too much weight.

LDA

Next, we use Latent Dirichlet Allocation (LDA) to perform topic analysis. LDA assumes that each document is a distribution of topics, while each topic is a distribution of words. What we are going to do is to find the topic/word distribution that can maximize the likelihood that we see our actual texts. After applying LDA to all the tweets, we find 5 interesting topics related to AAPL.



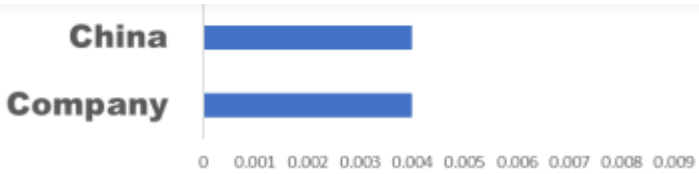
So,

> *What are people talking about when they are talking about AAPL?*

In general, we can see that AAPL has already been a great representative of the whole market. When people mention AAPL, they are talking about not only the operation of Apple itself but also macroeconomics, giants in the market, trading strategy, and….ads for Robinhood. Now, let's have a closer look at the topics.

Topic 1: Tim Cook and China

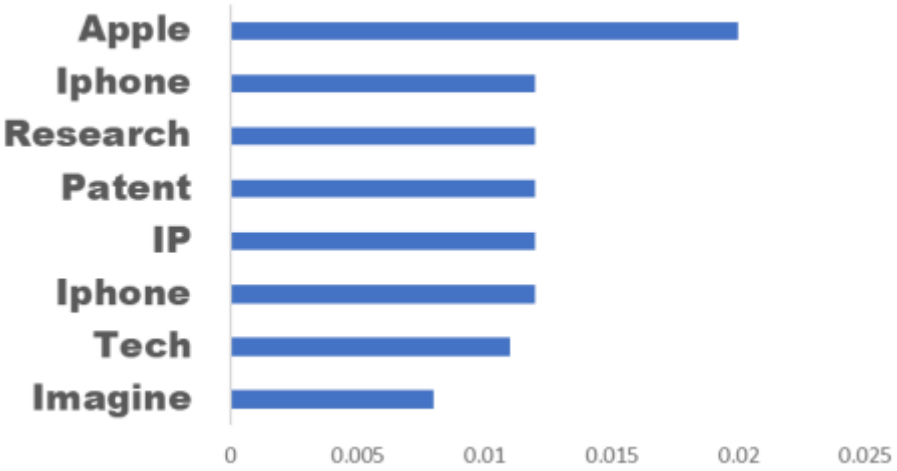
The high-frequency words in Topic 1 revolve around "Tim Cook", "iPhone" and "China". According to CNBC News, "Here's why Apple is so vulnerable to a trade war with China", the United States and China started a trade war in 2019. Donald Trump claimed to raise tariffs on Chinese imports. Apple was extremely vulnerable because China is responsible for most iPhone assembling and is Apple's biggest consumer market. It's not hard to imagine how much discussion there was about a big tariff hit on Apple.



Probability of Words

Topic 2: Patent

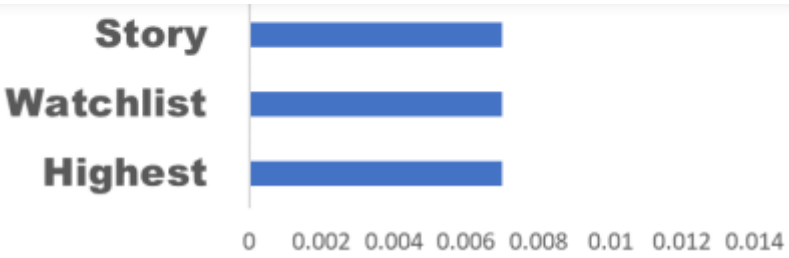
Topic 2 involves “Apple”, “iPhone”, “Patent” etc. The hottest topics in 2018 and 2019 related to patents are that Apple settled patent disputes with Samsung and Qualcomm. Besides, The U.S. Patent and Trademark Office granted Apple 52 newly granted patents in 2019.



Probability of Words

Topic 3: Trading

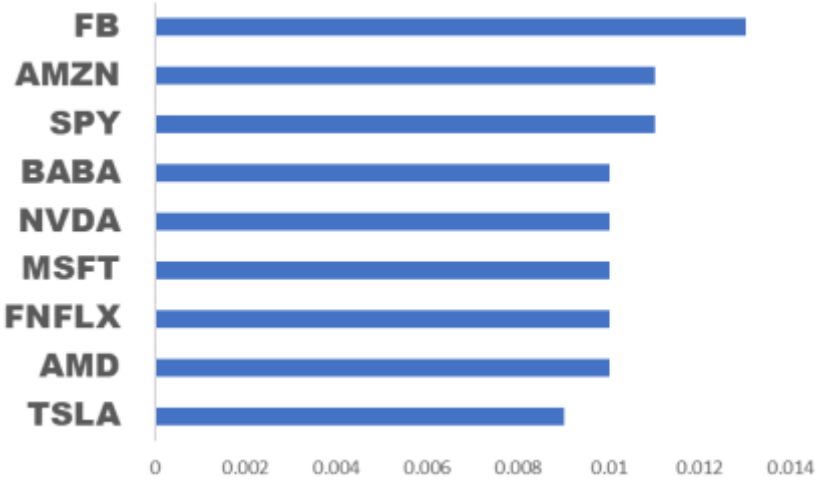
Apple is such a huge company that any shrewd trader would not miss it.



Probability of Words

Topic 4: Top company tickers/hot discussion

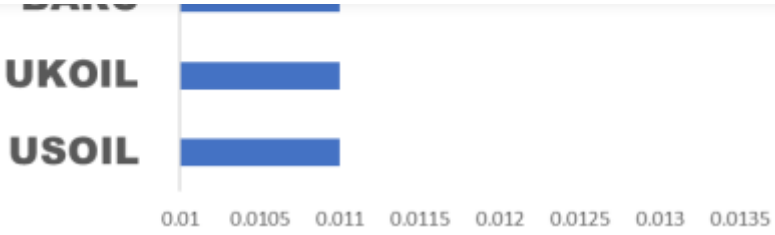
These are the big companies' tickers, which are often tagged and discussed on Twitter.



Probability of Words

Topic 5: Oil battle

Oil affects all things, Apple is no exception.



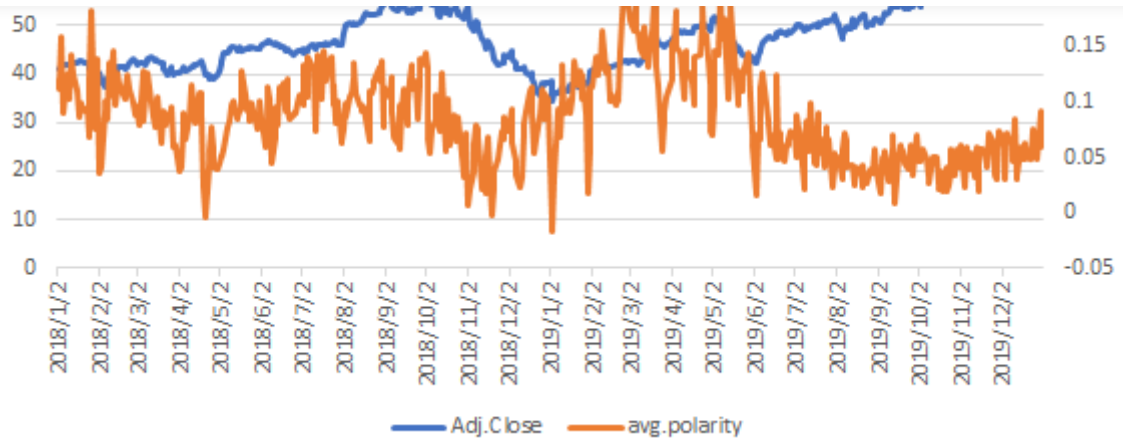
Probability of Words

We now know what main words the public talks about when linking to AAPL. Let's move on to a sentiment analysis over words.

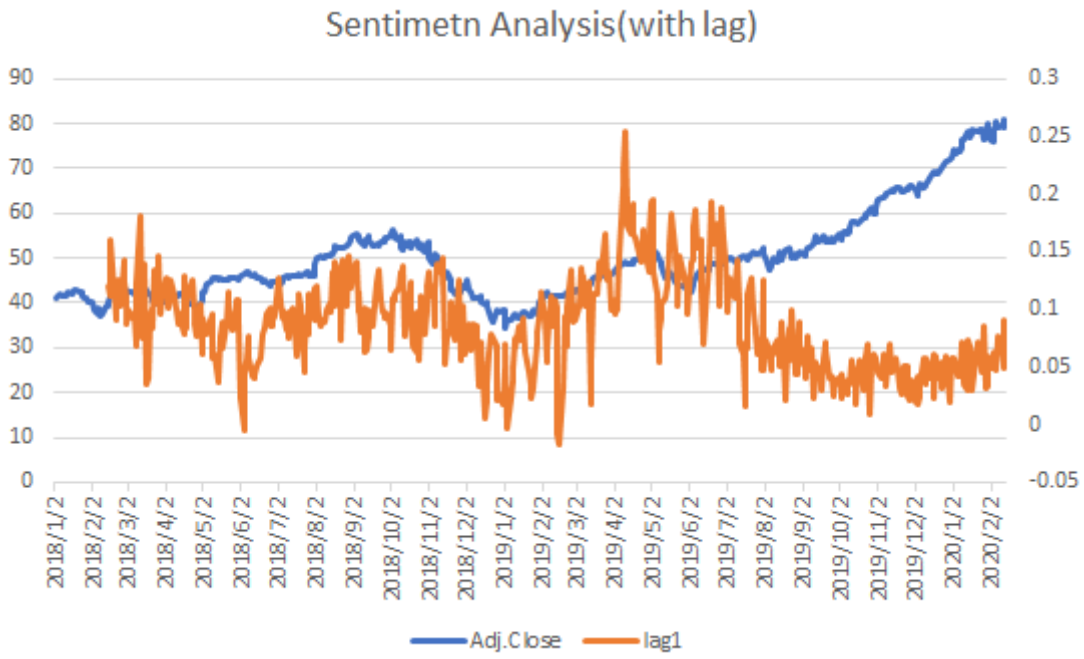
Is there a connection between market sentiment and the stock return of AAPL?

To analyze sentiments in the market, we decided to use polarity scores to represent the emotions.

Using tweets with ticker AAPL, we collected texts that were sent every day and grouped by their date. Scanning through these texts, we assigned polarity scores to words that can be identified as positive words or negative words from the same day and then summed up the value to be the polarity score of that specific day. This was conducted to data of each day from 2018 to 2019 and plotted against the market close price of AAPL.



As shown in the graph, although not a perfect match, there is a similarity between the two from the beginning of 2018 to mid-2019. It might be an indication of lag for the market sentiment to actually reflect on the stock prices. With that, we created a one-month lag variable and plotted it against the AAPL again.



With a one-month lag, there seems to be an overlay. This might imply that sentiment accounts for some variation of the AAPL stock price, though a fork at mid-2019.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   42.559      0.734   57.98  < 2e-16 ***
lag1          45.698      7.020    6.51 2.87e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.773 on 323 degrees of freedom
Multiple R-squared:  0.116,     Adjusted R-squared:  0.1132
F-statistic: 42.38 on 1 and 323 DF,  p-value: 2.873e-10

> |
```

As we can see from the result, the P-value looks very good. It turns out the lag polarity score has some relation with AAPL stock price. Sentiments exist in the market and have effects on AAPL stock prices, though this effect cannot account for a large portion of variations of the stock prices. To a certain degree, the stock prices vary with our emotions, and there is a time lag before this effect is reflected in stock markets.

Although we can use sentiments to build a prediction model of stock markets, factors influencing stocks are more than just sentiments. It could be influencer effect, politics, etc. Stock is fickle.

Reference

Dataset:

https://www.kaggle.com/omermetinn/tweets-about-the-top-companies-from-2015-to-2020

Topic Modeling and Latent Dirichlet Allocation (LDA) in Python | by Susan Li | Towards Data Science

https://www.cnet.com/tech/mobile/apple-and-samsung-finally-settle-their-patent-dispute/

https://www.cnbc.com/2019/04/16/apple-qualcomm-settle-royalty-dispute-sources-say.html