Stock Returns
An Analysis of Case Studies and Twitter Analysis to Predict Stock Prices

CS 4624: Multimedia, Hypertext, and Information Access
Final Report

Virginia Polytechnic Institute and State University
Blacksburg, Virginia, 24061
May 6, 2020

Instructor: Dr. Edward A. Fox
Client: Ms. Ziqian Song

Group Members: David Heck, Charlie Nguyen, Sayan Ray, Simon Shi, and Yannik Sood
# Table of Contents

Table of Figures ........................................ 2  
Table of Tables ......................................... 3  
Executive Summary .................................... 4  
Introduction ........................................... 5  
Requirements ......................................... 5  
Roles ..................................................... 6  
Design ................................................... 7  
Implementation ....................................... 9  
Testing and Evaluation ............................... 13  
User Manual ............................................ 18  
Developer’s Manual .................................... 21  
Timeline ................................................ 29  
Problems ............................................... 29  
Solutions ............................................... 30  
Future work ........................................... 30  
Acknowledgements .................................... 30  
References ............................................. 30  
Appendix ............................................... 31
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Example stock Twitter feed from Gilead</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>Gilead stock chart</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>APIs in our Python program</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Importing CSV files from Google Drive</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>An example of a list of keywords with stock ticker</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>CAR values for stock categories</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Salient terms</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>Keywords from social media</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>Snippet of code for CAR values</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>Code snippet for event sentiment score</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>Code for stable values</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>Imports from code</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>Punctuation in file</td>
<td>23</td>
</tr>
<tr>
<td>14</td>
<td>More example code</td>
<td>24</td>
</tr>
<tr>
<td>15</td>
<td>Our design for project</td>
<td>25</td>
</tr>
<tr>
<td>16</td>
<td>Project methodology</td>
<td>26</td>
</tr>
<tr>
<td>17</td>
<td>Project methodology in depth</td>
<td>27</td>
</tr>
<tr>
<td>18</td>
<td>Frequent words in case studies, positive</td>
<td>30</td>
</tr>
<tr>
<td>19</td>
<td>Frequent words in case studies, negative</td>
<td>31</td>
</tr>
<tr>
<td>20</td>
<td>Words correlation with CAR values</td>
<td>31</td>
</tr>
<tr>
<td>21</td>
<td>YUM stock, negative</td>
<td>33</td>
</tr>
<tr>
<td>22</td>
<td>HPQ stock, negative</td>
<td>34</td>
</tr>
<tr>
<td>23</td>
<td>GM stock, positive</td>
<td>34</td>
</tr>
<tr>
<td>24</td>
<td>FOSL stock, positive</td>
<td>35</td>
</tr>
<tr>
<td>25</td>
<td>All Negative Sentiment</td>
<td>36</td>
</tr>
<tr>
<td>26</td>
<td>All Positive Sentiment</td>
<td>37</td>
</tr>
</tbody>
</table>
Table of Tables

Table 1: Analyst Ratings Data from given data 14
Table 2: Stock prices subcategory data 14
Table 3: Stock picks subcategory data 15
Table 4: Salient terms 17
Table 5: Salient phrases 18
Executive Summary

The Stock Returns Project was to assist the research of our client, Ziqian Song, to analyze the language used in the financial news and social media discussions surrounding stock-related events, and to derive meaningful insights from this data.

Our end goal was to build meaningful tools that can help the client analyze information surrounding the events that may, in the future, predict a stock price move.

The project involved collecting, preprocessing, and analyzing textual data as well as working with stock data from the Wharton Research Data Services (WRDS). Data collection included 49 main categories and 335 subcategories of corporate events with 4.6 million related news and press releases. Following data collection, one of the tasks included identifying 20 influential events for case studies. These are events that have many news reports and tweets surrounding them. We picked 10 stocks that saw significant increases in stock price (surge stocks) and 10 that saw significant decreases in stock price (plunge stocks) to run our data analysis on.

Once we selected our 20 companies to evaluate, we used Python modules to scrape Twitter and Google data relating to each company and their specific case study event. By collecting different predictive features (e.g., emotions, top words, topics), we could find valuable correlations between the events and their discussion online. In our findings, we identified words that appeared the most for both surge and plunge stocks, sentiment on Twitter surrounding each respective event, and larger market trends that surrounded each event.
Introduction

Today, there are trillions of dollars put into various markets around the world [1]. As technology evolves, there are a large portion of stock trades that are automatically executed by computers, which mainly operate on technical analysis.

Software that can accurately predict stock movements based on major news events would revolutionize how stocks are traded. Events could be positive, like a major acquisition that caused stock prices to jump, or negative, like poor earnings reports, causing stock prices to fall. Predicting the weight of each event and the potential move has huge application in the real world.

Our project was aimed at analyzing data from different case studies, with each study surrounding a news event that caused a major surge (stock price increase) or plunge (stock price decrease). Specifically, we were looking for social media discussions and news articles, analyzed this data, and compared it with the stock prices. Our group worked on 20 case studies with thousands of data values on the companies’ stock price, news events, and tweets, yielding encouraging results.
Requirements

After many meetings within our group members, our professor, and our client, we identified the following tasks to accomplish:

1) Identify 20 influential events by looking into news and Twitter discussions. There are 10 surge events, and 10 plunge events.

2) Identify patterns by retrieving old tweets and articles surrounding the date of each surge or plunge event, and looking for keywords, phrases, and sentiment.

3) Thoroughly document the methods we used, analyze our findings, and include a literature review.
Roles

To help with completion of this project, we adopted different roles. We have identified our strengths and weaknesses amongst the group and found the appropriate fit for each member.

David Heck was in charge of the case study research. He has a background in data analytics and Swift programming. Yannik Sood generated features like top words and topic generation, since he has previous experiences in Python, web scraping, and stock trading. Sayan Ray helped with the topic generation and would categorize stock events. He has done work in the data analytics, Python, and web scraping fields. Simon Shi worked on sentiment analysis and prepared the final report to submit it to VTechWorks. He would also help with extracting data since he has experience in data analytics. Charlie Nguyen would generate summaries and statistics about the data and inspect data that may be useful for stock prediction. He has data analytics and Python experience.

These roles helped us assist Ms. Song in her stock research.
Design

Our final deliverable to Ms. Song was an analysis of the findings about the big data that we received. While there were many approaches to collect data and analyze it, we used Python to extract the files using Jupyter Notebook. We imported multiple APIs including the Pandas library.

For collecting the data, we used TweetScrape from an API to acquire tweets and also filter out the keywords [2].

We used TextRank, an algorithm that is based off of PageRank, to choose keywords from different news posts from the company [3].

For the Latent Dirichlet Allocation (LDA), we will use the Natural Language Toolkit (NLTK) API, pyLDAvis, and genism. LDA is a generative statistical model that will allow us to observe the sentiment of a particular stock to determine what happened when the stock surged or plunged.

We also analyzed sentiment on all the tweets we collected, and visualized our findings.

The following shows an overall design of what our project will look like. Figure 1 shows an example of a stock. In this example, the hashtag GILD we searched for was for the Gilead company. In Figure 1, there was a lot of news around the coronavirus that we found. Gilead’s stock ticker is GILD.
First, we would extract keywords from Twitter using a tweet scraper [2] to analyze the stock’s sentiment before and after a stock’s surge or plunge.

Figure 2 shows Gilead’s stock chart. The chart has green and red candlesticks, which are a tool to represent movements of a particular stock. A green candlestick means that the stock went up for the day and a red candle means the stock went down for the day. If there was a major up movement or major down movement, we would analyze that area for both the stock prices and the Twitter feeds.
Then, we would be able to draw correlations of the stock price changes as well as the social media feeds we were able to pull from using the Tweetscrape API [2]. This way, our goals would be accomplished if we were able to successfully predict the fall or rise of stock prices. This was our general process for this project.
Implementation

The data we received from our client was split into categories. For example, there were price targets, analyst ratings, stock picks, and taxes categories. In each of the categories, there are subcategories within those categories that go in depth into the specifics of the companies’ category and what happened with their stock prices.

First, we extracted basic data using Java by inputting the CSV files of a select number of categories to get an idea of what the data was like and how many positive, neutral, and negative stock prices were. We decided to pick the following categories: analyst ratings, stock prices, and stock picks. Using these categories would give us the general sentiment from this data.

We also coded a Python program, using the pandas, numpy, and matplotlib APIs to help with pulling specific parts from the files. In Figure 3, you can see the top of our program that has 3 imports. We ran the program on Google Colab. This would be explained more in our developer’s tool section.

```python
import pandas
import numpy
import matplotlib
%matplotlib inline
frame = pandas.read_csv('/content/drive/Shared drives/CS4624 Stock/all_event_CSS_MCQ.csv')
frame = frame.drop('bins', axis = 1)
frame['bins'] = (frame.CAR>0).map(int)
frame
```

*Figure 3: APIs in our Python program*

In order to pull out the folders and files from the Google Drive link our client has sent us, we first uploaded our program to Google Colab. Then, we mounted our drive folder onto it with this line of code. In Figure 4, we mounted our Google Drive files in Google Colab.
Figure 4: Importing CSV files from Google Drive

Figure 5 is an example of a section where we extracted keywords from different stocks from social media. The key columns were the keywords_scores, keywords, and ticker_symbol. This data would allow us to get a score for each of the keywords for each stock [3].

![Table of keywords with stock ticker](image)

*Figure 5: An example of a list of keywords with stock ticker*
Figure 6 shows the CAR values we would use to determine whether a stock would be in the: category surge if the value is above 2.5%, plunge if the value is below -2.5%, and stable if the value is between -2.5% and 2.5%.

After collecting tweets from stock companies as in Figure 5, we were able to show the tweets and find the keywords.
Testing and Evaluation

A basic Java program was used to input the CSV file and count the number of ratings. There are 3 results of the ratings: positive, negative, and neutral ratings. Table 1 explains the ratings for the 3 subcategories in the ratings category: analyst ratings change, analyst ratings history, and analyst ratings set.

<table>
<thead>
<tr>
<th></th>
<th>Analyst ratings change</th>
<th>Analyst ratings history</th>
<th>Analyst ratings set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive ratings</td>
<td>99,766</td>
<td>30</td>
<td>35,068</td>
</tr>
<tr>
<td>Negative ratings</td>
<td>108,662</td>
<td>3</td>
<td>3,192</td>
</tr>
<tr>
<td>Neutral ratings</td>
<td>70,257</td>
<td>18</td>
<td>26,675</td>
</tr>
</tbody>
</table>

*Table 1: Analyst ratings from data*

We counted the number of stocks that had a gain or a loss in the stock prices subcategory and counted the number of stocks that had a stock buy or sell signal in the stock picks subcategory. Table 2 shows the number of stocks that had a gain and loss in the stock prices file. Table 3 shows the number of stock picks that were a buy and sell, which was in the stock picks file.

<table>
<thead>
<tr>
<th>Stock prices subcategory</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total stock gain</td>
<td>140,001</td>
</tr>
<tr>
<td>Total stock loss</td>
<td>130,032</td>
</tr>
</tbody>
</table>

*Table 2: Stock prices subcategory data*
<table>
<thead>
<tr>
<th>Stock picks subcategory</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total stock buys</td>
<td>3,983</td>
</tr>
<tr>
<td>Total stock sells</td>
<td>536</td>
</tr>
</tbody>
</table>

*Table 3: Stock picks subcategory data*

From the Python program that ran on Google Colab, we obtained count values for plunge, stable, and surge values. The number of plunge stocks told us which stocks fell, the number of stable stocks tells us which were steady, and the number of surge stocks tells us which stocks jumped up.

In Figure 7, we were able to find the 2-gram salient terms. They are split into plunge, stable, and surge categories.

![Figure 7: Salient terms](image)

In Figure 8, we have nouns and phrases pulled from social media.
Figure 8: Keywords from social media
User Manual

It is important for our client to have tools to be able to predict stock prices. So we have provided some data for investors and our clients to use. This will let them know if a stock is a buy or a sell.

We know that if the Composite Sentiment Score (CSS) value decreases the event sentiment increases, and if the News Impact Projection (NIP) value increases the event sentiment increases. We can check for those values to make a decision. Table 4 shows common salient terms while Table 5 shows common salient phrases.

<table>
<thead>
<tr>
<th>Salient 2-grams for Surge</th>
<th>Salient 2-grams for Plunge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long term</td>
<td>Integration expenses</td>
</tr>
<tr>
<td>Total compensation</td>
<td>Risk uncertainty</td>
</tr>
<tr>
<td>Executive compensation</td>
<td>2019 compare</td>
</tr>
<tr>
<td>Equity incentive</td>
<td>K acquisition</td>
</tr>
<tr>
<td>Award grant</td>
<td>Big data</td>
</tr>
<tr>
<td>CIC Plan (change in control)</td>
<td>Machine learn</td>
</tr>
<tr>
<td>Shareholder right</td>
<td>Gross margin</td>
</tr>
<tr>
<td>Compensation committee/plan</td>
<td>Negatively impact</td>
</tr>
<tr>
<td>Stock Awards</td>
<td>Percentage revenue</td>
</tr>
<tr>
<td>Compensation plan</td>
<td>Alternative Data</td>
</tr>
<tr>
<td>Incentive Award</td>
<td>Continental resource</td>
</tr>
</tbody>
</table>

*Table 4: Salient terms*
<table>
<thead>
<tr>
<th>Salient phrases for Surge</th>
<th>Salient phrases for Plunge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast actionable information</td>
<td>Private securities Litigation reform act</td>
</tr>
<tr>
<td>Dilute share</td>
<td>Depress company stock</td>
</tr>
<tr>
<td>Common adverse reaction</td>
<td>Debt level</td>
</tr>
<tr>
<td>Common stock</td>
<td>Mutual fund</td>
</tr>
<tr>
<td>Frequent serious adverse reaction report</td>
<td>Actual result differ</td>
</tr>
<tr>
<td>Close imbalance information disseminate</td>
<td>Street quant ratings</td>
</tr>
<tr>
<td>Close website data delay</td>
<td></td>
</tr>
<tr>
<td>Close MOC stock order imbalance</td>
<td></td>
</tr>
</tbody>
</table>

*Table 5: Salient phrases*

**Collection of Data:**
Collection of the data from Twitter required us to write a Twitter scraper using Jupyter Notebook. We were able to collect data based on what the company name was, the date, and the title of the article. We then needed to extract key data from the tweets and write it to a CSV file, so the tweet scraper also had functionality for keywords, and the tickers that each company had. Using the GetOldTweets3 module in Python, we were able to get the old tweets and write them to the CSV file.

**Analyze Dataset:**
The CSV files provided by the Wharton Research Data Services will be used as the input at this stage. In the pre-processing stage, we extract the indicators like CSS, NIP, etc. to build a bigger data set (output) that will be used for analysis in the later stages. This will be done by running Python on Jupyter notebooks. We will be using pandas, matplotlib, and other libraries. Next we will perform LDA analysis.
on the headlines and extract sentiment scores from the topics. This will be used to correlate the indicators with the sentiment score to find the causal factors.

**Case Study:**
For the implementation for the case study, we manually went through the CSV files and chose data points corresponding to a company and event. We chose these data points based on the CAR values for plunges and surges. The data points were then used to collect data based on the company and events around the time to figure out news events that occurred when the stock price had surged or plunged. Using keywords we were able to collect data (explained in the collection of data portion). Then the data would be used in analytics.
We read the data from the all_range_event_withfolderfilename_7days.csv in a Pandas dataframe. Next, we created 3 data frames from that using sampling of the Pandas dataframe. We used Python Pandas for this. In Figure 9, the CAR values are shown to determine the correct stock categories for surge, plunge, and stability.

```python
import pandas
import random
path = '/content/all_range_event_withTICKER_withfolderfilename_7days.csv'

events = pandas.read_csv(path)
surge = (events[events.CAR > 0.05]).sample(n = 1000)
plunge = (events[events.CAR < -0.05]).sample(n = 1000)
stable = (events[(events.CAR > -0.05) & (events.CAR < 0.05)]).sample(n = 1000)
```

*Figure 9: Snippet of code for CAR values*

Next, we collected the SOURCE_NAME, 'EVENT_SENTIMENT_SCORE','CSS', 'NIP', 'PEQ', 'BEE', 'BMQ', 'BAM', 'BCA', 'BER', 'ANL_CHG', 'MCQ' from the relevant folders. The client required us to put these indicator data for each of these events. Figure 10 shows what we did to collect the different sentiment scores.
Figure 10: Code snippet for event sentiment score

We did this for each of the 3 dataframes. Next, we used the source name and news headline to perform Google search in Python, scraped the most important (first link) from that and extracted the news article from them as the news. Content from them was added as a column for each of these dataframes. Figure 11 shows our code to do this.
Next we performed LDA analysis on the collected news to see the most important terms that can caution the user when they are reading the news articles. For this we used the following Python NLP libraries. NLTK and pyLDAvis were the main ones. Figure 12 shows our imports and libraries used to collect the news.
In Figure 13, this was how we dealt with punctuations in the news and how we took them out of our output file.

Figure 14 shows more example code to help with modeling.
This is our list and goals of each of the types of users that our system needs to support.

**Data Collection:** The user’s goal is to collect data using Twitter data based on recent events. Collection of data based on the event is categorized into company by data and event. So using the Twitter scraper that we created, the user should be able to pull data from Twitter and store it into files that will be used for analytics.

**Analyze Data:** Predict the stock prices and classify them as surge, plunge or stable. Extract the sentiment score from crawling news and press-release. Extract the sentiment score from Twitter. Extract the values of the causal factor that will be analyzed in the final project. Train a model to classify them in different categorical labels.

**Case Study:** Using the data collected from the Twitter scraper, the user should be able to sort the key events from the findings into different stock changes for plunges and surge values. These values will be used in the analytics of the data.

Our break down of each goal into units of tasks and subtasks, a combination of which makes up the goal.

```python
from gensim import corpora, models
import pyLDAvis.gensim
import warnings
warnings.filterwarnings('ignore')
pyLDAvis.enable_notebook()
corpus = corpora.MmCorpus('word_model.mm')
dictionary = corpora.Dictionary.load('word_model.dict')
lda = models.LdaModel.load('word_model.lda')

followers_data = pyLDAvis.gensim.prepare(lda, corpus, dictionary)
pyLDAvis.display(followers_data)
```

*Figure 14: More example code*
1) **Collect Data:** To collect data, we need a list of events (news headlines) to pass to the Twitter scraper so we can collect tweets.

1) Analyze Event and run Tweet Scraper. This task is dependent on the task of 2) collecting tweets through a tweet scraper, which is dependent on the system supporting the task of 3) collecting accurate and relevant event information.

![Diagram](Figure 15: Our design for project)

2) **Analyze Data:** To analyze the data, we need to collect the relevant tweets and format them.
1) Data Collection involves preprocessing tweets, which leads to 2) Performing LDA analysis for tweets and sentiment analysis of tweets, when done, can be used to 3) Perform advanced analysis to find the statistical dependencies.

![Project methodology diagram](image)

**Figure 16: Project methodology**

3) **Case Study:** Choose key cases to review and match a correlation between the stock value and the events that occurred during the time period.

1) Case Study is the study of different cases that show changes in stock value. 2) The task is to choose companies where the stock value had a large drop or gain. 3)
The third task is to collect data via Twitter during the time of the surge or plunge of the stocks. 4) The data is collected using keywords so there can be a correlation to the stock’s value. 5) Once the data is collected, it will then be analyzed.

This is our system represented as a set of workflows, as shown in Figure 17, with a list of workflows covering each goal.

![Figure 17: Project methodology in depth](image)
Timeline

This is our timeline we set at the beginning but we will modify it to fit delays and changes due to going online:

- Initial analysis of the raw data [February 14- Completed]
- Isolating Events [February 28- Completed]
- Keyword Extraction [February 28- Completed]
- Twitter Scraping [March 16- Completed]
- Web Scraping [March 16- Completed]
- VTechWorks Draft Submission [April 26- Completed]
- Semantic Analysis [April 20- Completed]
- VTechWorks Final Submission and Final Project Submission [May 6- Submitted]
Problems

Our main problem was having the TweetScrape [2] working to extract keywords from different tweets. This took a while but we were able to figure out by collaborating with a different group about this subject, the Twitter disaster group.

We were also forced to not meet up due to the coronavirus situation. We had to adjust meetings from home and communicate our ideas through Zoom. We were able to successfully get our work done the last couple weeks of the project and stayed focused on the project.
Solutions

From tweets that were collected based on the Case Studies, we used LDA to find words that frequently appeared in Case Studies that had a positive or negative CAR value. In Figures 18 and 19, there are two of the visualizations of the positive and negative events.

*Figure 18: Frequent words in case studies, positive*
While these visualizations show which words were most frequent, they do not explain which types of words were associated with the grouping of events. The table below highlights a few categories of words that appeared in either the group of the positive CAR events, or the group of the negative CAR events. All words in the ‘Associated Words’ columns appear in the corresponding set of top 30 words for either the positive or negative group. In Figure 20, we found a correlation between the words and CAR values.

<table>
<thead>
<tr>
<th>CAR Value</th>
<th>Word Category</th>
<th>Associated Words</th>
<th>Word Category</th>
<th>Associated Words</th>
<th>Word Category</th>
<th>Associated Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Monetary</td>
<td>Price, cash, value, million, billion</td>
<td>Assertive</td>
<td>Bring, acquire, make, take</td>
<td>Positive</td>
<td>Good, win, like, wow</td>
</tr>
<tr>
<td>Negative</td>
<td>Disagreements</td>
<td>Battle, dispute, refuse, mislead</td>
<td>Uncertainty</td>
<td>Would, could, chance</td>
<td>Legal</td>
<td>Patent, court, settlement, trial</td>
</tr>
</tbody>
</table>

**Figure 20: Words correlation with CAR values**
As can be seen in Figure 20, words that are associated with money, positivity, and assertive verbs were common in the positive events, while the negative events had words that were used to describe disagreements, uncertainty, and legal issues.

The increase or decrease in CAR value can be explained by investigating why those words appear for those groups. For the monetary words, valuations, revenues, and prices of companies and their products are usually referenced when earnings reports occur, or a product is released. The good news from these events typically drives up share price. The assertive words are verbs that describe actions the companies have made, and due to their connotation that implies decisive actions, are usually coupled with context that is beneficial for the company. These words are also used when mergers occur, where mergers that are perceived as beneficial end up driving share price up.

The positive words are perhaps the simplest to explain; positive sentiment towards a company reflects confidence in that company’s abilities, which will increase share price. The disagreement words are associated with two entities having a conflict, which in this case is two companies. A common cause is lawsuits, which can greatly negatively impact a company if legal restrictions are placed upon them, lowering their share price.

The uncertainty words are all words that imply hypothetical situations and unpredictability, and when these words are used to describe a company’s actions or event relating to a company, the uncertainty can cause a drop in confidence, lowering share price accordingly.

Finally, the legal terms are similar to the words associated with disagreements. Lawsuits, patent issues, and trials can be the result of company officials or actions by the company itself, and if the associated company is found to be in the wrong, the penalties that result can cause the share price to drop as confidence lowers or the company’s capabilities are lessened.
We collected the tweets for our case studies and grouped them by the dates. Next, we performed sentiment analysis on these grouped collected tweets to find the public sentiment about the event based on the date. After, we normalized those scores to standardize these scores. Here one can see the bar plots of the normalized sentiment scores based on the date.

In Figure 21, we can see the negative sentiment of the YUM stock. In Figure 22, we can see the negative sentiment of the HPQ stock. In Figure 23, we can see the positive sentiment of the GMstock. In Figure 24, we can see the positive sentiment of the FOSL stock.

Figure 21: YUM stock, negative sentiment
Figure 22: HPQ stock, negative sentiment

Figure 23: GM stock, positive sentiment
To conclude, Figure 25 shows an all negative sentiment bar graph and histogram. Figure 26 shows an all positive sentiment bar graph and histogram. We are able to see the different sentiment of the stocks from this data. The surge shows more positive skewed tweets and the plunge shows more negative and neutral skewed tweets at the time of the event.
Figure 25: All Negative Sentiment
Figure 26: All Positive Sentiment
Future work

We would like to continue testing different data and collecting more data to further analyze different stock data.

Acknowledgements

We would like to thank the VT CS department, Ms. Ziqian Song, and Dr. Edward Fox for their continued support in this project. To contact our client, her email is ziqian@vt.edu. We would also like to thank our classmates for providing constructive criticism on our milestone presentations.

References

