Tesla Inc. Stock Prediction using Sentiment Analysis

Amey Bhadamkar[1] and Sonali Bhattacharya[2]

Abstract

The stock market is a very volatile component of the financial domain. Accurate predictions of various stocks are a highly active area of research and analysis. Following the previous ML prediction techniques using Artificial Neural networks and fuzzy-based techniques, this research aims to extend the accurate prediction results. Since multiple qualitative factors go into the decision-making of a buy-sell of stock, a blend of algorithmic trading is at the cornerstone of the research. This research work aims to look into the unique relationship between Elon Musk's Tweets and Tesla's stock value. Exploratory Data Analysis was employed as the primary analysis method to better differentiate patterns within our dataset, which had been pre-processed to remove any stop words. Combining these methodologies and elements yielded a decisive conclusion with a clear correlation: an increase in the number of tweets/engagements corresponded to an increase in Tesla's closing price and vice versa.

Keywords: Tesla, Elon Musk, Stock Market, Sentiment Analysis, Machine Learning

JEL Code: G17

[1] *amey_bhadamkar@scmhrd.edu, Symbiosis Centre for Management & Human Resource Development, Symbiosis International (Deemed University) SIU , India*

[2] *sonali_bhattacharya@scmhrd.edu, Symbiosis Centre for Management & Human Resource Development, Symbiosis International (Deemed University) SIU , India*

1. Introduction

The American Stock Market is a highly complex system, where huge chunks and volumes of information. Data is generated instantaneously and constant change in small proportions with different factors and diversity. The ownership of companies is divided into the shareholders, who are the real owners of a company. The shareholders are the owners of the company's shares that are offered in public listed companies. The stocks symbolize the company ownership in parts and are the stakeholders in the company's profits and losses.

Regarding the stock markets, where the Stocks are traded, Efficient Market Hypothesis is a theory that proposes the patterns in stock market prices are random and dependent on the newly available information. The research aims to challenge this existing hypothesis and detect patterns and trends in Stock Market using various factors, like Sentiment Analysis and Time-Series Forecasting. The research paper contains the analysis using advanced Machine Learning algorithms implemented in the Python environment. The models built in here will be validated against the actual trends in the stock market.

Because it is influenced by numerous elements such as market attitudes, financial operations, political events, rumors, and news, and company dealings, the stock market is non-linear, non-continuous, and volatile. Making profits is possible in 2 ways when the stocks are bought at a lower price and sold at a higher price when the Market has gained value in the particular share. Another way of making a profit is by selling the unpurchased stock in the Market and buying it when the Market is going down, but this is restricted to a mode of trading called Intraday. The real challenge here lies in the prediction of the stocks market trends. To minimize the loss is also an optimization target to succeed in the stock market.

Predicting the future prices of stocks is not a humanly possible task hence the usage of high-end Machine Learning algorithms and forecasting techniques are used. While theorists say that the prices cannot be predicted, advances in Sentiment analysis using social media are gaining much traction and gaining relevant successful results with decent model accuracy.

Elon Musk, known for his extraordinary wealth, became the wealthiest person on the planet in 2020. Elon Musk founded x.com in 1999, which subsequently became eBay, Space X in 2002, and Tesla in 2003.

Despite all of Elon Musk's accolades, his most well-known endeavour is his success with Tesla. Tesla, which primarily manufactures electric automobiles and energy storage solutions, is the world's fastest-growing car brand.

Musk, on the other hand, was not the one who established Tesla. Instead, Martin Eberhard and Marc Tarpenning created the company in 2003, naming it after Nikola Tesla, a Serbian-American inventor.

Musk became the chairman in 2004 after investing more than $30 million in the company. Since then, Tesla is regarded as a corporation with limitless potential in terms of technological developments, with its stock skyrocketing under Elon Musk's leadership.

Elon Musk's achievements are astonishing, but he has been chastised for his frequent use of Twitter, regarded as disrespectful and unprofessional. New research suggests, however, that when used effectively, Twitter may be an effective marketing tool.

With the advent of Twitter as a popular social media tool in recent years, correlations between Twitter tweets and stock returns have been apparent. Elodie Michaux, for example, published research in 2019 that found a correlation between stock returns and tweets, even though she was unable to establish any association between tweet count and stock trading volume, indicating that tweets can affect stock prices. Furthermore, ScienceDirect, a respected science journal, presents several studies that achieved identical outcomes, implying that good sentiment in tweets and Twitter coverage will inevitably improve stock values.

It becomes clear from such study in multiple search engines and libraries that there is an indisputable link between the financial Market and Twitter. As a result, this research study aims to elaborate on this topic, focusing on Elon Musk's tweets and Tesla, Inc. stock returns. Furthermore, thanks to the advent of Sentiment Analysis, which extracts subjective emotion from text and words. Our research can comprehend the personal emotions behind tweets and replies and associate them with Tesla's stock price.

2. Literature Review

Literature Review may lead to biased inference due to various limitations of the study as the scope was restricted to certain entities. The research also has a more extensive future scope than the findings in the paper. Many insights and conclusions can be drawn from the other International markets; this paper only focuses on China’s stock market. E- Research is based very much on the Indian Market and LSTM techniques. The scope for neural networks and Deep Learning algorithms remains at a stretch. The analysis done for S&P 500 data was practical with Regression but should have been tested on the Neural Networks or Deep Learning Algorithms.

The dynamic Market has to be understood through delicate nuances of the stock market through various domains. The effects of interrelated domains and the changes in them cause an effect on the stock market prices. This Research Paper focuses on the essence of improving the existing models for Stock Market Shares price prediction. This will be facilitated by extensive data from Social Media via Sentiment Analysis. Sentiment Analysis on Social media has a strong correlation with the fluctuation of the prices. It uses Algorithmic trading and Deep Learning Algorithms, Artificial Neural Networks, and Fuzzy Logic as the research's cornerstone.

Based on the various Classification and Regression models used for accurate model predictions, LSTM, Artificial Neural networks, and Deep Learning techniques, they are further augmented with public sentiments from Social Media in addition to SM prices (Mehtab, Sidra and Sen, Jaydi 2019). According to (Anshul Mittal, Arpit Goel 2017), the Sentiment Analysis can be structured into four steps Methodology using Word lists, Tweet Filtering, Daily scoring, and Score Mapping.

In addition, the Sentiment Classification Methodology uses the ML approach and Lexicon Approach by (Aditya Bharadwaj, Yogendra Narayan, Vanraj Pawan m Maitreyee Dutta 2018). Along with the traditional methods of prediction modeling, a robust set of algorithms are proposed by (Seng, and Yang, 2017) wherein LSTM, Hybrid Models using SVM, KNN, and Random Forest is used in a Majority Voting Algorithm. The Taiwanese Stock Market (Seng, and Yang, 2017) has proposed an algorithm to calculate the sentiment orientation and score of the data with added information. The results have been integrated to construct an empirical model for calculating the Stock Market volatility. The proposed framework by (Kalampokis, et al., 2013) in the Greece Research environment has revealed that the relevant set of studies can be decomposed into small steps, and unique approaches can be followed.

Similarly, in the Tehran Stock Market, extensive research suggests a Lead-lag relationship in the stock price changes dynamically by (Sun et al, 2020). The forecasting-based investor behavior is more accurate only when the stock is stably leading the price change most of the time. A more news-driven approach using news-driven sentiment analysis and further augment the volatility models with sentiment factors to measure different types of sentiments in the India Market (Paramanik and Singhal, 2020). Moreover, Since sentiment Analysis plays such an important in the decision-making of investors (Ahmed, 2020) has aimed at understanding the mechanisms that affect the investor's demand for risky assets. Using this data and important decision affecting factors, large unstructured files collected from an online Social Media platform are linked to analyze the trends in the Indian Market by (Bhardwaj et al., 2015). Considering the types of shares and stocks generally traded on are OTC (Over The Counter stocks), behavioral portfolio investors invest in layers corresponding to an aspiration or a goal in mind. Penny stocks have different characteristics and trading behavior than OTC stocks (Nofsinger and Varma, 2014).

3. Methodology

Exploratory Data Analysis, or EDA, is an analytical method for grouping the output of a dataset according to a set of criteria. To put it another way, this kind of analysis looks for patterns in data and categorizes them properly. Based on the versions of data the EDA is divided into 4 types: a. Multivariate non-graphical, b. Multivariate graphical, c. Univariate non-graphical, and d. Univariate graphical EDA is largely used to assure that the data used here is absent of anomalies and outliers, allowing scientists and analysts to work with a more consistent and reliable data set.

3.1 Machine Learning:

Machine Learning (ML) was a primary computational technique used in this study. Simply explained, machine learning (ML) is a computer method in which a processor enhances its accuracy over time by experience and trial-and-error, similar to how a human brain might think. Analysts and data scientists frequently use Machine Learning to create software that can readily improve with new datasets and algorithms.

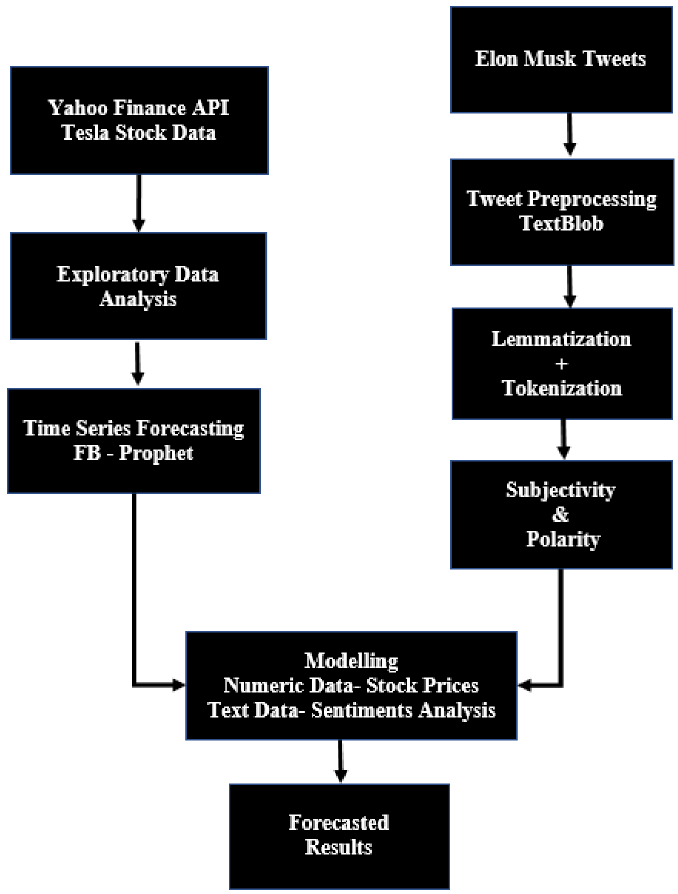

Figure 1. Methodology Flowchart

3.2 Machine Learning Algorithms:

This section discusses several prominent algorithms that are frequently used in Machine Learning. Because the computer "trains" its intelligence using the inputted algorithm, these algorithms are effectively the foundation of the code.

Different algorithms are designed for different sorts of Machine Learning; for example, one algorithm may be better suited for classification and another for prediction. Different approaches will have varying degrees of accuracy; thus, we must understand the physics of these algorithms to determine which algorithms to utilize next.

3.3 Time Series Forecasting:

When dealing with the unforeseen and unknown, there are bounds to be overcome.

Time series forecasting isn't perfect, and it's not appropriate or beneficial in every circumstance. Because there are no strict rules for when to utilize forecasting and when not to, it is up to analysts and data teams to understand the constraints of analysis and what their models can support. Not every model will fit every set of data or provide an answer to every inquiry. When data teams understand the business challenge and have the necessary data and forecasting capabilities to answer it, they should use time series forecasting.

Forecasting that is accurate works with clean, time-stamped data and can spot actual trends and patterns in historical data. Analysts can distinguish between random fluctuations and outliers, as well as determine accurate insights from seasonal variations. Time series research reveals how data evolves, and accurate forecasting can indicate the data's growing trend.

3.4 Sentiment Analysis:

Sentiment Analysis, also known as Opinion Mining, is a branch of humanities and linguistics combined with advanced data science and (AI) artificial intelligence. Sentiment Analysis' basic assumption is to take a set of linguistic data (in this case, Elon Musk's Twitter) and determine the emotions and subjective feelings behind those words. Sentiment Analysis is done in several ways. For example, a machine might learn to discern human emotions by reviewing enormous datasets of text. Sentiment Analysis was utilized because we predicted that emotions expressed in tweets could be linked to Tesla stock prices.

Sentiment Analysis is employed in almost every aspect of our lives. Sentiment Analysis is used in various ways in advertisements, campaigns, and commercials, all of which include pathos. Sentiment Analysis was used extensively in our study, with the subjective state behind Elon Musk's tweets/replies being applied to Tesla's stock, revealing novel correlations between the two.

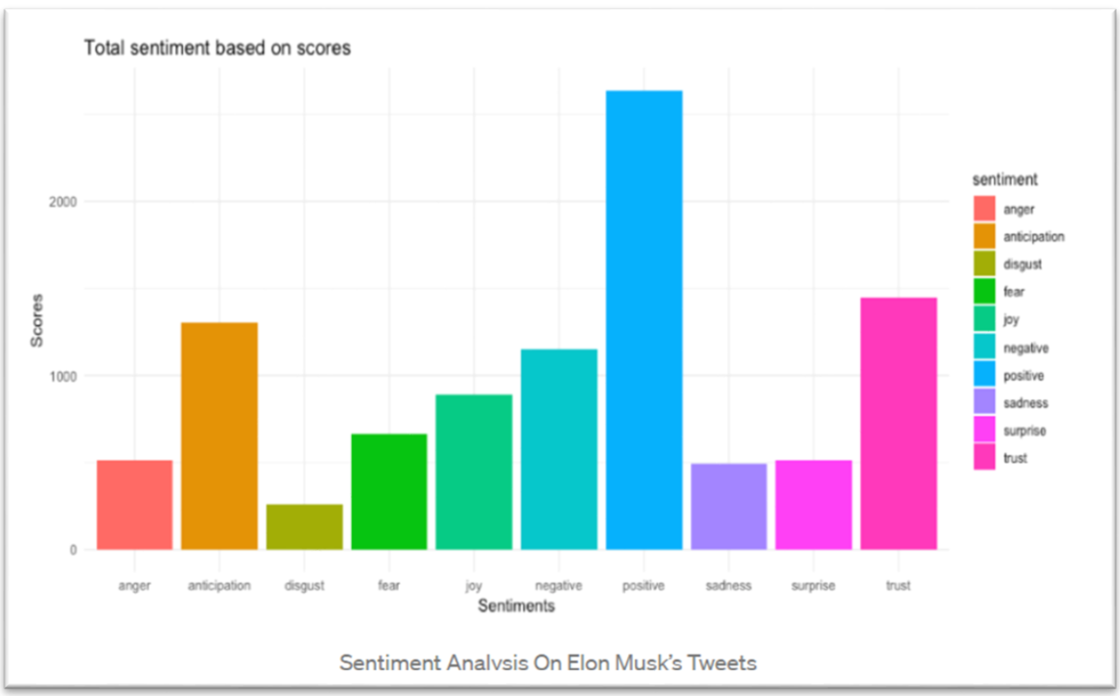

Figure 2. Sentiment Analysis on Elon Musk's Tweets

Each tweet on Twitter was divided into three categories based on its expected emotional language: positive, neutral, and negative. This was accomplished with Python and the Scikit-learn program. This machine learning was carried out using pre-processed tweets from the text column of the CSV file. We were able to link Musk's online sentiments to Tesla's potential stock market changes using such tactics, and we were able to go deeper into our hypothesis.

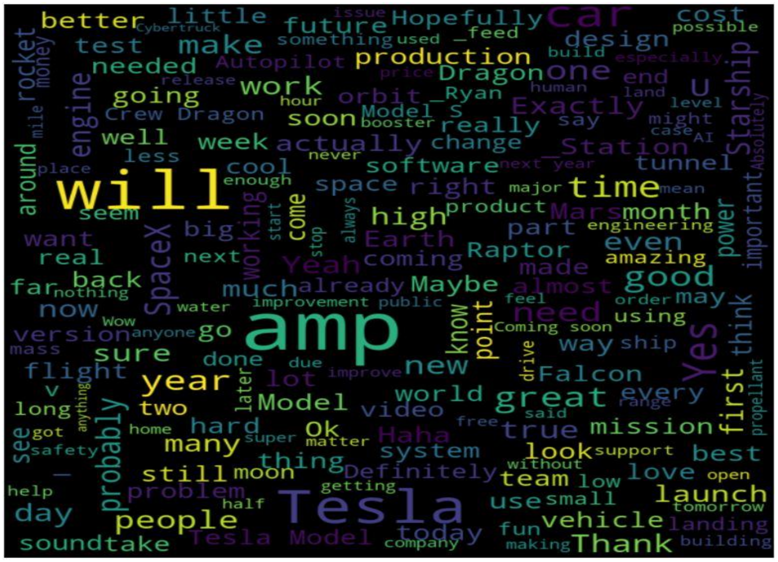

Figure .3 Sentiment Analysis – Word Cloud

4. Findings

4.1 Data Analysis:

The data presented in the research states the stock prices of the company Tesla Inc. Tesla went public on June 29, 2010. Through this IPO, the company had raised the US $226million. The stock was initially listed at US$17.00 per share. In today's date, the stock is listed around US$ 721 per share, with almost 4144% growth in 11 years of going public.

Tesla's stock is listed on the NASDAQ index. Since its inception, the current world's richest person, Elon Musk is the Chief Executive Officer and running its operations across multiple countries. Since taking the leadership of the electric vehicle company in 2008, Musk has defied the odds and routinely exceeded expectations.

Twitter is a sophisticated micro - blogging platform that has changed the way people talk—for better or worse—since its debut. Because of the millions of internationally active users, there is a plenty of data to harvest and use.

Elon Musk, the CEO of Tesla and SpaceX, has been embroiled in controversy as a result of his tweets. Elon Musk's tweets are primarily anchored in positive, trust, and anticipation attitudes. Elon routinely tweets positive things about his life, particularly regarding his businesses, SpaceX and Tesla.

The technique of quantifying and categorizing words in relation to a specific emotion or product is known as sentiment analysis. Stop words are used to eliminate words from the analysis that aren't useful. It was difficult to eliminate all of the stop words because they are the foundation of his tweets. The opportunity to join in on other people's conversations is one of Twitter's most appealing features.

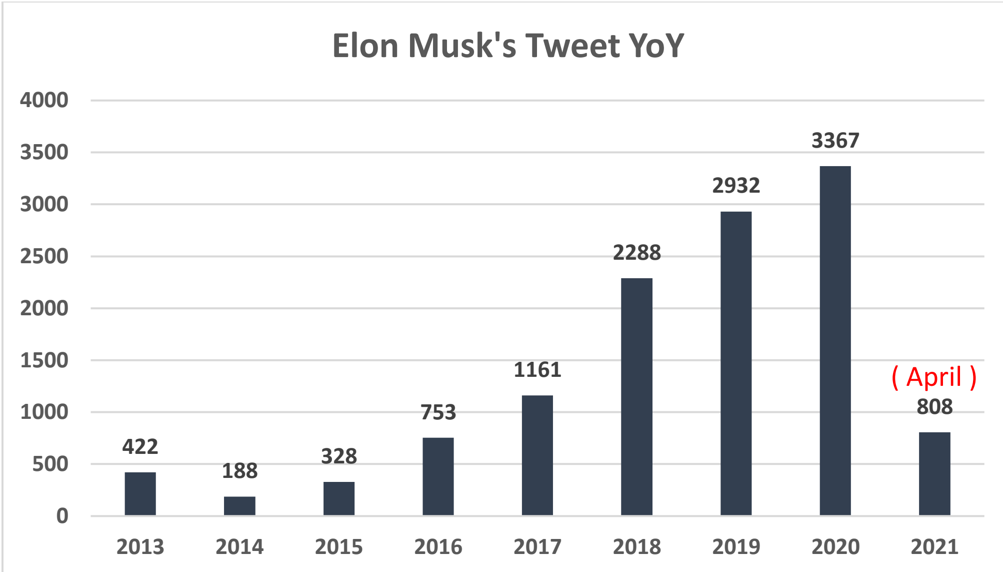

Figure 4. Yearly Tweet Pattern

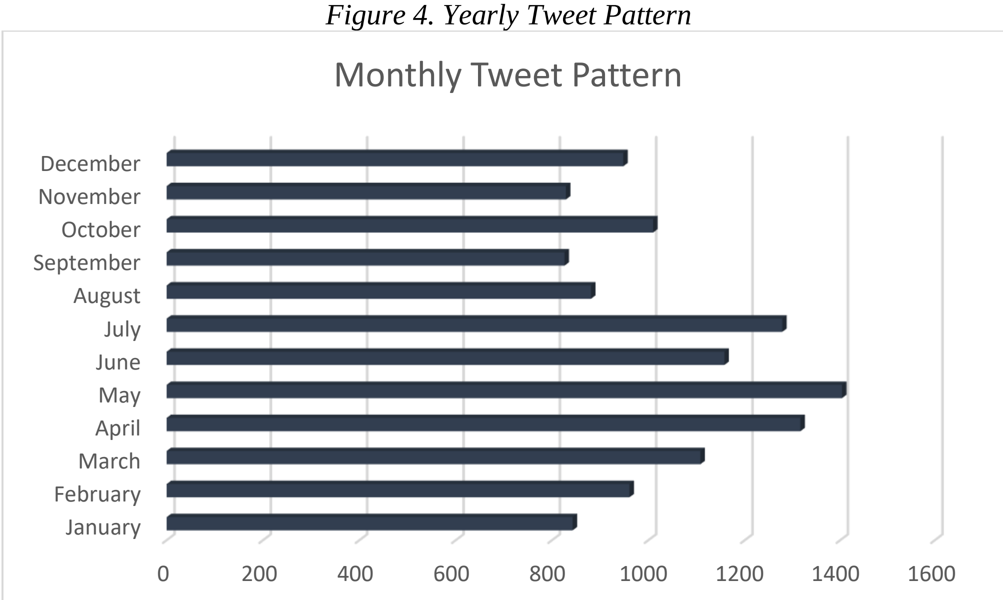

Figure 5. Monthly Tweet Pattern

The primary data that is collated for this analysis includes the detailed Tesla Opening, Closing, Day High, Day Low and Adjusted Closing prices. The stocks are listed from 1st January 2017. Also, we have used Elon Musk's Tweets analysis to support the sentiments projected by his tweets that have a positive and negative correlation on the TSLA stock prices. Through APIs tweets from Elon Musk (@elonmusk) have been extracted and performed Sentiment Analysis on the same from 2nd Dec 2018.The major objective for choosing a stock like Tesla Inc. is that it has been very volatile with its CEO Elon Musk's tweeting pattern.

The primary data collated for this analysis includes the detailed Tesla Opening, Closing, Day High, Day Low, and Adjusted Closing prices. The stocks are listed from 1st January 2017. Also, we have used Elon Musk's Tweets analysis to support the sentiments projected by his tweets that have a positive and negative correlation on the TSLA stock prices. Through APIs, tweets from Elon Musk (@elonmusk) have been extracted and performed Sentiment Analysis from 2nd December 2018. The principal objective for choosing a stock like Tesla Inc. is that it has been very volatile with its CEO Elon Musk's tweeting pattern.

4.2 Significant Instances of Stock Price Fluctuations:

1. April 1, 2018- April 4, 2018:
   Elon Musk tweeted on April fool's day that Tesla went bankrupt during a delicate financial period. This particular tweet had the stock fall steeply by 7%. After the recovery of the stock prices, it grew by a whopping 13.5% in a single day on April 4, 2018, after falling consistently for two days.
2. August 7, 2018:
   Elon Musk tweeted about taking Tesla Inc. private, and he had secured funding worth 420$. Tweet Content: I am considering taking Tesla private at $420. Funding secured.
3. June 17, 2018:
   Elon Musk lashed out at short sellers for the Tesla stock and predicted their short positions would have exploded in 3 weeks. This tweet passed a positive sentiment for all the long-term public investors and helped the day high reach 5.43% at US$ 74.76.
4. Key Highlights of Particular months of Tesla Stock Fluctuations
   A. Highest Consecutive Day Volatility – Dec 2020

Figure 6. Highest Volatility

   B. Mid Sept-Oct 2020:
   During this time, the COVID pandemic had worst hit US and the imposition of a lockdown was uncertain. During this time the Stock Market volatility was the highest in terms of the traded stock value.

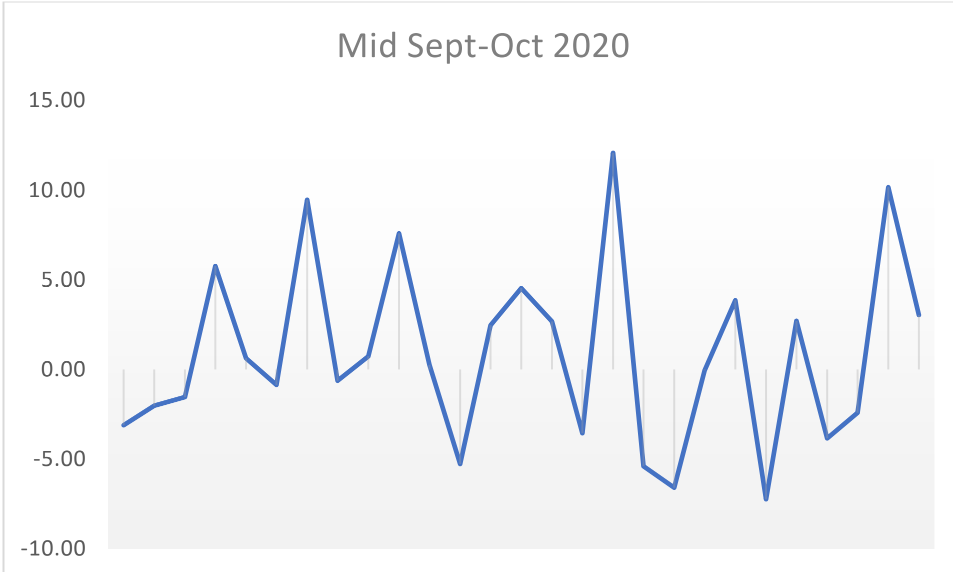

Figure 7. COVID Fluctuations

C. Covid-19 Pandemic:
The US automotive sector was severely hit by the pandemic and same story was depicted by Tesla Inc. with its share constantly dropping over the period of March and April 2020.

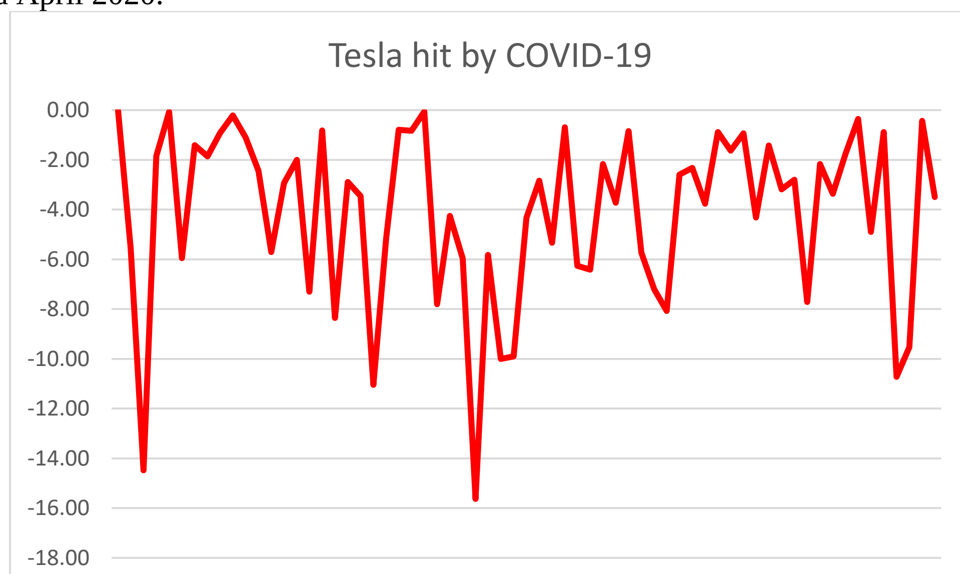

Figure 8. Effect of COVID on Tesla Share Prices

With thorough understanding and mapping of Tweets for Sentiment Analysis, we have concluded on a model score for the approximate percentage of Tesla Stock price volatility. The research was aimed at predicting an approximate rise and fall of Stock price for Tesla Motor. Inc

The application includes capabilities like as sentiment analysis, pos-tagging, and noun phrase extraction, among others.

TextBlob is a Python package with a straightforward API for interacting with its functions and doing basic NLP tasks. Because it acts similarly to Python strings, TextBlob is handy.One can alter and experiment with it just like in Python.

Tokenization, process of breaking down a text or a sentence into a series of tokens that roughly equate to "words." Tokenization is one of NLP's most important responsibilities.

Extraction of Noun Phrases:

We can extract the noun phrases from the TextBlob because we extracted the words in the previous step. When analyzing the "who" in a sentence, noun phrase extraction is critical.

Tagging of Parts-of-speech:

The method of marking text words based on their definition and context is known as part-of-speech tagging or grammatical tagging. It determines if a comment is a noun, an adjective, a verb, or something else. This is simply a more comprehensive form of noun phrase extraction, in which we wish to locate all elements in a sentence that are related to the part of speech

Inflection and Lemmatization of Words

Inflection is the method of assigning letters to the underlying form of a word to communicate grammatical meanings. TextBlob's word inflection is relatively straightforward, which means that the words we tokenized from a TextBlob can easily be transformed to singular or plural.

Advantages

1. Because it is developed on top of NLTK and Pattern, it is simple to use by giving an intuitive NLTK interface.
2. It uses Google Translate to perform language translation and detection

TextBlob is a popular text analysis toolkit that produces two values: subjectivity and polarity when used to evaluate the sentiment of a text.

Polarity and Sensitivity

Polarity is a number between-1 and 1, with-1 being extraordinarily negative and +1 being extremely positive. Subjectivity is a scale that ranges from 0 to 1 and refers to things like opinions, feelings, and even judgment. The text is more subjective the higher the number.

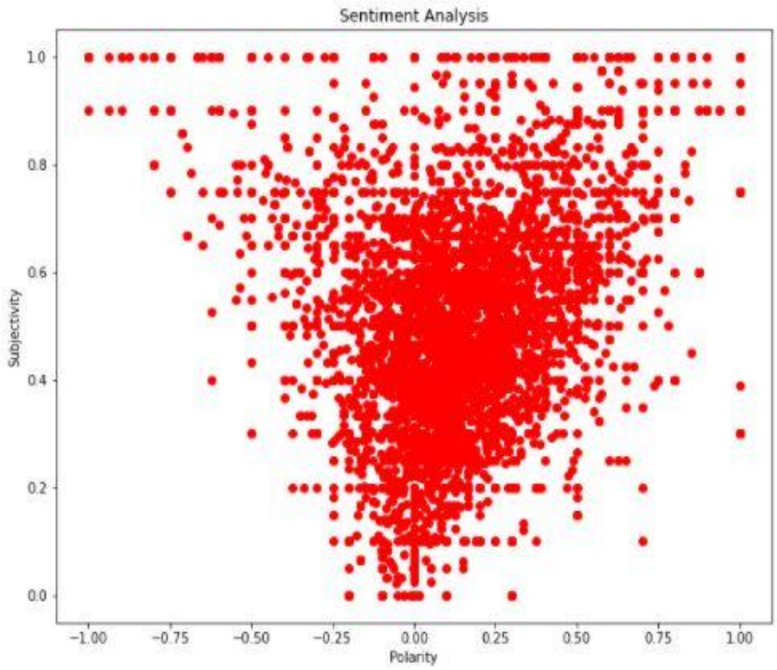

Figure 9. Polarity vs Subjectivity

FB Prophet:

The Prophet is a method for forecasting time series data that uses an additive model to fit non-linear trends to yearly, monthly, and daily seasonality and holiday impacts. Facebook's Core Data Science team published Prophet as open-source software.

The Prophet method gives users many options for tweaking and adjusting forecasts. By combining subject expertise with human-interpretable parameters, one can improve their prediction.

It works best with time series with substantial seasonal influences and historical data from multiple seasons. The Prophet is forgiving of missing data and trend shifts, and it usually handles outliers well.

Autoregressive models are the most often applied models for forecasting predictions. In a nutshell, the autoregressive model states that the output variable is linearly dependent on its initial values and an imperfectly predictable stochastic.

Understanding the math under the modelling of Prophet:

Y (t) = G (t) + S (t) + H (t) + E (t)

Where,

G(t) denotes Growth

S(t) denotes Seasonality (in yearly, monthly, weekly, daily)

H(t) denotes Holidays (Holidays impact businesses)

E(t) denotes the Error term.

FB Prophet employs a piecewise linear model for trend forecasting. Prophet model fitting is usually relatively rapid (even with hundreds of observations) and requires no data pre-processing. It also takes into account missing data and outliers.

Major Reasons of using Facebook Prophet are:

Accurate and Fast: It is accurate and generate results very fast

Reliable: Facebook Company itself uses Prophet for internal forecasting
Fully Automatic: Works with missing data & No need to perform extensive data preprocessing
Domain Knowledge Integrations: Forecasting can be made better by adding domain knowledge expertise like holidays & patterns

Implementation:
The FB Prophet model is an extensive model that encompasses the labelled data with mapped sentiments for the day. The results reflect a strong correlation between the sentiments gathered over Twitter and fluctuations in the stock prices.

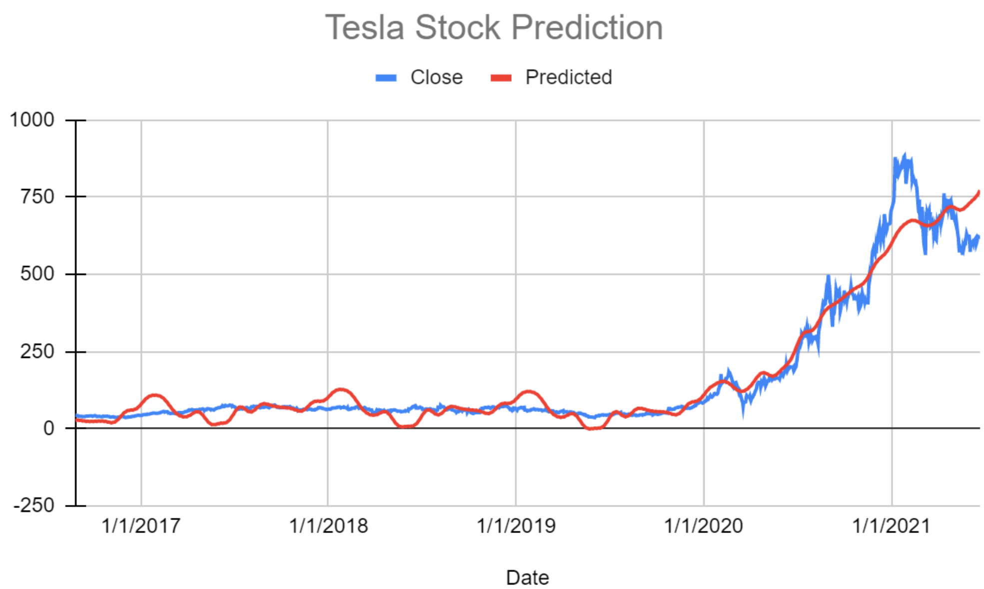

Figure 10. Tesla Stock Price Prediction

5. Conclusion

In conclusion, our research used a variety of functions to achieve our primary goal of identifying a clear association between Elon Musk's Twitter and the value of Tesla's stock.
We used a variety of tools to extract data to achieve this goal. We discovered that in the short run, the number of tweets Elon Musk wrote and his interaction marginally corresponds with the stock price of Tesla by examining the frequency of replies or tweets and the stock value varying over time.

In the long run, though, the association becomes more evident when looking at the statistics. In other words, when the tweets were evaluated across months or years rather than days, the association was more prominent. We discovered that Tesla's closing price changes had a direct, parallel association to Musk's engagement after evaluating all of the data.
Our findings are helpful to the general public since they reveal that Elon Musk's Twitter involvement is a good predictor of stock price rise. Though not entirely correct, the knowledge provided by the tweets is worth noting before taking any action.

However, we acknowledge in our research that there is a paucity of comprehensive data. In other words, the association identified in Tesla does not necessarily apply to other companies in general.

6. References:

Ahmed, B. (2020). Understanding the impact of investor sentiment on the price formation process: A review of the conduct of American stock markets. *The Journal of Economic Asymmetries*, *22*, e00172. https://doi.org/10.1016/j.jeca.2020.e00172

Bhardwaj, A., Narayan, Y., & Dutta, M. (2015). Sentiment analysis for Indian stock market prediction using Sensex and nifty. *Procedia Computer Science*, *70*, 85-91. https://doi.org/10.1016/j.procs.2015.10.043

Guo, K., Sun, Y., & Qian, X. (2017). Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market. *Physica A: Statistical Mechanics and its Applications*, *469*, 390-396. https://doi.org/10.1016/j.physa.2016.11.114

Huang, A., Wu, W. and Yu, T. (2020), "Textual analysis for China's financial markets: a review and discussion", *China Finance Review International*, Vol. 10 No. 1, pp. 1-15. https://doi.org/10.1108/CFRI-08-2019-0134

Kalampokis, E., Tambouris, E. and Tarabanis, K. (2013), "Understanding the predictive power of social media", *Internet Research*, Vol. 23 No. 5, pp. 544-559. https://doi.org/10.1108/IntR-06-2012-0114

Mittal, A. & Goel, A. (2012). Stock prediction using twitter sentiment analysis. Stanford University Working Paper

Moeini Najafabadi, Z., Bijari, M. and Khashei, M. (2020), "Making investment decisions in stock markets using a forecasting-Markowitz based decision-making approaches", *Journal of Modelling in Management*, Vol. 15 No. 2, pp. 647-659. https://doi.org/10.1108/JM2-12-2018-0217

Paramanik, R. N., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. *Procedia Computer Science*, *176*, 330-338. https://doi.org/10.1016/j.procs.2020.08.035

R. Nofsinger, J. and Varma, A. (2014), "Pound wise and penny foolish? OTC stock investor behavior", *Review of Behavioral Finance*, Vol. 6 No. 1, pp. 2-25. https://doi.org/10.1108/RBF-12-04-02

Ramalingam, V. V., Pandian, A., & Bhatt, J. P. (2018, April). Analysing News for Stock Market Prediction. In *Journal of Physics: Conference Series* (Vol. 1000, No. 1, p. 012026). IOP Publishing.

Riadi, S., Triono, S., Syahril, S., & Nofriansyah, D. (2019). Effectiveness of Metcaognitive Learning's Model in Engineering. *International Journal of Engineering and Advanced Technology (IJEAT)*, *9*(1), 4438-4443. ISSN 2249-8958

Sen, J., & Chaudhuri, T. D. (2016). An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice: a comparative study of the Indian consumer durable and small cap sectors. Journal of Economics Library, vol. 3, no. 2, pp. 303 - 326, arXiv preprint arXiv:1605.03930.

Seng, J.-L. and Yang, H.-F. (2017), "The association between stock price volatility and financial news – a sentiment analysis approach", *Kybernetes*, Vol. 46 No. 8, pp. 1341-1365. https://doi.org/10.1108/K-11-2016-0307

Sun, Y., Jin, Q., Cheng, Q. and Guo, K. (2020), "New tool for stock investment risk management: Trend forecasting based on individual investor behavior", *Industrial Management & Data Systems*, Vol. 120 No. 2, pp. 388-405. https://doi.org/10.1108/IMDS-03-2019-0125

Yao, C. Z., & Li, H. Y. (2020). Time-varying lead–lag structure between investor sentiment and stock market. *The North American Journal of Economics and Finance*, *52*, 101148. https://doi.org/10.1016/j.najef.2020.101148