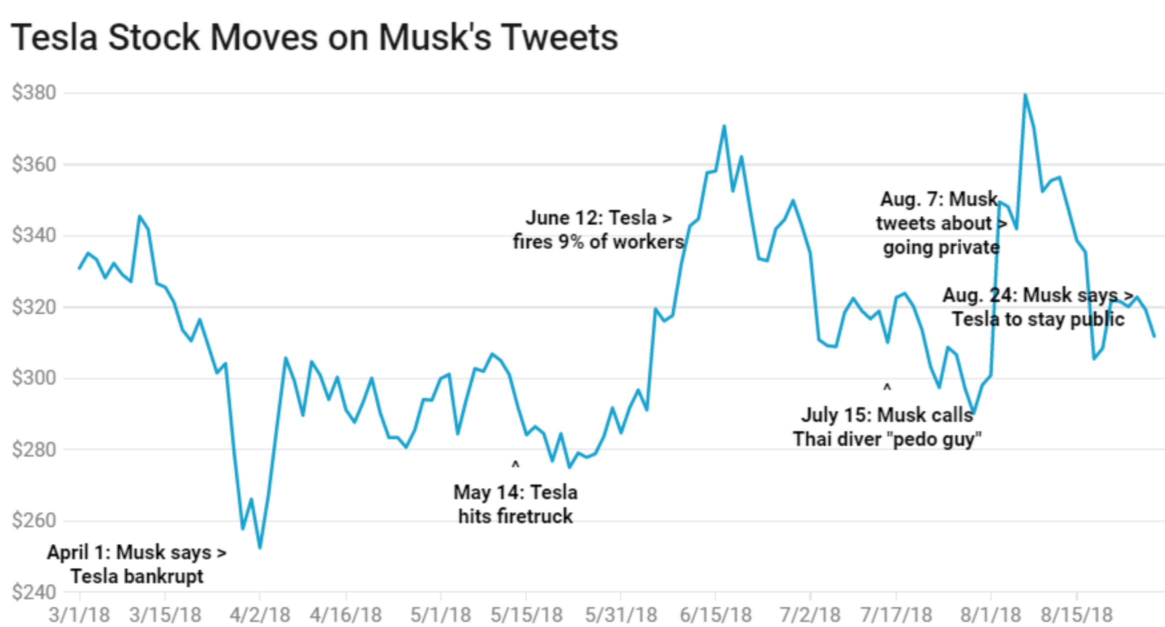
Sentiment Analysis of Elon Musk's tweets and its impact on Tesla's Stock Price

Introduction:

In an era where social media is increasingly reflecting and influencing behavior of other complex systems, behavioral finance has proved that financial decisions are driven by emotion and mood of the investors or people (Bollen, Mao, & Zeng, 2011) . Baker used the public opinion(using twitter tweets) and proved that there was a correlation between the moods of public expressed on Twitter and the way the stock market performs (Hasan 2019). Our project was inspired by the "theory" that Tesla stock price fluctuates to the tone of Elon Musk's Tweets. As data analytics students we wanted to put analytics and numbers to this theory to see if there is a correlation between the two. Elon Musk is the CEO of an electric car company, Tesla. We chose Elon Musk because he is the second most active CEO on Twitter in terms of the number of tweets, with more than 1000 tweets in a year on an average. The stock price of Tesla has seen a huge rise in the past after Elon Musk updated his twitter account about the self-driving cars and also, the Dow Jones and S&P indexes have dropped by almost 1% after the Twitter account of Associated Press falsely posted the message about an explosion in the White House (Hu, Jiao, & Zhu, 2013). Additionally, we aim to compare the stock price fluctuation of the Tesla stocks with the rest of the automotive industry to check for a similar trend. The aim of this project is to check for correlation between the Tesla stock price fluctuation and Elon Musk's Tweets and to further study stock price trend in the automotive industry to look for other possible trends that might be at play.

Background and Motivation:

There have been articles published about how some of Elon Musk's tweets have impacted the Tesla share prices. Below are a few examples:



Tesla Stock Sank With Elon Musk's Tweet Now Under Criminal Investigation

The $40 million tweet: Elon Musk settles with SEC, Tesla bears the brunt

A tweet sent by Musk suggesting Tesla would go private led to the charge of securities fraud.

Tweets by a CEO impacting the stock market was unheard of in the past and Mr Musk is known for the impact that his tweets have. Given the popular belief and our admiration for the electric car industry, we aim to find out the correlation between his daily tweets and the Tesla stock prices over a period of 7 years (from 2012- 2019). Tesla's share price has gone up from $20 to $260 since it's gone public (more than a 1000% increase). We are aware of the complex and dynamic environment that the stock market operates in and that there are several factors that are at play in determining the stock price movement. This project aims to explore the factor of social media sentiment in case of Tesla Motors.

Literature review

Study of various papers was carried out during the proposal making of this project. "Twitter mood predicts the stock market", (Bollen et al., 2011) . "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", (Pagolu et al., 2016). Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States (Kumar and Chong, 2018) , "Using Tweets to Predict the Stock Market", (Hu, Jiao, & Zhu, 2013). "The Effects of Twitter Sentiment on Stock Price Returns", (Gabrielle Ranco, 2015). "The Effects of Twitter Sentiment on Stock Price Returns", (Tahir Nasir, 2018). "Stock Prediction Using Twitter", (Hasan, 2019).

1. The first article is about Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. The company that they choose for this analysis is Microsoft. This paper uses applied sentiment analysis and supervised machine learning principles to the tweets extracted from twitter to analyze the correlation between stock market movements of a company and public opinion/sentiments as expressed on twitter. Opening and closing stock prices were used to classify the fluctuation in the stock price as an increase or decrease. If the opening stock price is more than the previous day's closing stock price, then it is classified as "0" otherwise it is classified as "1". Word2vec and Ngram, are two textual representations that are used to carry out the sentiment analysis. 355 instances with 3 attributes are used to classify. The tweets are classified into three categories: positive, negative and neutral. Experiments are conducted to determine the best window size of the stock price; best results are achieved when the sentiment value precedes 3 days to the stock price. Various models like random forest, logistic regression

and LibSVM is used as sentiment analyzer and correlation analyzer. For stock price and sentiment correlation, LibSVM gave a result of 71.82% with 90% of the data as training data.

This paper concludes that there is a strong correlation between stock price fluctuations of a company to the public opinions on the same as expressed on twitter. The study also concludes that with increase in the training data the model tends to perform better.

2. The second article aims at checking for correlation and predictive property of public's mood economic indicators. Mood tracking tools, namely Opinion Finder that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) are used for this analysis. This paper hopes to assess the effects of including public mood information on the accuracy of a “baseline” prediction model. A time series of daily DJIA (Dow Jones Industrial Average, is a stock market index that measures the performance of the top 30 stocks) using the closing-values from Yahoo! Finance. The correlation between DJIA and the mood trackers is checked for by using Granger causality analysis for past n days. A Self- Organizing Fuzzy Neural Network model is used to test the hypothesis that the prediction accuracy of DJIA prediction models can be improved by including measurements of public mood like positive, happy, calm, alert, sure, vital and kind. The Calm mood dimension shows a predictive value. OpinionFinder’s assessment, on the other hand, is not Granger causative of the DJIA. Twitter data for a year was taken to carry out this study. Direction of DJIA movement was predicted with an accuracy of 87.6%.

3. The third paper we read was more related to our project and was titled “Using Tweets to Predict the Stock Market”. In this project they intended to find the relationship between tweets of Elon Musk and the change of Tesla stock prices. There were a few trends to suggest that there might exist a relationship between these two. Tesla share prices saw a huge rise in price after Elon Musk updated his twitter about the self-driving vehicles also, the stock market indexes dropped after Associated Press falsely tweeted about an explosion in the white house. The content of the tweets data for this project was obtained from tweepy along with “favorites” and “number of retweets”. The financial data of Tesla’s stock prices was obtained from Yahoo finance and was labeled in a couple of different ways: -1 or 1 to denote increase or decrease in price in a trading day and this feature is used to construct a label. The stock prices are matched with the tweets of the same day. An SVM model with Gaussian kernel was used in the project. 3 groups of features were tried and evaluated, and all the features and labels were trained using SVM. Firstly, a dictionary is constructed according to the content of all the tweets. Secondly, an upper threshold of frequency Fh and a lower threshold Fl of frequency to consider only the medium frequency words is set. Thirdly, the “Indication Factor” $P(x=i|y=1)$ $p(x=i|y=-1)$ is computed and k words with largest “Indication Factors” are selected according to the labeling vector Y. Then, the vector v is established as vi to be the number of ith largest words appears in the tweet. The number of “favorites” and number of “retweets” of one day are averaged and the word vectors are aggregated. The next set of features used are sentiment scores of tweets and stock price movements. The tweets are labeled positive, negative or neutral to determine the sentiment scores. Lastly, time processing is performed and the number of days it takes for the tweet to impact the stock prices is calculated. This was the most reasonable approach as it had both reasonable accuracy and reasonable confusion matrix.

Approach

In this project we used supervised learning i.e. Random forest classifier. It is the best classifier to be used for investments with low correlation like shares, the correlation increases for a portfolio when compared to individual share values, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. Hence the correlation coefficient and precision of the model will be very high (Tony Y., 2019). The project aims at establishing a predictive model. We also studied the correlation between Tesla share prices and its competitors' share prices (Ford Motors and General Motors) to take a look at one of the external factors that might be at play in the stock price movement.

Data:

To conduct this analysis, we collected Elon Musk's tweets using Twitter API from 2012 to 2019. A total of 8863 tweets were collected. Tweets are not periodic in nature. This dataset comprised of the tweet content, date, time of tweets and retweets. Additionally, the share prices of Tesla over the same 7-year period was obtained from Yahoo! finance. This dataset consists of Date, opening price of the share, High (the maximum price of the share during the day), Low(the minimum price of the share during the day) and the closing price of the share along with the Total Volume of the shares traded during the day(includes both the number of shares sold and bought). This data is periodic in nature.

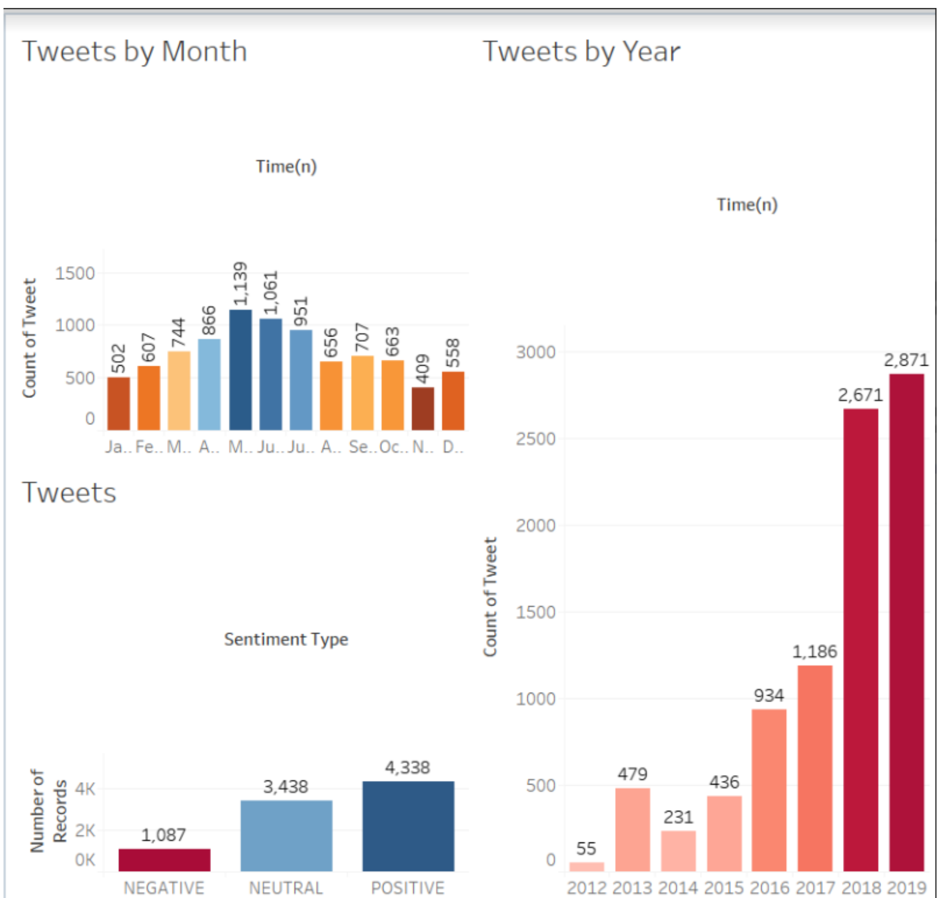


Fig 1 : Exploratory analysis of Elon Musk's tweeting habits from 2012-2019

Data Cleaning and Preparation

Combining Tweets and Stock Prices Data:

As mentioned above both the tweets dataset and the share price dataset are obtained from 2 different sources. Hence, the first step was to combine both these datasets into one comprehensive dataset. We combined both these datasets based on the “Date” label which is common between both datasets.

Missing Data:

Since the stock market does not operate on the weekends, the values of the stock price on Saturday and Sunday and on public holidays is missing. These were eliminated from the analysis.

Sentiment Analysis:

Rule based system of sentimental analysis was carried out on the collected tweets data. We used the Natural Language Toolkit (NLTK) in Python to perform sentiment analysis. The NLTK consists of a “Stopwords” package which consists of the lists of positive and negative words predefined. The code runs through every tweet and identifies the positive, negative and neutral words and assigns a score to each tweet and classifies it as a positive/neutral/negative tweet based on the scores. The output of the same is explained in Table 1:

Tweet	sentiment_neutral	sentiment_negative	sentiment_pos	sentiment_type
meltingice assuming max acceleration g com	0.655	0	0.345	POSITIVE
rt spacex bfr capable transporting satellites c	0.833	0	0.167	POSITIVE
bigajm yup	1	0	0	NEUTRAL
part https co fvu muhm	1	0	0	NEUTRAL
fly places earth mins anywhere cost per seat	1	0	0	NEUTRAL
rt spacex supporting creation permanent self	0.75	0	0.25	POSITIVE
bfr take anywhere earth less mins https co h	0.877	0	0.123	POSITIVE
mars city opposite earth dawn dusk sky blue	1	0	0	NEUTRAL
moon base alpha https co voy qew kl	1	0	0	NEUTRAL

Table.1

Below are the steps involved in implementing this algorithm.

a) Tokenization: Preprocessing of the data to remove special characters, stop words and nonSemitic words and arrange the words into an array
b) Negation handling: If a Token is preceded by a negator, its valence is flipped
c) Classification of the sentiment: Negative,Positive and Neutral.

Output of this analysis is a CSV file. Each tweet instance will correspond to a sentiment analysis value of “negative”, “positive” or “neutral”. Each tweet has a sentiment score. These scores were used to subdivide the sentiment categories as mentioned below.

High positive(HIGHPOS): sentiment_positive >= 0.3
Low positive(LOWPOS): sentiment_positive <0.3
High negative(HIGHNEG): sentiment_negative >= 0.3
Low negative(LOWNEG).: sentiment_negative < 0.3

We chose 0.3 as the cutoff after reading a few examples of sentiment analysis online which suggested that positive or negative scores of >0.3 makes it a high positive/high negative tweet respectively. This made it a five class attribute.

Stock Impact Calculation:

A tweet might not have an impact immediately on the stock market. The time that it takes to show as an impact on the stock prices is considered as three days from the day of tweet. So, for a tweet on the nth day we considered the stock prices of nth+3 day. The issue arises when we want to calculate impact for tweets tweeted on Wednesday and Thursday since n+3 th day will be Saturday and Sunday respectively. and therefor, for Wednesday we considered the impact day as Friday and for Thursday we considered the impact day as Monday. The impact is calculated as the difference between the opening price of the stock on the impact day and the closing price of the stock on the previous day. This gave us 3 outcomes, a rise in the stock price, fall or no change. Like the sentiment analysis, we converted this 3-class attribute into a 5-class attribute.

For stock price impact >= $3: HIGHUP,
For stock price impact < $3 & >$0: LOWUP,
For stock price impact = $0: NOCHANGE,
For stock price impact < $0 & > -$3: LOWDOWN, and
For stock price impact < -$3: LOWDOWN

Classification:

After performing the above steps, we had 2 nominal attributes with 5 classes each. As mentioned already we used Random Forest algorithm to classify the tweets and predict the stock price impact. We used 10-fold-cross-validation to split the original dataset into a training and a test dataset. The metric that we wanted to use to evaluate the model is accuracy, recall and true positives. We would also like to minimize the False Positive. If this model is used to make investments and financial decisions based on the prediction of stock price increase/decrease, minimizing the FP will minimize the cost of wrong decision making.

Results

Weka was used to run Random Forest on the prepared data. Output of this model is represented in table.2 and table.3

Correctly Classified Instances 3148	35.52%
Incorrectly Classified Instances 5715	64.48%
Kappa statistic	0.0015
Mean absolute error	0.287
Root mean squared error	0.3789
Relative absolute error	99.98%
Root relative squared error	100.02%

Tabel.1

a	b	c	d	e	<-- classified as
3039	1100	0	0	0	| a = LOWUP
2789	1090	0	0	0	| b = LOWDOWN
1444	670	0	0	0	I c = HIGHUP
1220	410	0	0	0	| d = HIGHDOWN
39	5	0	0	0	| e = NOCHANGE

Table.3

As we see in the confusion matrix the model does not learn high up, high down and no change class. To investigate this further, statistical study of the data is conducted. Table.4 shows the class distribution and the input feature distribution. Both, the class distribution and the sentiment type input feature is skewed. Outliers adversely affect the model's performance especially regression-based models.

TESLA Stock price impact-5 classes		Sentiment_type-5classes	
LOWUP	3150	HIGHPOS	2339
LOWDOWN	2899	LOWPOS	2001
HIGHUP	1512	NEUTRAL	3439
HIGHDOWN	1262	HIGHNEG	366
NOCHANGE	45	LOWNEG	723
TOTAL	8868	TOTAL	8868

Table.4

As the next step we carried out correlation attribute evaluation in Weka. This evaluates the worth of an attribute by measuring the correlation (Pearson's) between it and the class. Nominal attributes are considered on a value by value basis by treating each value as an indicator. An overall correlation for a nominal attribute is arrived at via a weighted average. The correlation coefficient for each of the attribute classes against Tesla stock impact is as shown in Table.5. As the numbers suggest there is little correlation between the attributes.

0.01604	3 HIGHPOS
0.01354	4 LOWPOS
0.01321	2 HIGHNEG
0.00331	5 LOWNEG

Table.5

As a process of reducing the shrewdness of the data, the number of classes of the class attribute is reduced to a two class attribute; i.e; "Up" and "Down". The input feature is also reduced to three class attribute of, Positive, Negative and Neutral. Random Forest and Multilayer Perceptron is used to run the prediction model.

The final outputs of Random Forest and Multilayer Perceptron are as below:

Random Forest:

Correctly Classified Instances	4660
Incorrectly Classified Instances	4203
Kappa statistic	0
Mean absolute error	0.4988
Root mean squared error	0.4995
Relative absolute error	100.03%
Root relative squared error	100.03%
Total Number of Instances	8863

a	b	<-- classified as
4660	0	I a = UP
4203	0	I b = DOWN

Table.6

Multilayer Perceptron:

Correctly Classified Instances	4614	52.06%
Incorrectly Classified Instances	4249	47.94%
Kappa statistic	-0.0001	
Mean absolute error	0.4986	
Root mean squared error	0.5004	
Relative absolute error	99.99%	
Root relative squared error	100.21%	
Total Number of Instances	8863	

a	b	<-- classified as
4660	466	I a = UP
4203	420	I b = DOWN

Table.7

Correlation between Tesla share prices and Ford, GM share prices

Weka o/p:

Ranked attributes:

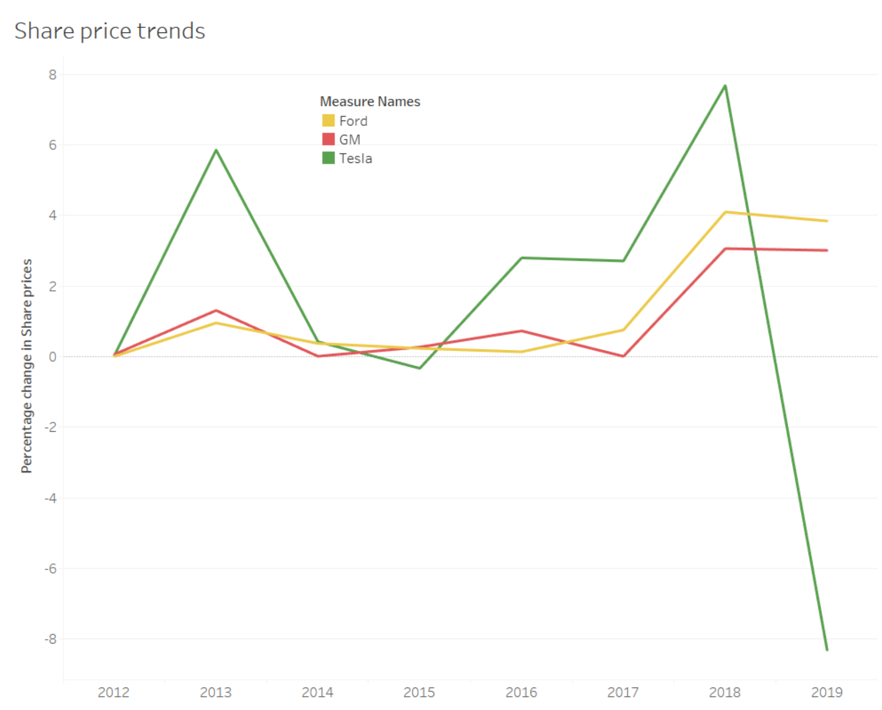
0.212	1 GM Stock Price Impact
0.205	2 FORD Stock Price Impac

Table.8

We ran the correlation of GM motors and Ford share prices with Tesla prices in Weka and we plotted the percentage of impact in share prices for Tesla, GM motors and Ford motors from 2012 to 2019 in Tableau. As seen from both the Weka output and the tableau visualization created using our dataset, there is a very low correlation between Tesla and its competitors share prices. GM motors and Ford are correlated to each other but there is no correlation of both these share prices with respect to Tesla.

.

Tableau visualization:



Discussion

We used rule-based sentiment analysis. This clearly has its limitations, especially in a domain specific application like the one that we are using in this project. To check for effectiveness of the NLTK sentiment analysis on tweets we carried out manual analysis of 600 tweets to check for relevance and effectiveness. 600 tweets that were manually analyzed are assumed to be the ground reality for the case of this study. NLTK output for the same 600 tweets have a 70% error rate when compared to the ground reality. Therefore, we see that there is a scope for improvement in the method of sentiment analysis of these tweets. This can also be one of the contributing factors to low correlation to the stock price impact.
As we see in the results section, Random Forest algorithm classified only LOWUP and LOWDOWN and the confusion matrix shown above was far from satisfactory. The model couldn't predict the stock impact with just stock sentiment as an input. Single feature-based models could be insufficient to learn about the stock impact. Single features may be useless until used in combination. Some phenomena can be described only in multi-dimensional space.
The stock price movement is a complex phenomenon. Plenty of factors affect the stock prices.The main reason behind the poor performance of the Random Forest model was the skewed distribution of the sentiment type. Multilayer Perceptron performed equally if not worse when compared to Random Forest with a precision rate of 34.18%, it was facing the same issue as Random Forest.
Hence, we decided to investigate this further and found out the correlation between various sentiment types and the Tesla stock impact and found that they had a very poor correlation. Thereby, we decided to reduce the number of classes in Stock impact to two classes; UP and DOWN and in the input feature Sentiment type to three classes; POS, NEG, and NEUTRAL.

This increased the precision of the model to 52.57%, which was a significant increase, given that the initial precision rate was as low as 38%. However, the model performance was still bad as the confusion matrix above suggests. It was classifying all the sentiments into "UP" stock impact class. The RMSE increased to ~0.5 which was not a big increase and was still a good value.
However, when we ran Multilayer Perceptron, although the precision didn't increase, the confusion matrix looked much better and the model was classifying both the "UP" and "DOWN" impacts. Hence, we decided to use this model for the cleaned data. The accuracy of our model is low because of the weak correlation between the input variables and the class variable.

Conclusion and Future work

All the tweets by Elon Musk's do not have an impact on Tesla Share prices. His tweets are about a wide range of topics varying from politics to his other companies, retweets and replies to his fans. Sentiment analysis using NLTK is not a very effective tool to identify and measure the magnitude of impact his tweets have on Tesla share prices. There is very low correlation between the sentiments of his tweets and the Tesla share impact. However, as shown before some of his tweets do have a significant impact on Tesla share prices. These tweets are few and infrequent. But still hold a scope for future work.

Future work: To overcome the limitations of a single feature training model we'd like to Identify and add some important attributes such as number of likes, number of comments, number of retweets and the demographics of people reacting to his tweets. Additionally, there is a scope for accounting the emoji sentiment value to improve the overall sentiment score representing the real world.

References

Bollen, J., & Mao, H. (2011). Twitter Mood as a Stock Market Predictor. Computer, 44 (10), 91-94. doi:10.1109/mc.2011.323

Domeniconi, G., Moro, G., Pagliarani, A., & Pasolini, R. (2017). Learning to Predict the Stock Market Dow Jones Index Detecting and Mining Relevant Tweets. Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management.

Firth, M. (1976). Share Price Behaviour. Managerial Finance, 2 (3), 294-318.

Hasan, K. S. (2019, January 03). Stock Prediction Using Twitter. Retrieved from https://towardsdatascience.com/stock-prediction-using-twitter-e432b35e14bd

Nisar, T. M., & Yeung, M. (2018, February 09). Twitter as a tool for forecasting stock market movements: A short-window event study. Retrieved from https://www.sciencedirect.com/science/article/pii/S2405918817300247

Nisar, T. M., & Yeung, M. (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. The Journal of Finance and Data Science, 4 (2)

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) . doi:10.1109/scopes.2016.7955659

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. Plos One, 10 (9). doi:10.1371/journal.pone.0138441

Roslan, M. A., & Rahiman, M. H. (2018). Stock Prediction Using Sentiment Analysis in Twitter for Day Trader. 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC)

Baker HK, Kent Baker H and Nofsinger JR (2011) Behavioral Finance: An Overview. Behavioral Finance . DOI: 10.1002/9781118258415.ch1 .

Bollen J, Mao H and Zeng X (2011) Twitter mood predicts the stock market. Journal of Computational Science . DOI: 10.1016/j.jocs.2010.12.007 .

Kumar S and Chong I (2018) Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States. International journal of environmental research and public health 15(12). DOI: 10.3390/ijerph15122907 .

Pagolu VS, Reddy KN, Panda G, et al. (2016) Sentiment analysis of Twitter data for predicting stock market movements. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) . DOI:10.1109/scopes.2016.7955659

Hasan, K. S. B. (2019, January 3). Stock Prediction Using Twitter. Retrieved December 12, 2019, from Medium website: https://towardsdatascience.com/stock-prediction-using-twitter-e432b35e14bd

Hu, Z., Jiao, J., & Zhu, J. (2013). *Using Tweets to Predict the Stock Market*. Retrieved from https://pdfs.semanticscholar.org/39f0/33a4797ef3726dd5527d740cbdc6fd9936c9.pdf