# Twitter Sentiment Analysis to Predict Stock Performance

Mentees: Zoey Yao, Elizabeth Sun, Abigail O'Neill, Varun Agarwal
Mentor: Tanvi Khot
University of California, Berkeley, Undergraduate Lab at Berkeley

## Background:

In the past decade, social media platforms have become a large influence on the world economy. The stock market in particular was impacted because the stock price is based on the present value of expected future dividends, which can be easily shifted by public perceptions. Our hypothesis is that there is a positive correlation between positive reactions to tweets about a company and stock prices, and vice versa.

## Research Question:

In order to learn more about how stock prices fluctuate, we focused on one potential contributing factor: public opinion, and more narrowly, Twitter and social media posts from the S&P 500. We proposed our research question: How strong is the correlation between positive tweets and stock prices rising for top companies and how can this data best be modeled to predict stocks?

Our research question will be explored from two aspects:

1. We predict that there is a positive correlation between positive reactions to tweets sent by company heads and stock prices the next week, and vice versa.
2. We will be building a model to observe this relationship. The model will be able to predict future stock prices based on a tweet that was posted in the same time block and its attributes, specifically its likes, shares, and content.

## Data:

Our references included datasets of stock prices on top fortune 500 companies such as Google, Apple, and Amazon. Along with that, we referenced tweets made by and about those top companies through another dataset. We then combined these sets into time blocks, the stock prices affected, and the tweets about these companies. We also added extra features extracted from the two main features such as the like count, share count, and sentiment of a tweet. Finally, we predicted values by splitting the tweets and stock values into their respective companies.

To build the model, after feature extraction, we used a LSTM model as well as other regression models to predict future stock prices. We also utilized tools and libraries like Pandas, Numpy, Tensorflow, etc. to help build our model.

## Results:

***RIDGE, Lasso, Elastic Net***:

We cross compared the results yielded by OLS, Ridge Regression, Lasso Regression, and Elastic Net, while analyzing the cross validation error of regression results versus tuning parameters.

Graphs below in the order of Ridge (Figure 1) (best score = 296.249, best alpha = 1), Lasso (Figure 2) (best score = 295.835, best alpha = 1), and Elastic Net (Figure 3) (best score = 295.835, best alpha = 1, best l1 ratio = 1) for Amazon.

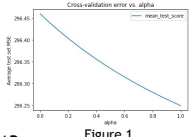


Figure 1

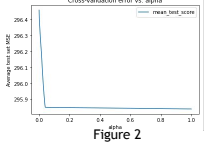


Figure 2

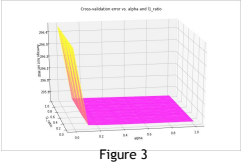


Figure 3

***LINEAR:***

Graphs below display (Figure 4) the sentiment and actual change of stock prices and the line of best fit for the graph, and (Figure 5) the residual plot for Google's stock prices. In Figure 5 we can see that the data is roughly centered around zero. We also applied linear regression to several other company tables and received similar results. We also calculated the RMSE for all test predictions, receiving values ranging from 10.99 to 19.29.

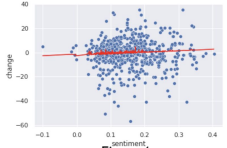


Figure 4

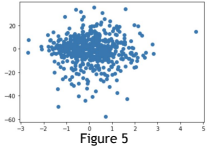

Figure 5

***LSTM:***

Plotted below (Figure 6) is the predictions using the LSTM models and the actual values, receiving a loss of 0.0392. The blue line represents the actual values of stock prices and the orange line represents the predicted values of stock prices.

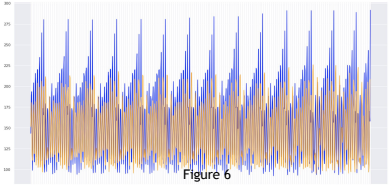

Figure 6

## Conclusion:

Among all the different models, some show some correlation/ positive relationship between our features used and the stock prices, while others do not best fit the datasets. Our linear regression models show that a linear relationship could possibly be a good fit between positive reactions or sentiment to tweets and stock prices. After calculating the RMSE values, the difference between the true values and the predicted values did have a relatively small RMSE, however when visualizing the line of best fit through the graph, as the line looks relatively flat it might not be the best model to fit to our data. Looking at the ridge regression model, the visualizations did show some promise as to proving a strong relationship between stocks an tweets, however the RMSE values were quite high. Finally, looking at the visualizations for the LSTM model, the loss values seem the strongest, and the visualization for stock prices seems very similar. We conclude therefore that the LSTM model was the strongest model to demonstrate the relationship between stocks and tweets. If we had more time, we would research more into the LSTM model and improve on the features we used in order to prove the relationship further. The research done through our code and external papers leads us to infer some relationship between stocks and tweets relation to those companies.

## Future Research and Limitations:

While the open-source datasets did not provide the most complete information, we are looking forward to extrapolating the past behavior of stock prices fluctuations and creating weighted functions to better fit the data to our current model. The tweet dataset didn't seem to have very relevant tweets, as many of the like and comment values were low, meaning that they might have not had too much impact. Because of this, instead of the specifics of a certain tweet we decided to instead count frequency, which may have limited how we analyzed the tweets and predicted stock prices based on that. In the future, one major direction might be joining with natural language processing to more closely interpret the social media sentiment.

## References:

Tweets about the Top Companies from 2015 to 2020
US historical stock prices with earnings data
Hadi Rezaei, Hamidreza Faaljou, Gholamreza Mansourfar, "Stock price prediction using deep learning and frequency decomposition."
Singh, Aishwarya. "Stock Price Prediction Using Machine Learning: Deep Learning." Analytics Vidhya, 25 Oct. 2018
Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (n.d.). The effects of twitter sentiment on stock price returns.