# Stock Prediction Based on Sentiment of News Articles and Past Stock Trend

Jiahao Zhang; Abhi Vora; Fengjun Li; Hongtao Zhao
Boston University, Department of Electrical and Computer Engineering

## Problem Definition

The aim of this project was to predict the next day's Stock Price of certain companies using Data Learning and make some money!

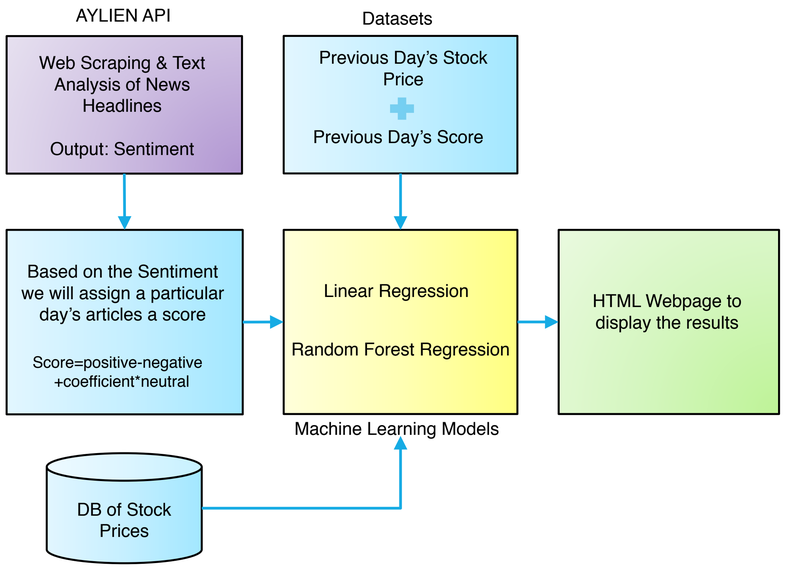


**Figure 1.** Overview of the System Architecture

## Methodology

In this project we wanted to analyze the impact of sentiment of news on a particular companies' stock price. To make our model more accurate we incorporated the stock price trend over the past 6 months.

Multiple modules implemented in our system are:

- Aylien API – Used to perform Web Scraping and Natural Language Processing to obtain a sentiment of each news article.
- Stock Price Database – 8 companies with 6 month's prices.
- Data Processing – About 5000 news articles in a 6 month period for each Company. Concise it to match the number of days, about 127.
- Machine Learning – Performed 2 models, Linear Regression and then Random Forest Regression.
- Webpage – A simple User Interface to view the Stock Price Prediction.

| | | |
|---|---|---|
| 26 | positive | 2017-11-17 12:40:08+00:00 |
| 27 | positive | 2017-11-17 12:37:52+00:00 |
| 28 | positive | 2017-11-17 10:09:48+00:00 |
| 29 | negative | 2017-11-17 06:46:09+00:00 |
| 30 | positive | 2017-11-17 05:27:26+00:00 |
| 31 | negative | 2017-11-17 04:00:29+00:00 |
| 32 | positive | 2017-11-17 03:06:00+00:00 |
| 33 | positive | 2017-11-17 03:03:36+00:00 |
| 34 | positive | 2017-11-17 02:24:26+00:00 |
| 35 | positive | 2017-11-16 23:13:45+00:00 |
| 36 | positive | 2017-11-16 21:46:05+00:00 |
| 37 | negative | 2017-11-16 21:41:19+00:00 |
| 38 | positive | 2017-11-16 21:28:18+00:00 |
| 39 | neutral | 2017-11-16 21:23:51+00:00 |
| 40 | positive | 2017-11-16 21:05:48+00:00 |
| 41 | positive | 2017-11-16 20:52:13+00:00 |
| 42 | neutral | 2017-11-16 20:26:21+00:00 |
| 43 | positive | 2017-11-16 19:51:47+00:00 |
| 44 | negative | 2017-11-16 19:42:09+00:00 |
| 45 | positive | 2017-11-16 19:42:09+00:00 |
| 46 | positive | 2017-11-16 19:42:09+00:00 |

| | |
|---|---|
| 2017/11/17 | 20 |
| 2017/11/16 | 14 |
| 2017/11/15 | 30 |
| 2017/11/14 | 2 |
| 2017/11/13 | 5 |
| 2017/11/10 | 52 |
| 2017/11/9 | -6 |
| 2017/11/8 | 0 |
| 2017/11/7 | 12 |
| 2017/11/6 | 1 |
| 2017/11/3 | -16 |
| 2017/11/2 | 11 |
| 2017/11/1 | 7 |
| 2017/10/31 | 8 |
| 2017/10/30 | 0 |
| 2017/10/27 | 11 |
| 2017/10/26 | 12 |
| 2017/10/25 | 1 |
| 2017/10/24 | 21 |
| 2017/10/23 | 23 |

**Table 1.** Data Processing: News Sentiment and Score from API

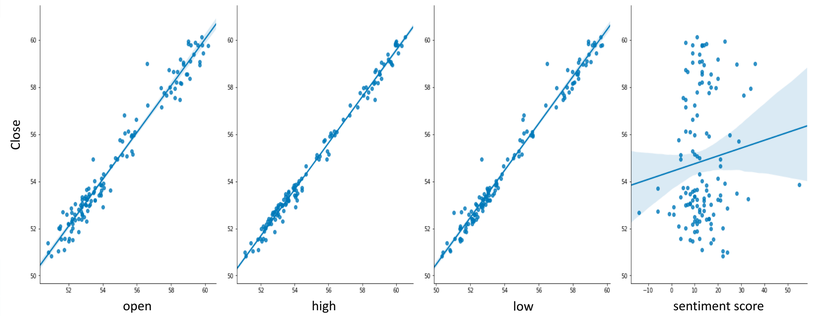


**Figure 2.** Relationship between Features and Stock Price

## Results

We have tested 32 trading days' data with our prediction.

- For previous Linear Regression model, there were only two features: Sentiment score, Closing Price. The R-squared co-efficient was 0.02.
- Random Forest Regression model, five more features were added. Out of the 8 companies, in 4 companies our prediction ranged from ± 5%.

| Date | Real Price | Predicted Price | Error Rate(%) |
|---|---|---|---|
| 12/05/2017 | 275.54 | 275.92 | 0.13 |
| 12/04/2017 | 277.97 | 272.88 | 1.83 |
| 12/01/2017 | 271.38 | 275.05 | 1.35 |
| 11/30/2017 | 276.8 | 268.42 | 3.03 |
| 11/29/2017 | 269.3 | 267.00 | 0.87 |
| 11/28/2017 | 267.99 | 266.14 | 0.70 |
| 11/27/2017 | 265.58 | 265.59 | 0 |
| 11/24/2017 | 265.88 | 265.55 | 0.13 |

**Table 2.** Comparison between Real Closing Stock Price and Predicted Closing Price of Boeing Inc.

## Conclusions

- In this project we learnt Data Processing and Machine Learning Techniques which will be helpful for the future.
- We used Aylien API for Web Scraping and Sentiment Analysis, Random Forest Regression as machine learning model and a webpage to display the predicted result.
- The predicted price and error rate is acceptable but can be improved in the future.

## Current Hurdles & Future Work

- Our current work is handicapped by the Aylien API as we have a cap on the news articles we can access in a given time span.
- We are also reliable on the API to give us accurate results, in terms of news articles as well as its sentiment analysis.
- In the future we plan to perform our own Sentiment Analysis using either N-Gram Algorithms or Neural Networks.

## Acknowledgments

- We would like to thank Professor Osama for his professional advice and consistent encouragement throughout the whole project.
- We would also like to thank TAs Shruti, Prateek for their patience and extraordinary support.