Sensitive Stonks

Analyzing Stock Market Prices with Twitter Sentiment

April 23, 2021

Carrera, Santiago; Hua, Helen; Koltai, Andrea; Seoane, Maria; Stevens, Stephanie

Introduction & Motivation

The objective of our project is to create a tool which illustrates the relationship between stock price and aggregated tweet sentiment through time, as well as suggesting Twitter accounts to follow for insight on specific stocks. This can be a useful tool for amateur investors wishing to incorporate this information in their investments, professional investors looking to gain a competitive advantage, and also companies interested in understanding how public sentiment of their company relates to their performance.

New online investment platforms have democratized the access to the stock markets. We believe that our tool will help new or amateur investors to select who to follow, and also how a particular stock reacts to tweets. If a stock is very volatile with respect to positive or negative tweets, the user can be aware of this, and be proactive when positive or negative tweets start to appear online. For professional investors, this tool can also complement their market knowledge and help their decision-making process. Finally, our tool can be useful for companies to gauge their stock sensitivity to Twitter engagement.

Problem Definition

It is widely known that public sentiment can influence stock performance, but there is currently no solution available to the amateur investor, especially one that doesn't have a background in Natural Language Processing (NLP) and access to vast amounts of data. Additionally, understanding and identifying twitter accounts with high levels of engagement to a particular stock maybe provide increased benefit to the user as they will gain insight into the context of sentiment trends.

Survey

Sentiment analysis is a widely used technique to discover the overall public opinion or feelings towards a particular product or topic. The analysis involves NLP and machine learning algorithms to determine the polarity (positive, negative or neutral) of a sentence, or classify it into more descriptive sentiment categories (i.e. anger, fear, happiness). Twitter, the micro blogging platform with its 180 million daily users [1], plays a remarkable role in social media and has become a popular resource to extract insights about people's mood. For instance, Luo et al. performed sentiment analysis on Twitter to analyze the opinions of HPV vaccinations and the results demonstrated the different trends across different years. [2]

The appropriateness of using sentiment analysis on stock markets can be questioned by the theory of market efficiency. Malkiel concluded in [3] that although the stock market is in general efficient in its pricing of securities, inevitably from time-to-time pricing inaccuracies arise, and will exist for short periods of time, due to either irrational behavior or slow transmission of information to be reflected in prices. Furthermore, Nofsinger [4] argued that the financial economic decisions are significantly influenced by shared emotions. The increase of social mood is associated with optimism and can lead to extreme overconfidence, in contrary the decrease in mood is associated with pessimism, fear and lower risk-taking level.

Bollen et al. [5] conducted causality analysis of the time series of mood profile of daily Twitter feeds on Dow Jones Industrial Average (DJIA). Gross-Klussmann et al. [6] studied the relationship between sentiment of tweets from stock market experts and its correlation with the indices stock returns. Although both studies found statistical evidence for correlation between tweet sentiment and market indices, these analyses were conducted on market average rather than providing insight on individual stocks.

While these results are promising, several studies have since performed various analyses on individual stocks. Pagolu et al. [7] conducted a predictive analysis to reveal the relationship of tweet sentiments and the stock prices of Microsoft. Although their results were promising, the test cases were selected randomly, thereby questioning the predictive power and future use their model. Sul et al.[8] revealed correlation between the cumulative sentiment of tweets of individual firms by Twitter accounts with fewer followers and their effect in the following day's stock returns. Souza et al. focused their sentiment analysis [9] on five retail companies, conducting causality analysis and using an auto-regressive model to predict stock returns based on tweet sentiment. They also used newswires as a proxy for market sentiment.

Additionally, Shah et al. reviewed a series of techniques used for stock market analysis, ranging from traditional statistical methods, such as ARIMA and regression, to machine learning and sentiment analysis [10]. Shah summarized that sentiments from news articles and media can drive market fluctuations and provide insights to how the market reacts.

Kolchyna et.al. [11] used the combination of lexicon approach and supervised learning algorithms. Similar to another experiment conducted by Kolchyna et.al [9], and Jurek et al. [12], there is a proposal for an improved sentiment analysis, using not only the traditional lexicon approach, but enhancing it with an ”evidence-based combination function” to try to give a sense of the intensity of the sentiment. Da Silva [13] pointed out that the ensemble methods (such as the majority vote of different supervised learning algorithms) in combination with opinion lexicon scores, improved the classification accuracy of tweets into positive and negative categories. Da Silva also suggests that these results can be further improved by using additional clustering methods.

Furthermore, Hogenboom et al. [15] revealed the importance of emoticons in correct classification of tweets. Although their research was based on a small set of Dutch tweets, they reported a significant improvement in classification accuracy by considering the sentiment score of emoticons.

In addition to studying the effectiveness of sentiment analysis in stock performance, there have been a large number of studies on how to appropriately visualize the results of sentiment analysis. As illustrated in papers [16, 17, 18, 19], there are a variety of visualizations that have been used to show the results of a sentiment analysis in the past. Traditional bar, line, and doughnut charts seem to be the most prevalent, while some studies have taken different approaches like word clouds, bubble charts, and ribbon charts.

Proposed Method

With the rise of social media in the last decade, sentiment analysis appears to be a compelling method for stock market analysis to exploit these market anomalies. While various sentiment analyses of social media posts in relation to the stock market already exist, the novelty of our approach is the introduction of an easy-to-use user interface that enables users to select a company and a time period of interest to compare its historical stock price with the sentiment of related tweets. Moreover, it recommends Twitter accounts to follow based on the company chosen, in order to better exploit the information that these accounts present.

Data Gathering and Filtering

Twitter data was gathered by processing existing datasets hosted in The Internet Archive[22] for tweets. This data source uses the ”Spritzer” version of Twitter grabs; a 1% random sample of all tweets. After filtering the data set to keep relevant tweets, we additionally combined it with a Kaggle dataset on financial tweets [25].

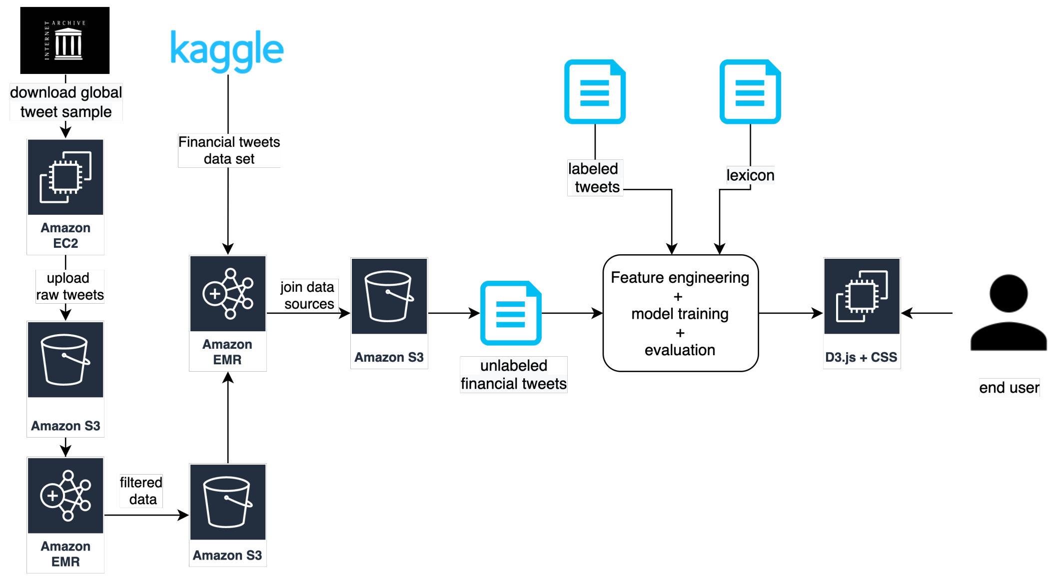

Figure 1: Data gathering and filtering

Due to the large volume of data, we decided to process it using Amazon Web Services (AWS). An Elastic Cloud Compute (EC2) instance was setup in AWS and the data was downloaded to the instance using a torrent service. All available tweets between July 2020 and January 2021 were downloaded and later uploaded to Amazon Simple Storage Service (S3).

Using Apache Spark running on an AWS Elastic Map Reduce (EMR) cluster, we performed multiple filtering steps to obtain our final dataset. The list of companies and their respective industry and symbol to query for were obtained from a GitHub repository[24]. Available tweets were filtered to include only English-language tweets, containing either # or $ symbol followed by an S&P 500 company symbol. The volume of data processed was around 500GB, and totalled approximately 1 billion tweets, and was processed using a 5 instance m5.xlarge cluster.

In order to acquire the financial information needed, we decided to use Yahoo! finance. Using the yfinance[23] python library, we obtained and stored the opening price, highest price, lowest price closing and adjusted closing price, and trading volume for the day for each S&P500 company for the given time period.

Sentiment Analysis

The dataset used to train the model was from Sentiment140 [20], which contains approximately 1.6 million tweets gathered using the Twitter API and labeled with positive or negative sentiment, along with additional metadata such as the date of the tweet and the user that posted the tweet. Using a labeled dataset saves time and resources needed to label a large dataset of tweets.

A series of functions and steps were performed to clean the text of the tweets. Website links, any characters except alphanumeric characters, hashtags, and punctuation were all removed. Then we calculated lexicon-based features, such as counting the number of positive and negative words in a tweet and the number of negating words. The text was then tokenized as the final data cleaning step.

In our approach of the sentiment analysis model, we first transformed the tweet text to numerical features using the TfidfVectorizer package in scikit-learn for Python. The algorithm converts a collection of text to term-frequency-inverse document frequency (TF-IDF) features, a numerical representation of the tweets. The features were then used to fit a Naïve Bayes classifier to determine if the sentiment of the tweet was positive or negative [21]. The model achieved 75% to 80% accuracy on the training data. The model was then used on the stock tweet data as described in the Data Gathering section to predict the sentiment of the stock related tweets.

Twitter Recommendation

In addition to the sentiment analysis of the tweets, we also added a feature to show the top recommended twitter accounts to follow related to the selected company. First, the "movement of day" for the stock is calculated, based on whether the stock price increased or decreased between opening and closing. This is joined with our cleaned dataset including predicted sentiment labels, the tweet, date, user account. We also filtered out any accounts that had less than 150 tweets. To rank the accounts, we created a score for each account based on the tweet sentiment and movement of the day using simple Pearson correlation. The Twitter users with the highest correlation were selected to be recommended in the dashboard.

Visualization

The front end of our project is an interactive dashboard which allows a user to select a particular stock from a drop-down menu containing companies in the S&P 500 and a time period between 2015 and 2020 to display the desired data. While research suggests there is no "right way" to visualize sentiment analysis, we have chosen to keep our visualizations as simple as possible. We used a dual axis bar and line chart since they are more widely understood by a broader audience. This will hopefully reduce confusion and misleading results. We utilized D3.js, HTML, and CSS in building these visualizations because of their versatility and ease of use. The visualization is hosted on an AWS EC2 instance to simplify the access to the dashboard.

In the dashboard, when the user selects a stock and a time period of interest, they will see a visualization like the one shown in Figure 2 below. We have incorporated a line chart representing the selected company's daily close price for the desired time period. A tooltip is included on the line chart to show the stock price on a particular time point. This allows the user to see the historical relationship of overall Twitter sentiment to stock price. On the same visualization, the bars represent the count of positive and negative sentiment tweets for a particular day for the time period selected by the user. In addition to our visualization, we included a list of twitter accounts to follow in a column to the right. Provided with the user account recommendations are the number of related tweets from the user and a link to follow the respective twitter account.

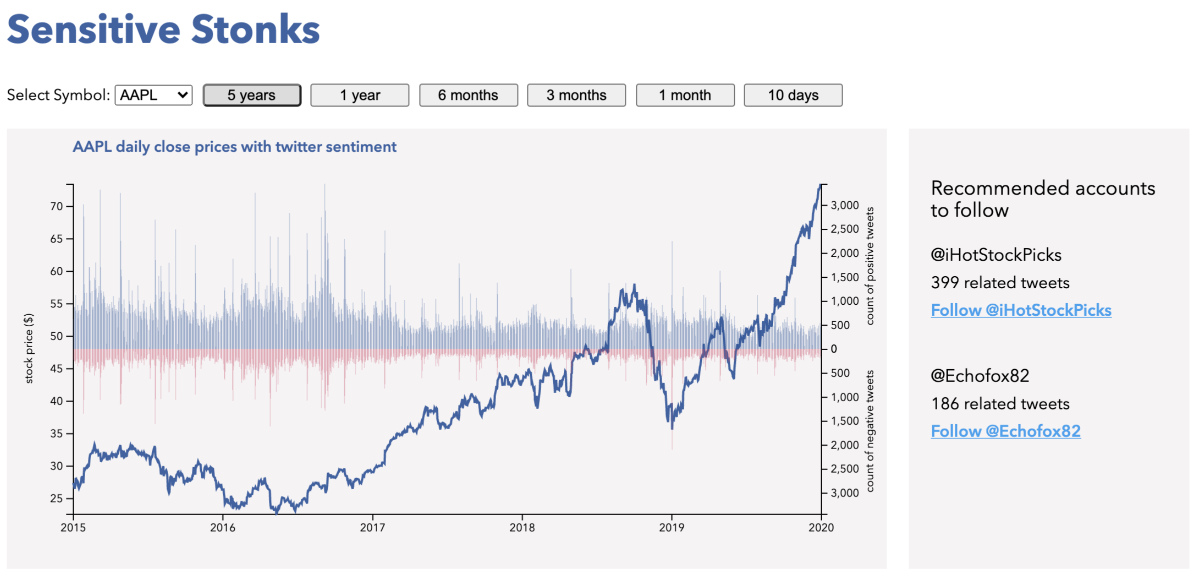

Figure 2: Interactive dashboard for stock and sentiment analysis, with twitter recommendations on the right side

Innovation and Contributions

All team members have contributed similar amount of effort, which can be shown in Table 1, for each member's contributions.

Overall, we believe our approach will be both successful and useful because of the following innovations:

- Utilizing a new combination of NLP models by combining the research of past experiments
- Providing the user with recommended Twitter accounts to follow for a more in-depth view the sentiment of a particular stock
- Creating a web application to provide access to broader audience

Group Member	Contributions
Andrea	Data Cleaning, Sentiment Analysis Model, Stock and Sentiment Visualization
Helen	Research, Data Cleaning, Sentiment Analysis Model, Final Report and Poster Writing
Maria	Data Gathering, Twitter Recommendation Model, Twitter Recommendation Visualization
Santiago	Data Gathering, Data Infrastructure, Model running
Stephanie	Research, Progress and Final Report Writing, Poster Writing, Survey Creation

Table 1: Member contributions

Experiments & Evaluation

Data Gathering

When retrieving the data from the Internet Archive, we tested several approaches including limiting our search to only tweets tagged with 'finance' or stock', however we found that this severely restricted the number of returned tweets. Subsequently, we decided to remove this criterion and were able to retrieve a larger number of tweets.

An unexpected issue we encountered was related to the common use of some of the ticker symbols. For example, “TXT”, “PSA”, “NOW” are words that are used in other contexts. Moreover, when filtering for single letter tickers like “V” or “K”, we found a large number of tweets that had no relation to our subject matter.

Moreover, of the initial 1 billion tweet sample, after the data was filtered we found out that the number of tweets we can actually use is very small, with most companies only having number of tweets in the hundreds. Given this reality, we decided to enrich this small dataset with an additional Kaggle dataset [25] that has a large number of tweets that have already been identified as related to our topic of interest. This dataset is limited to tweets for Amazon, Apple, Google, Microsoft, and Tesla between 2015 and 2020, hence these companies will be the more interesting ones to analyze, rather than the full S&P 500 we initially envisioned.

Modeling

In our model creation process, we evaluated three different models and tested different hyperparameters within those models. For the algorithm evaluation, we tested a Naive Bayes (NB) classifier, Support Vector Classifier (SVC), and Neural Networks. The models were also evaluated using a sample set (n = 10,000) of the training data set. All of the models performed similarly, with test accuracies between 70 and 75%, shown in Table 2. With the similar performance across the models, the model was chosen based on the speed of the algorithm. The NB classifier was the fastest with a runtime of 0.935 seconds, whereas the SVC and neural network models were significantly slower with runtimes over 200 seconds using the sampled training set. Another consideration was comparing the training accuracies. Both SVC and Neural Networks had training accuracies between 85 to 95%, showing signs of overfitting, whereas the NB classifier had similar accuracies for both training and testing.

Model	Training Accuracy (%)	Testing Accuracy (%)	Runtime (s)
Naive Bayes	79.65	73.6	0.935
Support Vector Classifier	86.50	73.4	233
Neural Network	92.11	70.65	235

Table 2: Model accuracies and run time

For our data cleaning pipeline and transformation, the TF-IDF vectorizer was used to convert the text to numerical features. Within the TfidfVectorizer function, the maximum number of features to include in the TF-IDF matrix is a hyperparameter that can be tuned. Using the sample training set, a range from 100 to 10,000 maximum features was tested for the vectorizer. With the Naive Bayes classifier as the model, we found that the testing accuracy plateaus at around 72.3% using 10,000 maximum features in the TF-IDF vectoriser, shown in Figure 3. Therefore, in our final Naive Bayes model, 10,000 maximum features was used.

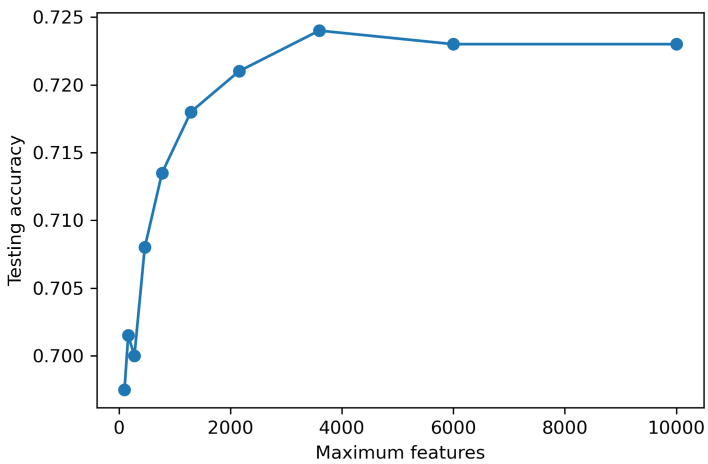

Figure 3: Naive Bayes testing accuracy using different maximum number of features in TF-IDF vectorizer

Visualization Design

While designing our visualizations for the application, our team discussed the best way to view the potentially complex information in the most succinct way. As stock price changes over time, we wanted something simple to visualize the trend for the price, leading to using a line graph to plot the stock price. We also wanted to visualize both the negative and positive sentiments of the tweets in the visualization, rather than a calculated measure from the sentiments to make it simple to understand. The twitter recommendation section had two possible locations in the dashboard, on the right side of the graph or below the graph. We decided to place the twitter recommendation on the right side, as it would keep it level with the graph, whereas placing it below the graph would make it possible to not notice the twitter recommendations.

Evaluation

We have developed a survey that can be sent to users to determine the overall effectiveness of and satisfaction with the web application. This survey attempts to identify a user's motivation for using the site (novice/casual stock trader, experienced trader looking for a competitive edge, or company looking to understand its sensitivity to overall Twitter sentiment). The survey also asks the user if a stock purchase will be made as a result of the information on the application as well as a rating of how helpful they found the information provided. Based on the results of our survey, we found that 89% of respondents were novice stock traders, 56% of whom did not have a particular stock in mind. When asked how helpful the information found on the site was, on average, the respondents rated our site as a seven out of ten, with 75% of users responding with at least a seven. Additionally, 40% of respondents said they would make a stock purchase based on the information found on the site.

Additionally, we wanted to understand the correlation of each company's daily stock change with the daily overall Twitter sentiment. We found that actual correlations were quite low for each of the companies studied.

Stock	Change Within Day	Change from Previous Day	Change to Next Day
AAPL	0.163	0.249	0.066
GOOGL	0.145	0.193	0.042
MSFT	0.139	0.122	0.021
TSLA	0.056	0.110	0.027
AMZN	0.146	0.167	0.055

Table 3: Correlation of Daily Stock Change to Twitter Sentiment

Conclusion

Due to the time constraints of the project, there are many items we did not implement. Future iterations of this project should include a wider variety of companies, a larger dataset of tweets, a more sophisticated, comprehensive method for identifying stock-related tweets, cryptocurrencies, and an automatic refresh of the data and underlying models. We could also include other social media networks in our analysis and dig deeper into the true correlation of Twitter sentiment (i.e., does Twitter sentiment drive stock performance, or does stock performance drive Twitter sentiment?). Additionally, we could include more visualizations in the dashboard, such as a heatmap to illustrate the correlation of daily stock price changes and Twitter sentiment, and a choropleth map highlighting tweet geographical locations.

It is evident, by our analysis, that individual stocks have very different relationships to positive and negative Twitter engagement. For some companies, it is not necessarily the existence of positive or negative tweets, but perhaps the volume of tweets, that seem to occur alongside a stock price change. While it is difficult (and inadvisable) to make inferences on a stock's future performance based on Twitter sentiment alone, we believe our project is a worthwhile tool in any investor's toolkit.

References

[1] Twitter Daily Active Users statistics: https://www.businessofapps.com/data/twitter-statistics/

[2] Luo, X., Zimet, G., & Shah, S. (2019). A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008-2017). Human vaccines & immunotherapeutics, 15(7-8), 1496-1504. doi: 10.1080/21645515.2019.1627821

[3] Malkiel, B. G. (2003). The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1), 59-82. doi:10.1257/089533003321164958.

[4] Nofsinger, J. R. (2005). Social mood and financial economics. The Journal of Behavioral Finance, 6(3), 144-160. DOI: 10.1207/s15427579jpfm06034

[5] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8. DOI: https://doi.org/10.1016/j.jocs.2010.12.007.

[6] Groß-Klußmann, A., König, S., & Ebner, M. (2019). Buzzwords build momentum: Global financial Twitter sentiment and the aggregate stock market. Expert Systems with Applications, 136, 171-186. DOI:https://doi.org/10.1016/j.eswa.2019.06.027.

[7] Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016, October). Sentiment analysis of Twitter data for predicting stock market movements. In 2016 international conference on signal processing, communication, power and embedded system (SCOPES) (pp. 1345-1350). IEEE. doi: 10.1109/SCOPES.2016.7955659.

[8] Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. Decision Sciences, 48(3), 454-488. DOI:https://doi.org/10.1111/deci.12229.

[9] Souza, T. T. P., Kolchyna, O., Treleaven, P. C., & Aste, T. (2015). Twitter sentiment analysis applied to finance: A case study in the retail industry. arXiv preprint arXiv:1507.00784.

[10] Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 26.

[11] Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:1507.00955.

[12] Jurek, A., Mulvenna, M. D., & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics. Security Informatics, 4(1), 1-13. doi:10.1186/s13388-015-0024-x.

[13] Da Silva, N. F., Hruschka, E. R., & Hruschka Jr, E. R. (2014). Tweet sentiment analysis with classifier ensembles. Decision Support Systems, 66, 170-179. https://doi.org/10.1016/j.dss.2014.07.003.

[14] Coletta, L. F. S., da Silva, N. F. F., Hruschka, E. R., & Hruschka, E. R. (2014, October). Combining classification and clustering for tweet sentiment analysis. In 2014 Brazilian Conference on Intelligent Systems (pp. 210-215). IEEE. doi: 10.1109/BRACIS.2014.46.

[15] Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., & Kaymak, U. (2013, March). Exploiting emoticons in sentiment analysis. In Proceedings of the 28th annual ACM symposium on applied computing (pp. 703-710). doi:10.1145/2480362.2480498

[16] Lu, Y., Hu, X., Wang, F., Kumar, S., Liu, H., & Maciejewski, R. (2015, May). Visualizing social media sentiment in disaster scenarios. In Proceedings of the 24th international conference on world wide web (pp. 1211-1215). DOI:https://doi.org/10.1145/2740908.2741720

[17] Gregory, M., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., & Turner, A. (2006, July). User-directed sentiment analysis: Visualizing the affective content of documents. In Proceedings of the Workshop on Sentiment and Subjectivity in Text (pp. 23-30).

[18] Kucher, K., Paradis, C. and Kerren, A. (2018), The State of the Art in Sentiment Visualization. Computer Graphics Forum, 37: 71-96. https://doi.org/10.1111/cgf.13217

[19] Duan, D., Qian, W., Pan, S., Shi, L., and Lin, C. (2012, September). VISA: a visual sentiment analysis system. In Proceedings of the 5th international symposium on visual information communication and interaction (pp. 22-28).

[20] Go, A., Bhayani, R., Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.

[21] Lilleberg, J., Zhu, Y., Zhang, Y. (2015, July). Support vector machines and word2vec for text classification with semantic features. In 2015 IEEE 14th International Conference on Cognitive Informatics Cognitive Computing (ICCI* CC) (pp. 136-140). IEEE.

[22] Internet Archive Search: collection:twitterstream. (n.d.). Internet Archive. Retrieved March 22, 2021, from https://archive.org/search.php?query=collection:twitterstream

[23] yfinance. (n.d.). PyPI. Retrieved March 28, 2021, from https://pypi.org/project/yfinance/

[24] datasets/s-and-p-500-companies. (n.d.). GitHub. Retrieved March 22, 2021, from https://github.com/datasets/s-and-p-500-companies

[25] Tweets about the Top Companies from 2015 to 2020. (2020, November 26). Kaggle. https://www.kaggle.com/omermetinn/tweets-about-the-top-companies-from-2015-to-2020?select=Tweet.csv