Jacob Tarnowski

4/21/2020

STAT 375

Dr. Ricca

Executive Summary

For this project, I decided to look at the effect the amount of web traffic surrounding a company affects their stock price. The idea for this study comes from the attempt to model consumer confidence. I believe consumer confidence is largely what dictates the stock market and is a function of information. The only way one can have an opinion on a company is if they have some sort of information on them. Whether it is an ad, a news article, or a video about the company, I believe these pieces of media are what determine consumer confidence. Now, as for why I am choosing to look at web traffic, it is quite obvious that there has been a shift of how people consume media from, newspapers and TV, to social media and the internet. Ideally, I would like to get Twitter data surrounding a company, but that proved to be more of a challenge than previously thought. That data would work the best because not only do we see that is generating web traffic, but it would also give a general idea of whether to web traffic is due to something “good” or “bad”. However, I had to settle for web traffic data, which is just information on how frequently a search term is used. I was able to obtain daily web traffic data from Google Trends. I was then able to get the stock market data required for this experiment from Yahoo Finance. I decided to use Apple’s stock for this experiment because they have a relatively large presence in the media. The next step in this project was to find the changepoints in the web traffic data. This was done by building a model based on this data and finding the points that were outside the 95% confidence interval of the model. I then took these points that were outside of the 95 percent confidence interval and placed them into groups based on what date they occurred and the web traffic value that was assigned for that specific day. By doing this I was able to get dates in which the trend of web traffic had distinctly changed. These dates where the trend changed are the dates that I am interested in, specifically the ones in which the trend increases. As I said earlier I believe that an increase in web traffic should contribute to an abnormal rise or fall in the stock market. After finding these changepoints I then looked at Apple’s stock price ten days after each of these changepoints and found the largest change in stock price within these ten days. I then labeled an abnormal change in the stock price as an increase or decrease of $4.50 (the average maximum change in stock price within ten days). With this information, I will be able to tell if the amount of web traffic a company is generating leads to a change in stock price.

Unfortunately, my experiment outcomes were rather poor. Only about five of the 20 plus change points I examined turned had changes of $4.50 or greater within ten days. However, I have some theories as to why my experiment did not turn out how I would have liked it two. One flaw I see in my experiment is that my web traffic data is not encapsulating all of the web traffic data surrounding a company. This is because the data is only showing me the popularity of the search term Apple. This does not include the search terms that are associated with Apple, such as the terms iPhone, Ipad, or Macbook. I choose to only use the term Apple because it was the most generic, but in hindsight, I could see how this could very well be the least used search term surrounding Apple. If a new Apple product comes out they aren’t going to google “Apple”, they are going to Google something more along the lines of “iPhone”. Given this fault is my data collection, if I were to go back and recollect this data I would like to make a general index of Apple’s web traffic that would be a combination of all search terms. By doing this I believe I would have a much better idea of what Apple’s web traffic is like. Another improvement that I could make to this experiment I touched on earlier, which is using Twitter data rather than data from Google Trends. In addition to the reasons, I provided earlier Twitter data would also help with the issue I was just discussing with “Apple” not being the only search term associated with the company. With Twitter data I would not have to make a combined index, all I would have to do is pull down tweets with all of the search terms associated with the company and store them. I would then look more for word frequency within these tweets to evaluate web traffic. One last tweak I would make to this experiment is that I would not use Apple stock, instead, I would use oil. I would make this adjustment because of the polarization of oil. Unlike Apple, oil companies are not in the media regularly. The only time oil companies have a large presence in the news is when something is happening with them. Unless something is expected to happen to the oil prices, you usually would not see an oil company name in a news headline. This would provide a greater distinction between changepoints in web traffic data. With Apple, it was difficult to decide when and where was the actual change point because of the perpetual media attention apple gets. I also see there being a stronger relationship with increasing web traffic and oil stock, again because when oil’s media attention rises, it generally means the event is fairly significant. An example of a significant event in the oil industry would be a failed deal with OPEC. Therefore, I believe that the changepoints in oil web traffic data will be a better predictor of abnormal changes in stock price than they were for Apple. Overall, my study did not turn out how I wanted it to, but I believe these improvements I suggest could make vast changes to the general outcome of my study.