Aneesha Sreerama : sreerama.a@husky.neu.edu

Alexander Deris : deris.a@husky.neu.edu

DS2000

April 3, 2020

Reacting to the News in order to Forecast Company Performance

Problem Statement and Background:

The objective of this analysis is to identify a correlation between the content in news articles and the performance of the company in the month that followed. To observe the effect of the news, we decided to use Apple Company as the focus of our project. In order to accomplish this task, we utilized an API designed by Alpha Vantage to collect data from the New York Stock Exchange. Using an HTTP request, we compiled a wealth of financial information about Apple Company from, market highs & lows, monthly opening prices, and monthly closing prices. This API was created with the intent to analyze the stock market data for companies on the New York Stock Exchange. In addition, we used the New York Times API in order to search for articles that best fit our filters. This API provides us with metadata, or precise information regarding topics such as headlines, abstracts, images, dates, authors, etc. This application was created to analyze news regarding particular topics, events, people, and popular trends.

**Introduction and Description of the Data:**

My motivation for choosing this particular project was to have a better understanding of the intricacies of analyzing text to make real-world decisions. Algorithmically-driven trading over stock exchanges and information collection to forecast business growth/consumer trends are largely driven by data. The impact of media coverage on decisions made by businesses regarding

upcoming products to legal activities affect the performance and perception of companies. Quantifying the effect media coverage has through the use of news articles could provide more insight into what types of events have the largest impacts on the value of a company. This investigation is important as its prevalence grows in today's society. The enormous amount of data generated can be used to help draw conclusions about our responses to events that we could not trace before. Through the use of 20 years worth of data acquired from the above mentioned APIs, I analyzed how the content of articles written in the span of one month impacted the company's performance in the month that followed

Methods:

To collect information about the monthly change in price for the Apple Company stock I used the Alpha Vantage API to collect the 'Monthly Open' and 'Monthly Close' for each month. I converted the respective columns of the data frame into two lists and subtracted the 'Monthly Close' value from the 'Monthly Open' value. This gave me the change in the Apple Company stock for each month in the 20 year timeframe.
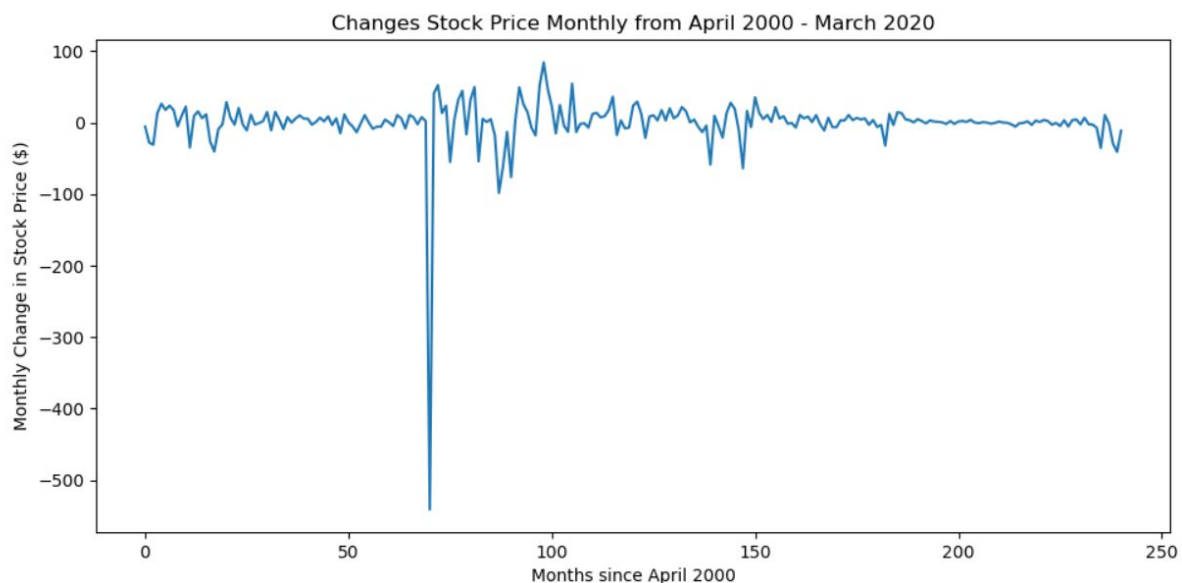
To analyze the articles I started with the headlines. Headlines are the first thing that we see and react to when we choose which article to read. They are the basis of determining which articles appeal to us and allude to the content that it unfolds. For this reason, I wanted to categorize the news articles about Apple based on the content they allude to. While looking through several articles, I noticed a general pattern that the articles seem to have. The articles centered on Apple tended to be of one of the following categories:

1. Discussions about Apple's latest products included:

    a. The company's new technology

    b. Product reveals

    c. The general atmosphere of how products were received

    d. New market spaces that Apple was venturing in

2. Apple vs Competitors included:

    a. Comparisons between Apple's performance to their main competitor: Microsoft

    b. Lawsuits and settlements that Apple was dealing with

    c. Markets that various tech companies were working to forge into their empires

3. Market Performance Articles included:

    a. Press releases

    b. Speculation on how their new products were factoring into their sales growth

    c. How recessions impacted major companies and stock exchanges alike

4. Undescriptive Headlines:

    a. Did not directly reference any of the prior categories

    b. Tended to be really short

c. Did not reveal significant information about what the article was about

5. Unrelated to Apple Company:

   a. These articles tended to be related to the "Big Apple" or New York due to the manner in which the New York Times API functioned. This API returned the Top 10 Articles that best matched the search filters that I implemented. Accordingly, when Apple tended to have less coverage in a month articles unrelated to Apple would be a part of the data. This category was created to filter articles that were unrelated to Apple.
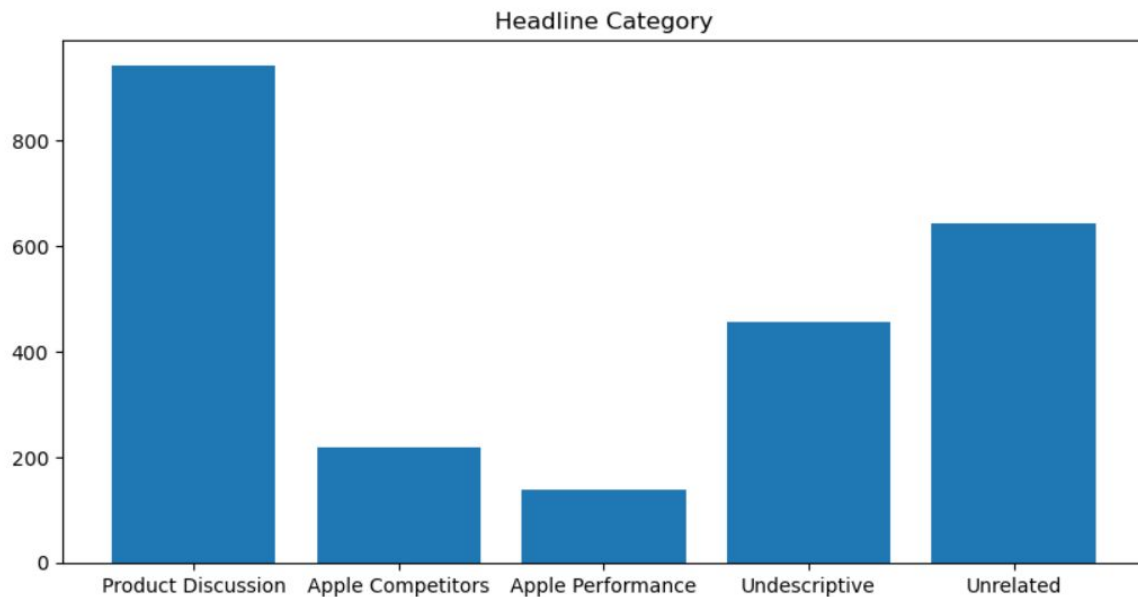
Using these categories, I wrote the categorize_headlines function that assigns a category to the headlines based on a dictionary of keywords I compiled based on reading the articles myself.

I proceeded to analyze the abstracts of articles as they summarize the most important aspects of the article in a concise manner. I designed a few functions for each of the categories mentioned above. I wrote a function to test four aspects of the article content.

1. Competitor Analysis:

   a. Searched for the word 'Microsoft' as this company was usually mentioned to compare Apple's products with those of Microsoft's

   b. The function searched for articles where Apple either filed or received a lawsuit

   c. The function searched for legal jargon in the sentence and the positioning of proper nouns within the sentence to identify the cases where Apple was negatively affected by a legal matter

2. Product Analysis:

   a. The function searched for articles that were centered around new Apple products

   b. Determined if their technology was facing problems or issues

   c. Determined if the articles purely discussed an Apple product

3. Performance Analysis:

   a. The function searched for words associated with income statements and the associated trends that are generally followed by it such as 'increasing' or 'bearish'

   b. Determined the general trend of the market

4. Unrelated:

   a. Verified if the article was irrelevant to Apple company

Using the data from the article and headline analysis, I assigned a score of -1 for every negative event and 1 for every positive event. I ran my function and received an accuracy of about 55%. I

started weighing the categories in effort to find numbers that tended to increase the total

accuracy of the final analysis. Through this tedious process, I found that certain categories of

headlines and articles tended to have a larger impact on the performance of the company as I was

able to increase my accuracy by about 7%.

Results, Conclusions, and Future Work:

The first question this project attempted to answer was what were the subjects of the

articles and headlines that had the largest impact on the value of the company. The headlines that

tended to center around the discussion of Apple products and Apple's general market

performance had a larger impact on the Apple stock. By increasing the weight of these two

headline categories, we were able to increase the accuracy of our predictions. Similarly, the

articles focusing on problems associated with Apple's technology and lawsuits and antitrust laws

had a large negative effect. Once again, by increasing the weight of these articles, we increased

the accuracy of our predictions.

The second question this project strived to answer was is there a correlation between the

content in news articles and the performance of the company in the month that followed and if

we could forecast the performance of Apple in the future. Based on our final results we found

that there was a weak correlation between the articles published and the performance of the

company. However, this answer is only based on the way that we approached this problem.

Our model provides us with an accuracy of 62.91%. Forecasting is definitely a possibility but

this model would require a lot of improvement. The improvements we can make can largely

build on the shortcomings of our project. We did not capture the details of the articles to the

extent necessary to have a higher accuracy. The analysis only identifies what the subject of the text was based on. I would do a deeper analysis on the text, break it down to more categories, and expand the dictionary of keywords. In addition, this project only attempts to determine if the company stock will rise or fall in the following month. This project does not capture the magnitude of the change. In a future project I would try to look for very specific events and create a more extensive system for determining how to assign both a magnitude and direction to these changes.