

# TNL STOCK PREDICTOR

## Final Report

Tanushri Singh, Nickolas Ramos, Lawrence Evangelista  
CS 6375 Term Project, May 2019  
The University of Texas at Dallas

### **1 Introduction:**

The premise for our term project is the prediction a specific company's stock prices for the next week given data obtained about the selected corporation. This stock price prediction is augmented by sentiment analysis of recently shared twitter posts in which individuals have mentioned the selected corporation. The stock predictions and accompanying tweet sentiment are served to users through a website. Our front-end website was developed using Vue.js to give our team an opportunity to learn a new web framework and gain experience using it. Because of the availability of various libraries such as LIBSVM and Tweepy that facilitate data collection and machine learning algorithm execution, Python was the primary programming language used in creating our back-end that carried out stock price prediction and tweet sentiment analysis. To facilitate data transfer between the front and back ends, we took advantage of Python's micro web framework, Flask, using it to serve our predicted data to web ports the front-end could easily make requests to. After tweet sentiment analysis and stock price prediction for the chosen corporation occur, the computed results are displayed to the user on the web page which include a tweet and a graphical representation of the predicted stock prices for the chosen company.

### **2 Data Collection and Feature Generation:**

All our stock data is collected at runtime by using the Alpha Vantage platform. It offers free APIs in JSON as well as CSV formats for real time as well as historical stocks. An instance of what incoming data looks like is shown in Figure 1 below:

```

{
  "Meta Data": {
    "1. Information": "Intraday (5min) open, high, low, close prices and volume",
    "2. Symbol": "MSFT",
    "3. Last Refreshed": "2019-05-03 16:00:00",
    "4. Interval": "5min",
    "5. Output Size": "Compact",
    "6. Time Zone": "US/Eastern"
  },
  "Time Series (5min)": {
    "2019-05-03 16:00:00": {
      "1. open": "128.9200",
      "2. high": "129.0300",
      "3. low": "128.8000",
      "4. close": "128.8900",
      "5. volume": "1163055"
    },
    "2019-05-03 15:55:00": {
      "1. open": "128.9300",
      "2. high": "128.9700",
      "3. low": "128.8600",
      "4. close": "128.9150",
      "5. volume": "373828"
    },
    "2019-05-03 15:50:00": {
      "1. open": "128.9000",
      "2. high": "128.9500",
      "3. low": "128.8500",
      "4. close": "128.9300",
      "5. volume": "249300"
    },
    "2019-05-03 15:45:00": {
      "1. open": "128.9000",
      "2. high": "128.9200",
      "3. low": "128.8000",
      "4. close": "128.8994",
      "5. volume": "242266"
    }
  },
},

```

```

intraday_5min_MSFT.csv
1 timestamp,open,high,low,close,volume
2 2019-05-03 16:00:00,128.9200,129.0300,128.8000,128.8900,1163055
3 2019-05-03 15:55:00,128.9300,128.9700,128.8600,128.9150,373828
4 2019-05-03 15:50:00,128.9000,128.9500,128.8500,128.9300,249300
5 2019-05-03 15:45:00,128.9000,128.9200,128.8000,128.8994,242266
6 2019-05-03 15:40:00,128.8400,128.9000,128.8100,128.9000,173054
7 2019-05-03 15:35:00,128.8000,128.8900,128.7400,128.8400,186009
8 2019-05-03 15:30:00,128.6300,128.8300,128.6203,128.8000,208237
9 2019-05-03 15:25:00,128.7050,128.7300,128.5800,128.6230,273396
10 2019-05-03 15:20:00,128.7163,128.7500,128.6000,128.7090,164525
11 2019-05-03 15:15:00,128.7600,128.7800,128.6350,128.7200,673897
12 2019-05-03 15:10:00,128.7200,128.8000,128.7150,128.7640,122364
13 2019-05-03 15:05:00,128.7600,128.7600,128.6201,128.7150,500720
14 2019-05-03 15:00:00,128.8200,128.8700,128.7200,128.7600,133755
15 2019-05-03 14:55:00,128.9500,128.9500,128.7100,128.8130,449523
16 2019-05-03 14:50:00,128.9250,129.0200,128.9250,128.9550,134597
17 2019-05-03 14:45:00,128.9750,129.0200,128.8900,128.9300,145036
18 2019-05-03 14:40:00,128.9650,128.9950,128.9125,128.9800,133378
19 2019-05-03 14:35:00,129.0047,129.0500,128.8950,128.9600,384117

```

Figure 1: Alpha Vantage data - examples of company stock data in JSON and CSV formats as taken from Alpha Vantage

### 3 Method:

We used Support Vector Regression as our main metric to predict our stock prices for a chosen date. The team first went on to try other machine learning techniques such as Neural Nets, Support Vector Machines, and Logistic Regression. None of them seem to perform as well as SVR was on stocks. SVR was the optimal choice since we were working with continuous values and SVM is better supported for classification. For this project our goal was not to try and minimise the error rate but instead to fit the error within a certain threshold. So once the data is retrieved from Alpha Vantage it is then trained on an SVR model which is then tested to see what the predicted value should be for the chosen date.

$$Y = wx + b$$

$$\text{Solution: } \min \frac{1}{2} \|w\|^2 \text{ and}$$

$$\text{Constraints: } y_i wx_i - b \leq \varepsilon, \quad wx_i + b - y_i \leq \varepsilon$$

$$\text{Since NonLinear:- } y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) * \langle \phi(x_i), \phi(x) \rangle + b \rightarrow y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) * K(x_i, x) + b$$

### 4 Feature selection:

Not only did we utilize our main method to predict data for a chosen date in the next week, we also added a bias that is introduced by content shared on twitter for the chosen company. If most of the tweets shared are positive then the overall sentiment for the chosen company would be set to positive and vice versa. This sentiment would play a small role in the overall calculation of the predicted price. This is because, if most reviews about any given company are mostly negative, the stock prices are sure to drop in the coming days.

### 5 Results:

The results outputted by our project's Python back-end consist of a predicted price for the selected company's stock, the pulled tweet and a categorization of its sentiment as positive or negative, and the company's stock prices from previous days for graphical representation. As shown in Figure 2 below, our front-end currently displays all of the above back-end outputs. Verification of our project's price prediction accuracy is a little difficult to measure as doing so requires waiting several days to compare our predictions against the actual stock prices as the dates specified for prediction pass.

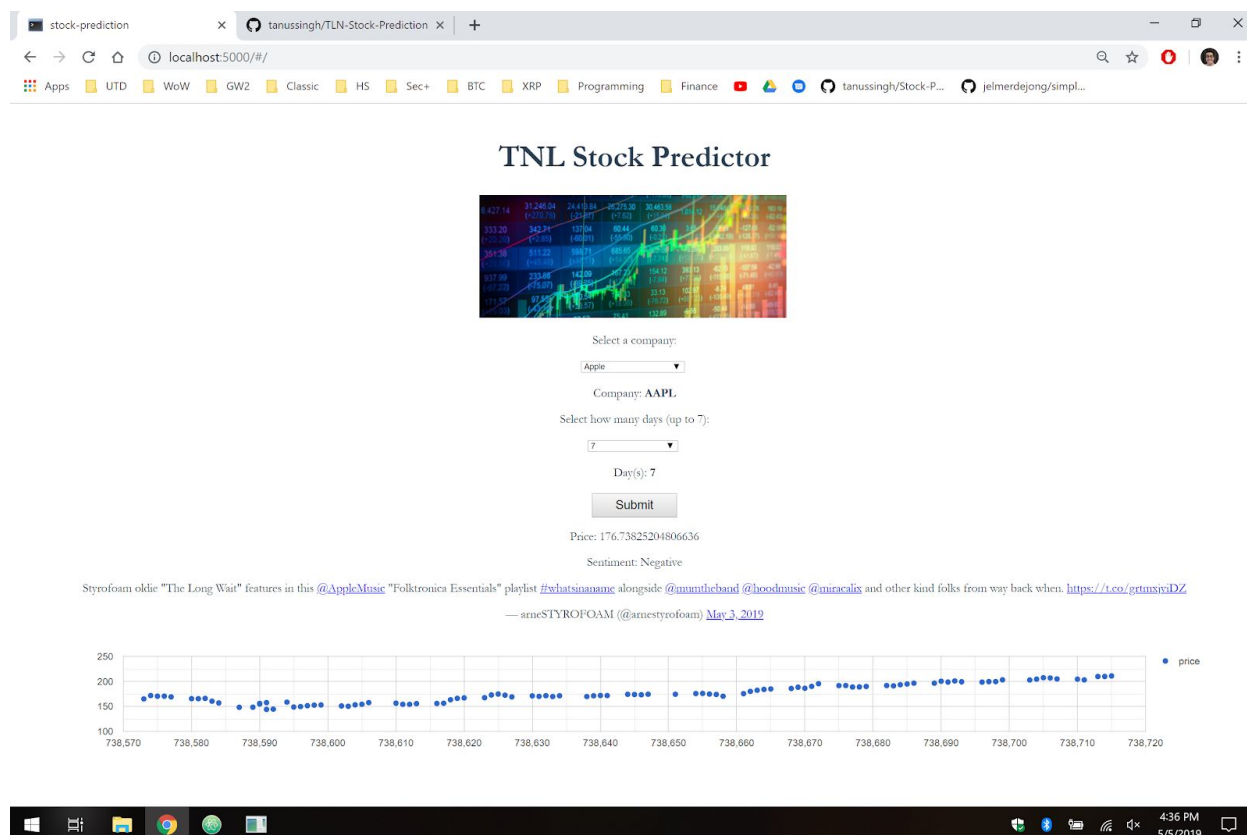


Figure 2: Front-end display of prediction results - Above we can see the selection of Microsoft and a four day prediction returning a stock price, tweet sentiment, the corresponding tweet for the output, and a graph.

## 6 Accuracy:

For the accuracy, we will test the prices that we predicted and compare them to their actual prices. As shown below, the accuracies are well over 98% for most cases.

	Apple	Microsoft	Google	Capital One
5/2/2019 (predict)	\$210.84	\$128.08	\$1181.72	\$92.76
5/2/2019 (actual)	\$210.49	\$126.77	\$1161.35	\$92.81
Accuracy	99.8%	98.7%	98.3%	99.9%

5/3/2019 (predict)	\$210.79	\$127.26	\$1173.55	\$93.12
5/3/2019 (actual)	\$211.32	\$128.57	\$1178.31	\$93.74
Accuracy	99.7%	98.9%	99.5%	99.3%
5/6/2019 (predict)	\$204.38	\$126.49	\$1166.36	\$92.15
5/6/2019 (actual)	\$206.93	\$127.42	\$1182.89	\$93.11
Accuracy	98.7%	99.3%	98.6%	99%
5/7/2019 (predict)	\$205.78	\$126.36	\$1180.37	\$92.31
5/7/2019 (actual)	\$204.67	\$125.32	\$1173.17	\$91.74
Accuracy	99.5%	99.2%	99.4%	99.4%

## 7 Challenges:

Some of the challenges we faced were during the development of the front end code. As all of us had little to no experience with front end development, this was a great opportunity to learn front end development. We used Vue, which is Google's front end development library for JavaScript. Whilst using Vue, we faced many issues; these include writing the components, building the website, and configuring the front end with the back end.

## 8 Future Work:

We would love to add additional companies to the current roster that the user can choose from since that would make the project more widely applicable to comparing the changes of different company stock prices, possibly attracting a wider array of users. If we are able to get a domain and properly host the website, it can potentially be ventured upon by new users who are attempting to learn stock trends and have ambitions to invest in the sector. Given extra development time to expand the scope of data and

algorithms used, the project can most certainly be extremely useful for daily users as well as franchises.

## **9 Acknowledgments:**

We acknowledge Professor Anjum Chida in our project for guiding us and answering any questions and queries that we had. We would also like to give credit to Twitter's Developer Platform and the Alpha Vantage Platform for having a simple and easily inclusive platform which helped us further develop the scope of our project. We also acknowledge C.-C. Chang and C.-J. Li since we have used LIBSVM while coming up with the optimal model.