PROJECT: SENTIMENT ANALYSIS FOR STOCKS

BY: TAI NGO

DATE: 11/19/2020

DSC-680 – BELLEVUE UNIVERSITY

Introduction

In the age of information technology, many resources have been poured constantly into various industries to push our society toward the next level. Among the prominent industries, technology has always been the main driving force. Technology has the ability to improve lives and productivity by minimizing waste and looking for better ways to manage resources. The consumers and the market will decide whether a new technology or an invention is good or bad. A successful technology or service has to serve the consumers in the most practical and affordable way. It has to bring values. Creating a small piece of equipment that is useful and affordable is like picking up an expensive gem underneath the ocean. The complexity and functionality of a device become irrelevant when its purpose is vague and confusing, and it will probably be rejected by the consumers. Due to the rise of many electronic devices, the internet and their widespread acceptance of the market, a new type of oil-like resource has been adopted into the market, which is called 'data'. The rise in the value of data may come from the rise of social media and the way they collect the data from their users. It is descriptive, massive and it has a very clear purpose; in most cases, it is used for targeted advertisement. In a very short amount of time, perhaps it is just a little bit more than ten years, the value of data, whether it is some basic personal information or habits of the users on the internet, becomes increasingly expensive. It is not a surprise to see data being compared to oil, which is still the primary source of the world's energy. However, the main difference is that oil is limited, but data is not. Oil takes millions of years to form and become useful, data takes milliseconds to be generated. Data is the new oil, and it is not just a figurative sense, but a literal sense. Many machines need real-time data feeding into it constantly to work properly. Once people start to figure out that data is just too important to pass on, they create new ways to mine, use and integrate it into their businesses. The number of ways that people can take advantage of data

increases over time, but there is an interesting topic that should get people excited and constantly check their data. This project will demonstrate a way to use real-time data, or daily news articles to be more specific, to make monetary decisions and rely on a technique of data science called 'sentiment analysis' to gain valuable insights from the stock market.

Sentiment analysis is a technique in data science that has been around for quite some time. Many people like to refer to it as some sort of applications that explore the public opinions or emotions toward a certain subject. As a data science student, when I learn about a specific subject and ask myself how do I apply a sentiment analysis on this subject, I realize that sentiment analysis is not just an ordinary data science project, it is an art of utilizing data. There are many applications that require data to function properly, and the stock market is one of them. It is not a surprise that the people who are very good at playing with the stock market become very wealthy, knowledgeable and intelligent. It is estimated that very successful singers or movie actors have assets worth around 100 million US dollars up to 1 billion US, but successful hedge fund managers or investors have far more than that. People who play the stock game well make very good money. Item-trading techniques have been created for hundreds of years to make predictions about the price movement of stocks. Back in the 18th century, a Japanese rice trader invented the Heikin Ashi, which is a candlestick charting technique; many traders are still using this technique nowadays. Different techniques allow people to look at stocks from different angles, which help them gain knowledge of what works and what does not, and the winning percentages of each technique. The question now becomes how does a sentiment analysis provide any valuable insights towards a stock price movement. Due to the way stocks are being traded and exchanged, algorithms play a big part. People have gradually figured out that the algorithms rely on news, especially news about political situations, pandemic, public sentiments and so on to determine the sell and the buy signal. Many people seem to firmly believe that news is an important feature in the stock-exchange algorithms, for which many investment firms and market makers are using. Understanding and categorizing the news in real-time allows retail investors to gain a serious advantage compared to their peers, who only rely on technical analysis and do not fully understand what is driving behind the market. The sentiment analysis helps the users shed light on what may be the next move of the algorithms. It may not tell you directly at what point you should enter a trade or buy a stock, but it gives a clear picture of what is happening on a particular day without the users having to scroll through every news articles. In addition, the sentiment analysis in this project provides valuable insights at the macro level. Investors and traders can form their own extrapolation with the information at hands and include this tool to their arsenal before making an important monetary decision. We may have a second thought

about purchasing more stocks when the sentiment analysis shows that the negative news is surging.

Summary of Business Problem/ Hypothesis

The stock market is arguably one of the most profitable ways to accumulate personal wealth by investing long-term or trading short-term. Forming a strategy or technique to consistently beat the average market return is the goal that many investors and traders are seeking. Many techniques have been created and refined for hundreds of years to obtain the highest winning percentage plays. However, there is no single technique that can reliably be used to beat the market. Experts in this field have been combining multiple tools and strategies including machine learning algorithms to help them make decisions. For most investors, they have been putting their monthly pay check into their 401K plans on stocks that they have no knowledge about or ETFs that reflect the overall market such as the Standard & Poor 500. While this is definitely a recommended choice for most people, it is far from being the best one. For those who already have some knowledge in the stock market and want to add a useful tool into their arsenal, this project can help those people solve this problem. Should I buy more stocks today or should I sell? How about tomorrow? Imagine if you know every news there is to know about the stock market or any particular stock on a given day, would that be a significant aid to help you make your decision? This project shows the solution to answer all of these questions. Applying the sentiment analysis on the stock market allows the investors to see the market at a macro level and makes wiser decisions. Playing with the timing of the market is like playing with fire, but having a sentiment analysis will make the timing much easier.

Questions

1. What kind of information does the dataset contain?
2. How is the dataset going to be accessed?
3. What are the relevant and important features that need to be extracted?
4. How is the dataset relevant to the project or help the business?
5. How will the data be cleaned?
6. What is the method for data visualization?
7. Which stocks are going to be investigated?

8. What are the libraries or modules in Python required to perform the analysis?
9. How does this project translate into benefits for businesses?
10. What kind of businesses could use the result of this project?
11. How do we expand this project further?
12. What are the downside of the projects?
13. Is this project scalable to more stocks?
14. How do we evaluate the result of the project?
15. How do we improve the result of the project?

Methods

The stock market is a very complex entity. There are thousands of stocks across various sectors being traded each second. In order to perform a real-time sentiment analysis on a particular stock, there needs to be a constantly updated dataset. For this reason, this project will not use a static file such as a csv file to build the application; instead, the sentiment analysis will be performed based on the news articles from Finviz.

To understand how the data in the project is utilized and why it is relevant to the business, we need to discuss about Finviz, which is a critical and essential tool for many professional and amateur investors. Finviz is a browser-based stock market research platform. Finviz stands for financial visualization. Investors can use Finviz to access to market information. Finviz offers free and premium options for investors to form ideas and find opportunities. Some features that Finviz offers are financial news, market maps, indicator tools, charts, portfolio tracking and screeners. Investors can easily see the snap shot of the overall market with map.

Figure 1. S&P500 Map by Finviz. Source https://finviz.com/map.ashx

In addition, users have more options to view the market across different sectors by using the bubble chart below.

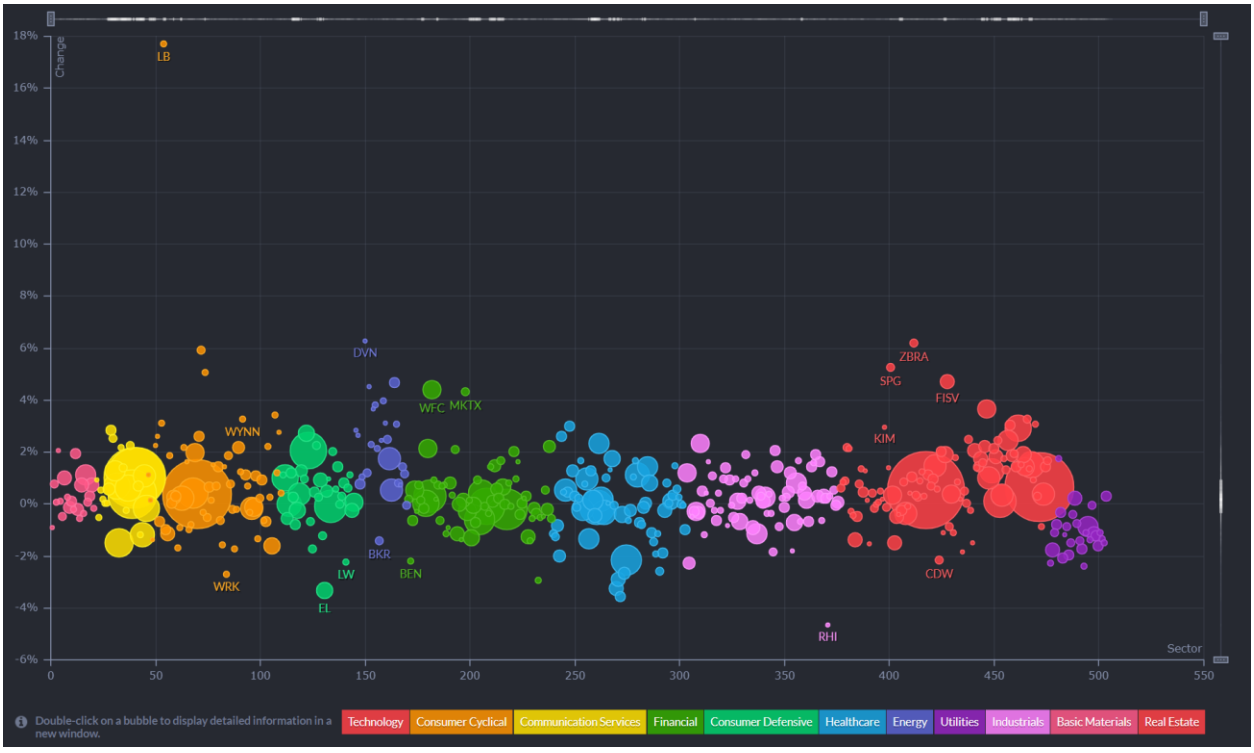

Figure 2. The S&P500 bubble chart by Finviz. Source: https://finviz.com/bubbles.ashx

The amount of free resources and information from Finviz is massive. It is very informative and helpful when it comes to market research. However, it is unreasonable to expect a normal person to digest all of this information on a daily basis. This is why this project will be very helpful to those who want to understand the sentiment analysis of the entire market within less than five minutes.

To perform a sentiment analysis on the stock market, we need relevant information. Finviz provides a number of articles that cover 'everything there is to know' about any particular stock. The financial articles are posted on a daily basis and continues to be updated on Finviz website. All we have to do is to scrap the data from the website. To achieve this task, we are going to use BeautifulSoup. BeautifulSoup allows us to gain access to the table of news articles. We can use Python to arrange and put the news into an empty dictionary. Once the news is collected, it can be parsed into an empty list, where all of the information will be filtered. After examining the data, four important columns are chosen: stock symbol, date, time and the content of the news article. In the next step, the sentiment analysis can be performed because the data is already cleaned and organized.

After obtaining the news from Finviz by using BeautifulSoup


```
{'QQQ': <table border="0" cellpadding="1" cellspacing="0" class="fullview-news-outer" id="news-table" width="100%">
 <tr><td align="right" style="white-space:nowrap" width="130">Nov-19-20 04:04PM  </td><td align="left"><div class="news-link-
container"><div class="news-link-left"><a class="tab-link-news" href="https://finance.yahoo.com/news/stock-market-news-live-u
pdates-november-19-2020-231410518.html" target="_blank">Stock market news live updates: Dow turns positive, shaking off earli
er declines as virus spread triggers more restrictions</a></div><div class="news-link-right"><span style="color:#aa6dc0;font-
size:9px"> Yahoo Finance</span></div></div></td></tr>
 <tr><td align="right" width="130">08:31AM  </td><td align="left"><div class="news-link-container"><div class="news-link-lef
t"><a class="tab-link-news" href="https://finance.yahoo.com/news/initial-jobless-claims-week-ended-nov-14-2020-coronavirus-pa
ndemic-194055455.html" target="_blank">Jobless claims: Another 742,000 Americans filed new unemployment claims last week</a>
</div><div class="news-link-right"><span style="color:#aa6dc0;font-size:9px"> Yahoo Finance</span></div></div></td></tr>
 <tr><td align="right" width="130">06:00AM  </td><td align="left"><div class="news-link-container"><div class="news-link-lef
t"><a class="tab-link-news" href="https://finance.yahoo.com/news/vaccine-developments-argue-for-more-fiscal-stimulus-morning-
brief-110038807.html" target="_blank">Why vaccine developments argue for more fiscal stimulus: Morning Brief</a></div><div cl
ass="news-link-right"><span style="color:#aa6dc0;font-size:9px"> Yahoo Finance</span></div></div></td></tr>
 <tr><td align="right" style="white-space:nowrap" width="130">Nov-18-20 05:31PM  </td><td align="left"><div class="news-link-
container"><div class="news-link-left"><a class="tab-link-news" href="https://finance.yahoo.com/video/dow-suffers-worst-loss-
3-223113415.html" target="_blank">Dow suffers worst loss in 3 weeks  YF Premium is bullish on Home Depot (HD)</a></div><div c
lass="news-link-right"><span style="color:#aa6dc0;font-size:9px"> Yahoo Finance Video</span></div></div></td></tr>
 <tr><td align="right" width="130">02:33PM  </td><td align="left"><div class="news-link-container"><div class="news-link-lef
```


After parsing the data, it looks easier to read


```
[['SPY', 'Nov-16-20', '05:48PM\xa0\xa0', 'U.S. Markets Rally on Positive Vaccine News From Moderna'], ['SPY', 'Nov-16-20', '05:
41PM\xa0\xa0', 'Tesla is getting added to the S&P 500, shares surge on the news'], ['SPY', 'Nov-16-20', '09:22AM\xa0\xa0', 'Inv
estor Confidence Builds With Lockdown-Lite Approaches'], ['SPY', 'Nov-16-20', '05:59AM\xa0\xa0', 'How the stock market schooled
everyone in 2020: Morning Brief'], ['SPY', 'Nov-15-20', '11:38AM\xa0\xa0', 'Target and Walmart earnings, Retail sales: What to
know in the week ahead'], ['SPY', 'Nov-13-20', '03:45PM\xa0\xa0', 'Impacts Of A Biden Tax Plan'], ['SPY', 'Nov-13-20', '11:59AM
\xa0\xa0', 'How Airbnb Makes Money'], ['SPY', 'Nov-13-20', '11:36AM\xa0\xa0', 'Here are 20 stocks that may benefit as investors
rotate money into cyclical sectors'], ['SPY', 'Nov-13-20', '06:01AM\xa0\xa0', '5 years of market action in one year: Morning Br
ief'], ['SPY', 'Nov-12-20', '06:21PM\xa0\xa0', 'Stocks Tumble as Virus Surges to New Record Daily High']]
```


Sentiment analysis is the most import part of this project; however, every other part equally contributes to the success of the project. There are many ways that we can perform the

sentiment analysis on the financial news. For this project, we are going to use Vader NLTK. Sentiment analysis is the process of using computational power to determine whether a piece of writing is positive, negative or neutral. The second name for the sentiment analysis is called 'opinion mining'. For this project, we want to know what is the public opinion toward a certain stock. The simple logic is that if everyone is bearish, or very bearish about a particular stock, then it is 'likely' to be sold more than it is bought on a certain day. We have to remember that this kind of analysis has probability, and having a low probability does not mean it will not happen. It is very possible that when most people think the stock is going to go down, it goes up. To explain why Vader NLTK will be a good choice for this project, we need to understand what it is and what it does. VADER stands for Valence Aware Dictionary and sEntiment Reasoner. It is a lexicon equipped with a rule-based sentiment analysis tool. It contains words that are labelled according to their semantic orientation as positive, negative or neutral. An advantage of Vader is that not only it can tell if a word is positive or negative, it has a scale of positivity and negativity of each word. Another advantage of Vader NLTK is its ease to use in Python. We only need to load the nltk module and use the Vader function. Once we have decided a pretrained tool for our sentiment analysis, we can perform some tests to see whether this tool is good enough. Many tests are done on different sentences and phrases, and the results will be shown below. The Vader shows an accurate result from the sentences.

```
# perform a negative test by using vader
print(vader.polarity_scores("The infection rate is increasing. The economy will crash"))
```

```
{'neg': 0.252, 'neu': 0.748, 'pos': 0.0, 'compound': -0.4019}
```

```
# perform a positive test by using vader
print(vader.polarity_scores("The earning reports are great. The stocks will rise"))
```

```
{'neg': 0.0, 'neu': 0.661, 'pos': 0.339, 'compound': 0.6249}
```

The next step of the process is to perform data visualization. It is safe to assume that the users are busy and their time is precious. The tool needs to be able to send valuable insights in a way that the users do not need too much time to digest. Data visualization is an excellent way to achieve this task. Each stock or multiple stocks can be displayed on a bar chart. The bar chart will show the information of how positive or negative the public sentiment is about a stock. The users have the option to put any stock into the script and run it. Below is an example of how the result comes out.

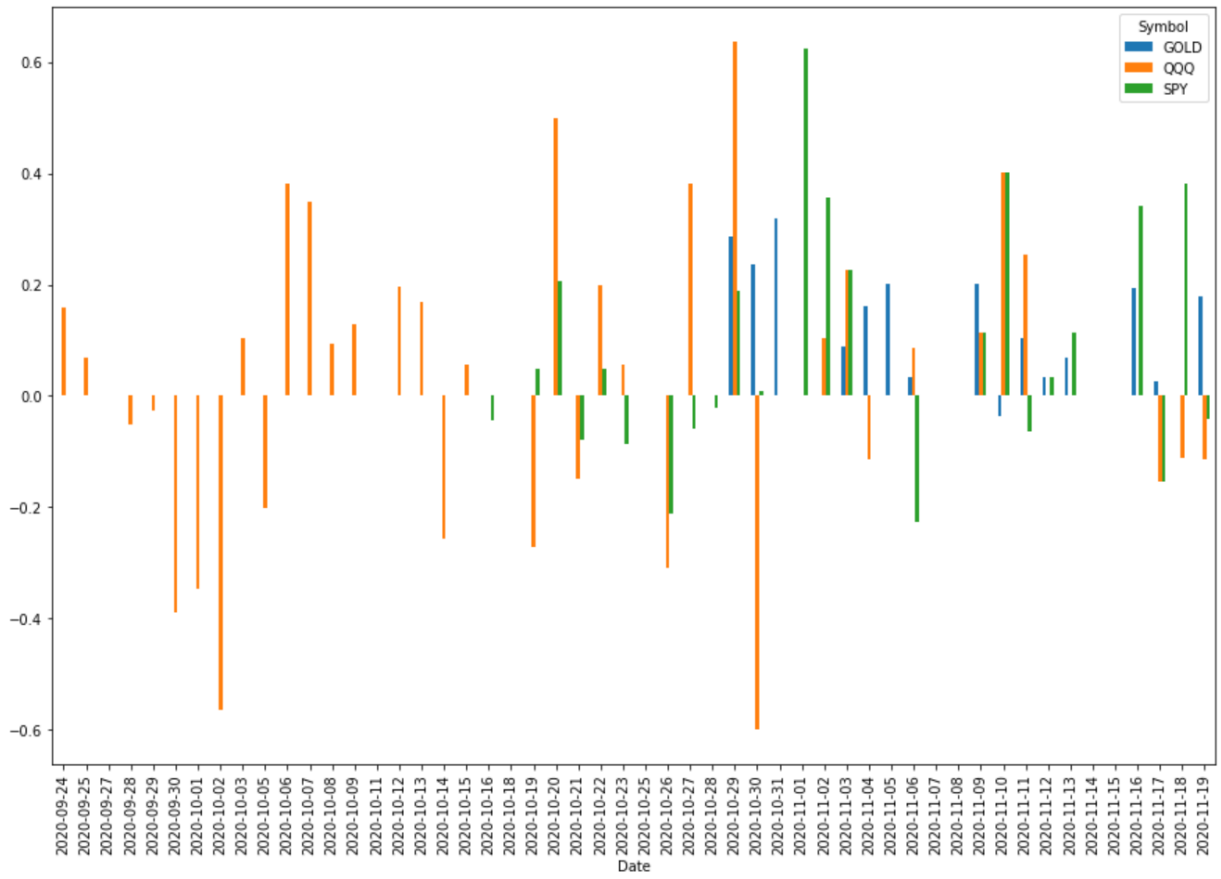

Figure 3. The sentiment analysis for SPY, QQQ and GOLD.

Figure 3 shows the sentiment analysis for three ETFs, ETFs are selected stocks from the same sector combined into one ticker. From Figure 3, we can quickly see how the sentiment for each stock changes over time and how it looks like on the last few days. Some people may refer to see just one stock. The following describes a quick analysis on SPY (S&P500 ETF) over the election week. We will see how the sentiment analysis can help investor predict the price movement of this stock.

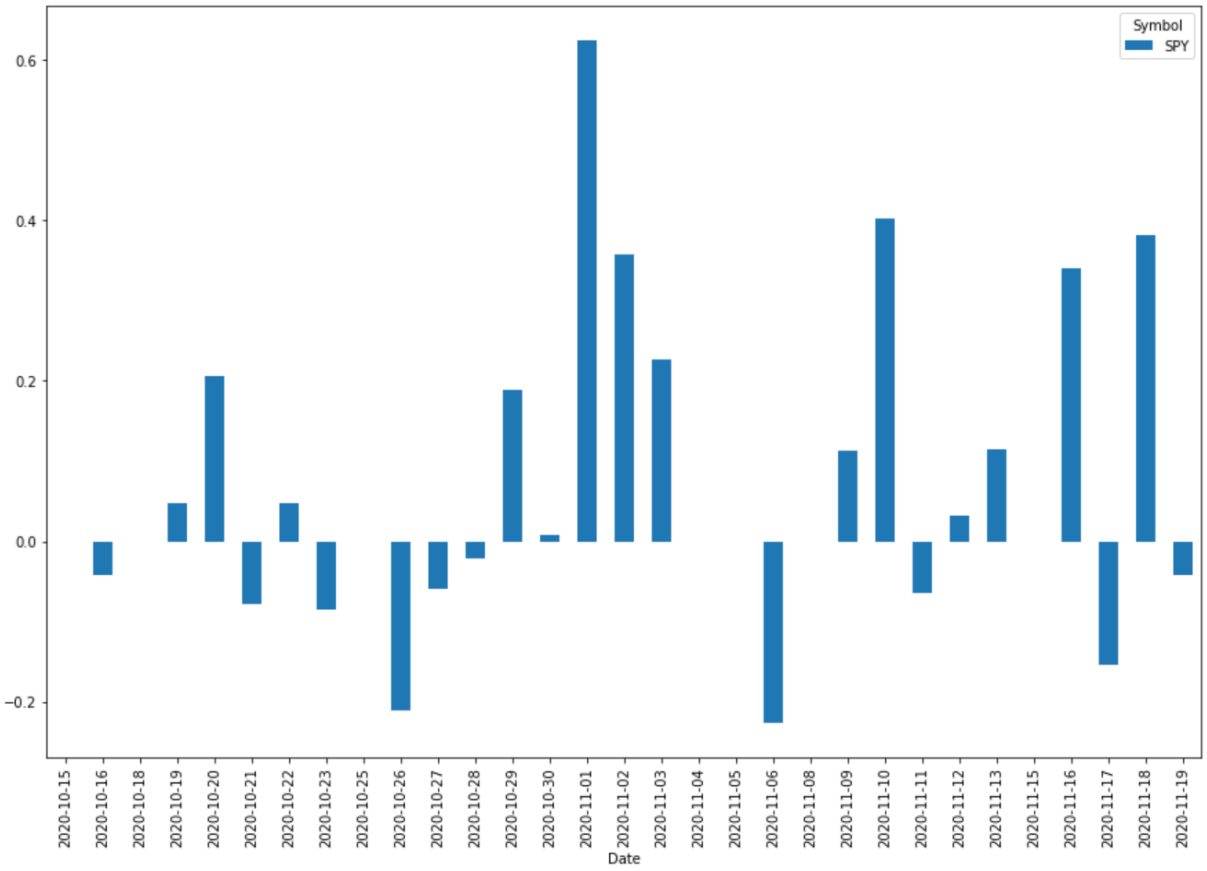

Figure 4. The sentiment analysis for SPY.

Figure 4 shows the sentiment analysis for SPY from the middle of October until the middle of November. The election week occurs during this period. As the stock market prepares for the election, experts expect a lot of volatility from November first to November 6. Historic data shows that stocks tend to dip before the election and starts to go back up after during the second or third week of November.

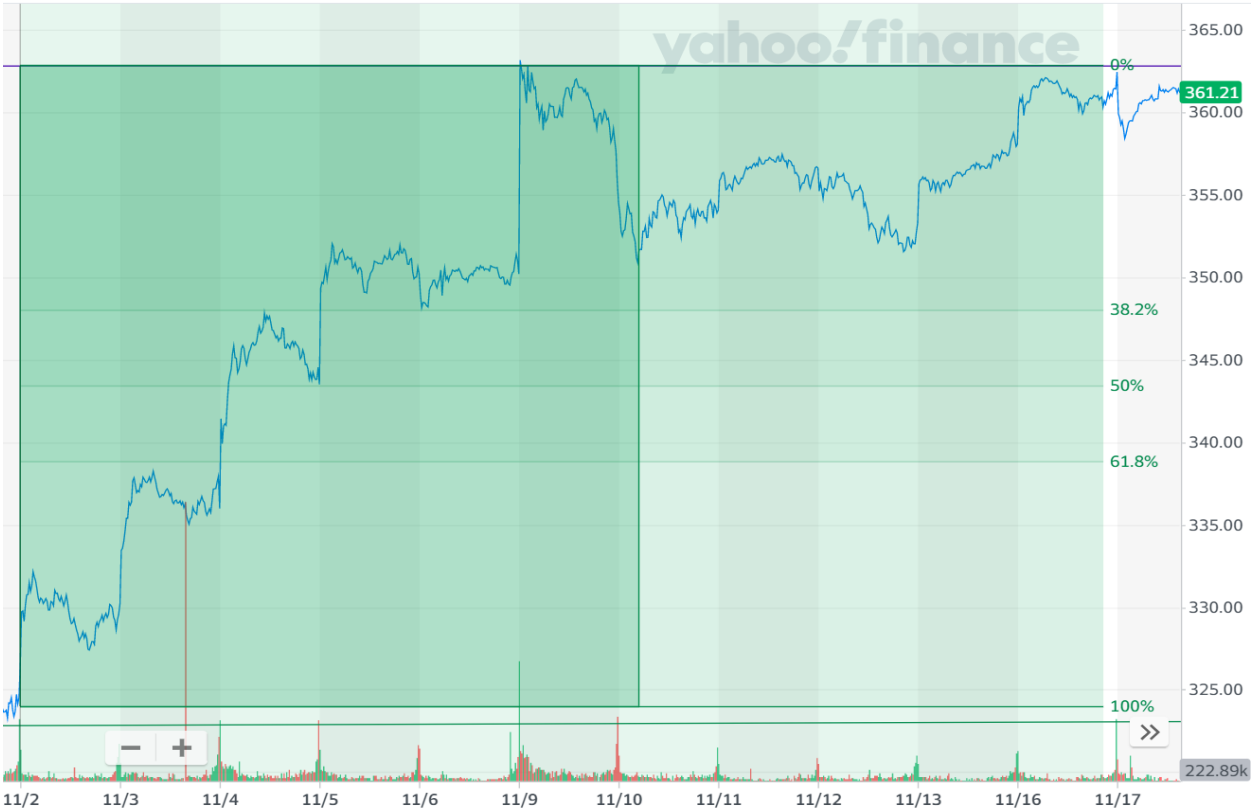

Figure 5. The price movement of SPY during the election week.

The stock went up significantly even before and during the election. By Monday of the next week, it had gone from 323 all the way to 363. This meteoric rise is not what a normal investor would expect. So how would our sentiment analysis help us make such an unreasonable prediction? We can examine from Figure 4 that the sentiment analysis for SPY during November first, second and third is a highly positive. It went down on November 6, which was Friday, as a result the next Monday peak did not last very long until it collapsed. Once again, SPY went up significantly on the 16th and down on the 17th, which also matches with the sentiment analysis.

The sentiment analysis can bring more useful insights to the table. It can aid investors in understanding the market by analyzing and rating the financial news and other relevant information from Finviz. The purpose of this is to be used as an additional tool even though it is fairly accurate on its own. We would not want to throw our money into the stock market when the sentiment analysis shows a high score for negativity. Using this sentiment analysis will open up more trading or investing strategies. The script takes only a few seconds to generate while the result is informative, quick and easy to understand.

References

1. Bakshi, R., Kaur, N., Kaur, R. March 17, 2016. IEEE. *Opinion Mining and Sentiment Analysis.* Retrieved from https://ieeexplore.ieee.org/abstract/document/7724305
2. Liu, B. May 2012. University of Illinois. *Sentiment Analysis and Opinion Mining.* Retrieved from https://www.morganclaypool.com/doi/abs/10.2200/s00416ed1v01y201204hlt016
3. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R. June 23, 2011. Columbia University. *Sentiment Analysis of Twitter Data.* Retrieved from https://www.aclweb.org/anthology/W11-0705.pdf
4. Maas, A., Daly, R., Pham, P., Huang D., Ng, A., Potts, C. June 19, 2011. Stanford University. *Learning Word Vectors for Sentiment Analysis.* Retrieved from https://www.aclweb.org/anthology/P11-1015.pdf
5. Lin, C., He, Y. Association for Computing Machinery. *Joint Sentiment/Topic Model for Sentiment Analysis.* Retrieved from https://dl.acm.org/doi/abs/10.1145/1645953.1646003
6. Pak, A., Paroubek, P. University of Paris. *Twitter as a Corpus for Sentiment Analysis and Opinion Mining.* Retrieved from https://lexitron.nectec.or.th/public/LREC-2010_Malta/pdf/385_Paper.pdf
7. Wilson, T., Wiebe, J., Hoffmann, P. University of Pittsburgh. *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis.* Retrieved from https://www.aclweb.org/anthology/H05-1044.pdf
8. Nasukawa, T., Yi, J. October 2003. Association for Computing Machinery. *Sentiment Analysis: capturing favorability using natural language processing.* Retrieved from https://dl.acm.org/doi/abs/10.1145/945645.945658
9. Medhat, W., Hassan, A., Korashy, H. December 2014. Ain Shams Engineering. *Sentiment analysis algorithms and applications: A survey.* Retrieved from https://www.sciencedirect.com/science/article/pii/S2090447914000550
10. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M. May 26, 2011. MIT Press. *Lexicon-Based Methods for Sentiment Analysis.* Retrieved from https://www.mitpressjournals.org/doi/abs/10.1162/COLI_a_00049
11. Ho, A., Menezes, I., Tagmouti, Y. University of Montreal. *E-MRS: Emotion-based Movie Recommender System.* Retrieved from http://www.ptidej.net/courses/ift6251/fall06/article/Projet%20Ai%20-%20Ilusca%20-%20Yousra.doc.pdf
12. Banerjee, S. *Collaborative Filtering for Movie Recommendations.* Retrieved from https://keras.io/examples/structured_data/collaborative_filtering_movielens/
13. CFI. *FINVIZ: Stock and Market Research Platform*. Retrieved from https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/finviz/

14. GeeksforGeeks. *Python – Sentiment Analysis using VADER*. Retrieved from https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/

Appendix