

Forecasting the Stock prices of Tesla using Time-series and Sentiment based Analysis with Machine learning and Deep Learning Methods

SHUBHKUMAR BHARATKUMAR PATEL

3077432

Submitted in partial fulfillment for the degree of
Master of Science in Big Data Management and Analytics

Griffith College Dublin
September, 2022

Under the supervision of Dr Osama Abushama

Disclaimer

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the Degree of Master of Science in Big Data Management and Analytics at Griffith College Dublin, is entirely my own work and has not been submitted for assessment for an academic purpose at this or any other academic institution other than in partial fulfilment of the requirements of that stated above.

Signed: ____Shubhkumar____

Date: _12-09-2020 ____

Table of Contents

Acknowledgements	Error! Bookmark not defined.
List of Figures	iv
Abstract	v
Chapter 1. Introduction	1
1.1 Problem Statement	3
1.2 Aim of the Study	4
1.3 Research Objectives	4
1.4 Research Questions	4
1.5 Document Layout	5
Chapter 2. Background	6
2.1 Stock Market Price Prediction using Time-series data	6
2.2 Stock Market Price Prediction using Public Sentiments.....	6
2.3 Twitter Data and Tesla Stock Price Prediction	9
2.4 Machine Learning Algorithms in Tesla Stock Price Prediction	10
2.5 Deep Learning Algorithms in Tesla Stock Price Prediction	12
2.6 Other Classification models in Tesla Stock Price Prediction	14
2.7 Summary of Research	15
Chapter 3. Methodology	17
3.1 Time series-based Analysis to predict Tesla Stock Prices	17
3.2 Sentiment Analysis to predict Tesla Stock Prices.....	23
Chapter 4. System Design and Specifications	28
4.1 Convolutional-Recurrent Neural Network Model	28
4.2 Convolutional-Long Short-Term Memory Model	29
4.3 Convolutional- Bidirectional Long Short-Term Memory Model	29
4.4 Random Forest Regressor	30
4.5 Catboost Algorithm.....	31
4.6 Deep Neural Network	32
Chapter 5. Implementation.....	34
Chapter 6. Testing and Evaluation.....	35
6.1 Experimentation for time series analysis and prediction of Tesla Stocks	35
6.2 Experimentation for sentimental analysis of tweets and prediction of Tesla Stocks.....	40
6.3 Discussion	43
Chapter 7. Conclusions and Future Work.....	46
References	47
Appendix I.....	Error! Bookmark not defined.

List of Figures

Figure 1. Tesla Sales Number Projection till the year 2025	2
Figure 2. Social Media and its dimensional impact on stock market	8
Figure 3. Framework for Forecasting Tesla Stock Prices using Time-Series Method	18
Figure 4 Pair Plot Of all attributes	19
Figure 5 Trend of Tesla Stocks by Years.....	20
Figure 6 Outliers in the data distribution	21
Figure 7 Total Volume analysis by Years.....	22
Figure 8 Total Volume analysis by Months.....	22
Figure 9. Framework for Forecasting Tesla Stock Prices using Sentiment Analysis from tweets.....	24
Figure 10 Negative sentiments for Training data	26
Figure 11 Positive sentiments for Training data	26
Figure 12 Conv-RNN Algorithm	28
Figure 13 Conv-LSTM Model	29
Figure 14 Conv-BiLSTM Algorithm	30
Figure 15 Random Forest Regressor Algorithm	31
Figure 16 Catboost Algorithm	32
Figure 17 Deep Neural Network Algorithm	33
Figure 18 High Stocks Value Forecast by Conv-RNN	36
Figure 19 Low Value Stocks Value Forecast by Conv-RNN	36
Figure 20 Open Value Stocks Value Forecast by Conv-RNN	36
Figure 21 Closing Value Stocks Value Forecast by Conv-RNN	37
Figure 22 High Value Stocks Value Forecast by Conv-LSTM	37
Figure 23 Low Value Stocks Value Forecast by Conv-LSTM.....	38
Figure 24 Opening Value Stocks Value Forecast by Conv-LSTM	38
Figure 25 Closing Value Stocks Value Forecast by Conv-LSTM.....	38
Figure 26 High Value Stocks Value Forecast by Conv-BiLSTM	39
Figure 27 Low Value Stocks Value Forecast by Conv-BiLSTM.....	39
Figure 28 Opening Value Stocks Value Forecast by Conv-BiLSTM	39
Figure 29 Closing Value Stocks Value Forecast by Conv-BiLSTM.....	40
Figure 30 Comparison between the actual closing price and the predicted closing price using Random Forest Algorithm.....	41
Figure 31 Comparison between the actual closing price and the predicted closing price using CatBoost Algorithm	42
Figure 32 Comparison between the actual closing price and the predicted closing price using Deep Neural Network.....	43
Figure 33 Comparison Of all Models executed for predicting stocks based on Time Series Analysis.....	44
Figure 34. Comparison Of all Models executed for predicting stocks based on sentimental analysis	45

Abstract

Given the recent increase in stock expenses, investing in Tesla stock is considered prudent. Equity claims may be risky and volatile depending on the company and the stock market. Investing in digital services is becoming increasingly popular as a result of global advancements in the internet and technology. Such business strategies rely heavily on online marketing initiatives. Clients have access to a wide range of investment options that can be evaluated based on their specific needs. Customers base their purchases on asset ratings. Every day, millions of customers post online reviews of these products and services. The goal of this research is to find the most efficient machine learning or deep learning-based system for predicting Tesla stock prices using time series analysis. Along with time-series analysis, sentiment-based analysis has been used to investigate how tweets affect the price of Tesla stocks because it has been discovered that market trends in this stock market are heavily influenced by news and Twitter posts because they integrate the various perspectives of all investors. We estimated the price of Tesla shares using Twitter statistics and other old archives, as well as time-lapse analysis and sentiment analysis. We achieved excellent results for time series analysis by utilizing the long-time quick epoch memory Conv-LSTM model. To determine the sentiment of tweets and forecast the price of Tesla stock, a set of sentiment analysis rules based primarily on Twitter is used by implementing Natural Language processing. Twitter-based estimation of stock rate is nicely predicted by the Catboost algorithm. The most appropriate model for valuing Tesla stock is identified after the evaluation of each algorithm on MAE. The model which achieves the lowest MAE value is considered the best performing model and thus can be employed in real-world situations for forecasting purposes.

Chapter 1. Introduction

Research in the financial market is extensively documented to enable the investors to understand the impact actually present based on the news (both positive and negative) and thereby understanding public sentiment with the best effort. While diving deep into financial analysis, the investigation of the stock market price is a challenging part due to its immense fluctuation and complexity. There are several approaches that have been documented as part of financial literacy to demonstrate the critical path of predicting the stock market price. Even more, these approaches mainly concern the historic as well as present information (both numeric and textual) to get a better prediction value. The financial organizations and trading industry is continuously evolving their systems to evaluate the economic value to make a productive investment and thereby get a good return. In a more specific way to understand the fact, the financial and market analyst who intends to predict the investment rate to enable a better investment process for the investors to gather information from the public based on certain intermediary steps. The information extracted from the news extensively put an effect on the performance scale of stock price and also promotes a link among magnitude and direction in response to the stock price and nature of the news [12]. According to the establishment and concurrent findings from the studies, it is regarded that the new mainly constitutes media coverage specific to different companies. Thus, it can be stated that financial market analysis can be regulated on the basis of the retrieved information and successively predict the future stock price from the past and present data. As per the statement from the Last driver license holder [26], it has been estimated that tesla sales will increase with the growth rate of 59%, where Tesla will sale 5.1 million EV vehicles by the year 2025. Sales projection of Tesla is shown in Figure 1.

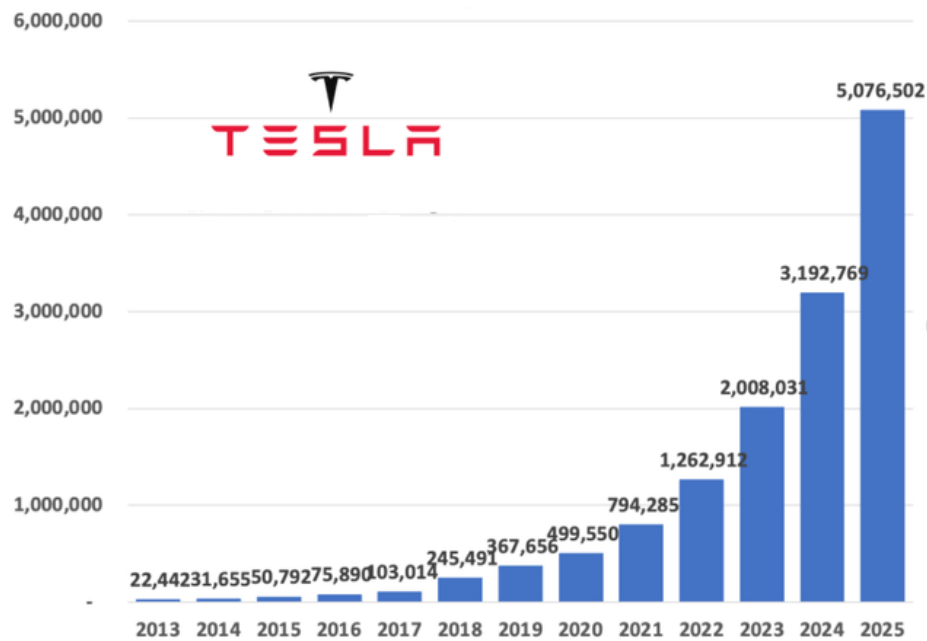


Figure 1. Tesla Sales Number Projection till the year 2025

Public sentiment and graphical representation of historic data are the two basic attributions that enable the analysis more efficient and productive. However, in recent times, researchers grab more attention and shown more interest in public sentiments that are obtained from social media, such as Facebook, Twitter, Instagram or blog. Among them, the most versatile and efficient information can be retrieved through Twitter comments, hence considered Twitter Sentiment Analysis. On the other hand, the historical data can be collected and abide by the condition of time series forecast, or Time-series Analysis. These two conditional analyses have been used vigorously for the prediction of stock prices and concurrently various classifiers are used to recognize the behavioural pattern of the data. Here, the most important discussion places an understanding of the classifiers that have been used for many years in order to predict the stock market price and yet they hold the same consistency. The most common method used is the classification algorithms such as machine learning.

The efficiency of the machine learning algorithms is well-known among researchers and experts which has introduced a wide range of applications in various fields for more than a decade and still holds its consistency. Through this approach, generalizability can be attained and thereby obtuse knowledge with the proper understanding of the

features the corresponding application actually represents. The prediction of the stock price is a challenging factor as has been discussed earlier and thereby effective and distinct approaches are needed. The main emphasis of stock price prediction in this study explains Tesla company stock price prediction based on both time-series data and Twitter comments with the incorporation of certain classifiers to ease the prediction process. Nowadays, the prediction of the stock market price is rapidly evolving with a distinct classification process of textual data that has been extracted from social media which significantly represents the sentiment of the public and even more the positive and negative impact it holds on the corresponding occurrence of events [14]. The observation and analysis of Twitter data have vigorously changed the old concept of stock market price prediction and even made it easier for experts and analysts to induce a positive approach to predict the future price that can help in a better investment process. In fact, it is an intriguing research-oriented field which has received attention for almost a decade. The conclusion drawn from previous studies shows aggregate statistics of public mood (in terms of opinion) based on Twitter data which might be correlated to the “Dow Jones Industrial Average Index” (DJIAI) [14]. Therefore, this paper would explore all the credentials, including the previous contributions of authors and researchers and how well the stock market price can be predicted, the rise and fall in the price over time, and also the correlation it persists with the viewpoints of public being exposed in Twitter about Tesla company stock.

While understanding the author’s opinion based on their pieces of work where the textual data from social media has gained more importance, the contribution of various classifiers shows its significance in classifying the behaviour of the texts, a visualization approach apart from data mining through natural language processing (NLP) technique. The wide sense of knowledge that has been gained with the conduct of this research study has explained various aspects and accordingly explores several algorithms and other techniques that can be helpful in the prediction of the stock price of Tesla. In such context, supervised principles of machine learning, deep learning algorithms, and other efficient methods have been explored.

1.1 Problem Statement

The study has revealed its importance in remodelling techniques for stock market price prediction which becomes a challenging and vital aspect of research. While looking deep into the matter, the main atrocities are found with its frequent changes or fluctuation rate due to which it becomes more imprudent to meet the actual price. Thus, this study has reviewed this issue and thereby promotes in-depth research on understanding the stock market price with a collateral discussion of the Tesla Stock Price.

1.2 Aim of the Study

The aim of the study is to analyse and predict the company's stock price for Tesla based on the events considering both the Time-series analysis and Twitter sentiment analysis accordingly with the use of machine learning algorithms and deep learning algorithms. In this approach, the main flow of the research would be the predicting stock price and hence consider suitable methods as stated above to meet the resultant outcome of the study.

1.3 Research Objectives

To meet the aim of the study, the following research objectives have been considered:

- To conduct a suitable analysis based on both time-series and public sentiments from Twitter comments to recognize the better approach
- To identify the various methods that can be used to predict the stock market price for Tesla in combination with the analytical concepts
- To determine the accuracy level of each method or classifier in the process of prediction and also find out the most convenient one among them
- To determine the feasibility of both machine learning and deep learning algorithms when utilized with natural language processing (NLP).

1.4 Research Questions

- What is the contribution of both time-series and Twitter sentiment analysis with the advent of Tesla Stock Market Price Prediction?
- How efficiently stock market price of Tesla can be calculated? Which approach will be more suitable to obtain optimal outcomes?

- Which algorithm accurately forecast the price of Tesla for both time-series and Twitter based sentiment analysis approach?

1.5 Document Layout

With the establishment of this chapter, the identification of main processes, suitable exploration of analytical concepts, need for pertinent analysis to determine the fluctuation and hence restructuring of the methods for prediction of the stock market price for Tesla have been developed. Since this paper has intended to correlate the past, present and future data, with the outcome that it would serve can help in enhancing a positive predictable motive for the investors and advertently promotes identification of whether the approach is convenient or not.

In the first chapter, the introduction about the stock market price prediction will be discussed along with the problem statement, objective of this study and research questions. In the Chapter-2, prior work related to different techniques related to stock market price forecasting will be discussed in detail. Chapter-3 in this report discusses about the research methods has been used for implementation of the project. Next chapter, Chapter-4 discuss about the design, specification and architectural part of algorithm used for this project. The tools, technologies and libraries used for the implementation of project has been discussed in detail in Chapter-5. Evaluation of obtained results will be discussed in the Chapter-6. Furthermore, Chapter-7 summarises the outcomes of this research along with the discussion of future possibilities.

Chapter 2. Background

2.1 Stock Market Price Prediction using Time-series data

The contribution to the prediction of stock market price has been considered one of the important phenomena which help in establishing proper decision-making steps within the financial domains. The specific elaboration of this prediction can be simulated on the basis of the types of data being used for analysis and accordingly retains the changes observed therein. In this section, the discussion would be based on time-series data that can be utilized for predicting the stock market price. It is one of the conventional approaches which is based on the data considering previous prices, however, the rate of precision can be considered insignificant since the price fluctuations are not only based on features affecting the cost structure of supplies but also based on the impact of news as well as the changes admissible to the company's growth [19]. With today's evolution in technology, the internet has gained considerable attention and become enriched with natural language processing (NLP) techniques like forums or social media comments which undergo data mining to create the text corpora [15]. Thus, with this establishment, it can be stated that the conventional system of data analysis becomes insignificant due to its inconsistency where its disadvantage lies and become a reason for the successful implementation of another convenient predictive analytical process, i.e. Sentiment analysis.

2.2 Stock Market Price Prediction using Public Sentiments

The fundamentals of sentiment-based stock market price prediction have gained significant support from researchers due to its high credible features and prediction capacity. To be specific, sentiment analysis is nothing but studying the sentiment of the public, and their behavioral motives which may be positive or negative, thus helping in the easy judgmental ability to rate the fluctuating price. The research that uses public opinion to predict the change in stock price develops a specific result that would be beneficial for investors in their investment process. In recent times, social media

comments are the key aspects of this prediction process, especially Twitter comments. Due to the dynamism, non-linear and abide by complexity, stock market price becomes inherently a difficult aspect to predict [3]. However, various debates have been made on the usefulness of public sentiments in analyzing the price and even more, in making a decision in economic domains for the trade market investors [3] [4].

The streaming data promotes an advantage through proper dealing with the continuum being collected on the basis of a real-time approach [4]. There has been a rising confrontation among the researchers to opine their views on the acceptance and rejection of the relationship that confirms a specific causality within sentiments & the trade in the stock market. Nevertheless, there is much advancement that has been made through computational methods that emphasized sentiment-based features but there is not a sense of maturity & performance [3]. Thus, to put an emblem on this sentiment-based analysis, various classifiers have been developed and introduced to get acquainted with this concept [5].

Taking note of the sentimental aspect of human beings, it is easier to determine the crucial aspects of the business market. In the case of financial and trade analysis, for example, stock market price, public sentiment can be beneficial and researchers often take notes on investors' sentiment for capturing their beliefs regarding future stock performance [7]. Investors often use social media information for exploiting the abnormal possibilities of returns. Under such circumstances, the contribution of Twitter data gains more importance due to its highly specific comments. Now considering social media, various dimensions also come into effect that influences the movement of the stock market.

The introduction of the first dimension explains the user which typically depends on the publisher's expertise & popularity [7]. It is because of this popularity and expertise that this shared information gets more importance on social media platforms and thereby can be beneficial in detecting the stock price. Coming to the second dimension, explains the message which is enriched with information and also contains convenient words that attract users [7]. Based on this dimension, the effectiveness of sentiment analysis becomes more approachable. Bestowed on the condition, these messages are

beneficial due to their positivity which significantly increases the trading activities prevalent in the stock market. The last dimension indicates the discussable part where the main effect becomes simulated based on the post which may have some positive effects on the corresponding trading activity [7]. From the precise note of the different dimensions of social media, it can be concluded that these can be helpful in predicting stock returns based on tweets.

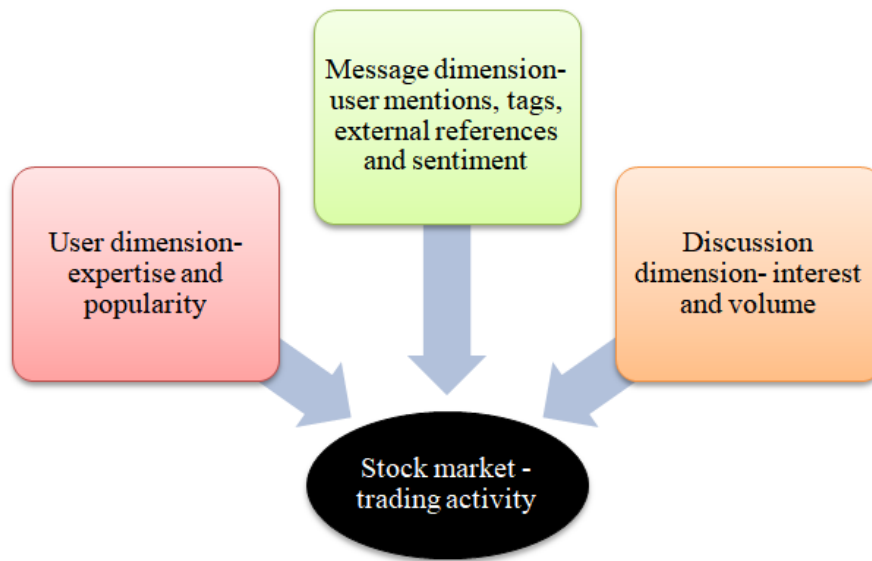


Figure 2. Social Media and its dimensional impact on stock market

There has been an extensive understanding of social media applications, especially Twitter sentiments in literature to predict the stock price (Hu et al., n.d.). The use of sentiment analysis based on the data from Twitter is significantly combined with the classification models in order to achieve a better predictive result. Apart from machine learning and deep learning which have gained much favor from researchers and experts, there are several other models like BERT- a transformer, etc. have been used significantly for the purpose of sentiment analysis which therefore combines with the data of stock price. Even more, the historical data as discussed in the previous section are used also in combination with the support vector machine (SVM) and ARIMA model to predict the prices [9]. Thus, it can be concluded that sentiment analysis has gained more importance than historical data being used single-handedly and is

deliberately become a suitable approach in predictive analysis in combination with classification models.

2.3 Twitter Data and Tesla Stock Price Prediction

The data obtained from Twitter generally contains public reviews which become an important asset for market researchers in predicting their stock market price. Since, it has already been discussed previously, that this study would consider the datasets for Tesla stock market price prediction, this section thus covers all the important credentials as obtained from previous studies. As discussed in the study conducted by [7], the author has used the Twitter datasets that mainly contain public tweets from a date range of 1 October 2019 to 30 December 2019. All these data contain information on Tesla Inc. In fact, a wide range of data has been gathered for capturing the proliferative sentiments of the public which contains general attitudes and not only the tweets concerning Tesla stock. In these approaches, one thing is clear public sentiments explain only Tesla stock which may be positive or negative based on the behavioral pattern of the text data.

The stock data of Tesla Inc. as mentioned in the study was obtained from the Bloomberg, via the Bloomberg terminal which contains both opening & closing prices. Here, to predict the movements in stock price, the necessary step being followed is a deduction of the opening prices from that of closing prices and hence obtaining the return [7]. Then accordingly, the classification models proposed in the mentioned study were utilized. This study has helped to discover the coherent steps followed in predicting the stock market prices for Tesla Inc. without showing any incompetency. While conducting any study that resolves around the prediction of stock market price, the dataset used is primarily trained to get a better evaluation result and hence used in sentiment classification models.

Upon delivering all this information, it can be ascertained that stock market price prediction although is a difficult task, the sequential use of various analytical concepts in combination with different classification models can be a better approach. It has already been established how the concepts draw a clear prediction process, especially

sentiment analysis through Twitter comments, however, in the later sections the author has also promoted a discussion on web-based classification models or simply classifiers for tackling the rising issue of frequent fluctuation in stock market price. To be exact, the suitable prediction of the stock price for companies like Tesla Inc. or others might be considered an unspoken wish on behalf of investors in the stock market. Though it is a complicated task, several theories have been configured for years to understand the stock movement, the effect of the sudden rises and falls along with the interest rates, social media information and news[10]. The further flow of information on different classifiers in combination with predictive analytical concepts has been established in later sections of this chapter which not only demonstrates the efficient performance of each classification model but also helps in determining the most accurate model for the Tesla stock market price prediction.

2.4 Machine Learning Algorithms in Tesla Stock Price Prediction

The implementation of Machine Learning Algorithms along with natural language processing (NLP) become a well-known and common classification model to predict the stock price based on historical data and Twitter sentiments. Many studies have contributed data for knowledge optimization on how exactly the precision and analysis can be done and concurrently explore the classification algorithm in stock market price prediction [10]. Prediction of the trends in the stock market price is an area which promotes significant interest to researchers with a motive to experiment with the complexity as well as dynamism showcased by the process. In fact, intrinsic volatility that is present in the stock market throughout the globe has become the main cause of the challenge faced therein [11]. Throughout the course of predicting and analysing the stock market price, forecasting & diffusion modelling may be effective but cannot be considered the panacea for encountering diversified problems, whether it is long-term or short-term or otherwise. Further acknowledgement shows that risk pertaining to the financial market has a correlation to forecasting errors which thereby need a prioritized measure to resolve in order to minimize the risk related to investment. Thus, the introduction of machine learning algorithms is a prime consideration by various authors that have even simulated data for research on these specific classifiers for minimizing

the forecasting error. Even more, through this approach, the issues related to forecasting receive tremendous attention as the classification problem [11].

Various classification models constituting this particular algorithm have been applied to mitigate a number of issues simultaneously occurring with the advancement of technology. However, apart from this, the models are also described as the classifier in predicting the behavioral pattern of the data needed to predict the stock market price in the trading industry. Even more, two or more classifiers are used in combination in recent times to be considered a novel approach that is defined as ensemble learning. The formation of predictive features from Tweets that can help in proper decision-making is what is categorically figured [18]. The predictive analysis that is achieved through a combination of classifiers and tweets for sentiment analysis can be a better approach that supports decision-making. In one of the studies conducted by [18], the predictive features from Twitter data have been discussed with the use of graph theory along with an itemization of datasets and the association rule theory that forms and retrieves several features from the datasets. According to the evaluation, it shows that quantitative characteristics indicating semantic frequent itemized sets are beneficial in the “predictive regression models” along with specific target variables.

As discussed by [14] in their study, equity traders seem to search for specific tools that can help in maximizing returns as well as minimizing risks, then let it be a fundamental aspect of techniques for technical analysis. The introduction of well-known sentiment classifier is a big and significant approach by researchers for extraction of sentiment from news headlines and from tweets. The use of different classifiers is then considered for fundamental analysis along with deep classification models or an ensemble algorithm. In the above-mentioned study, four specific machine learning classifiers have been used for the fundamental analysis along with deep LSTM model architecture which includes a “decoder-encoder LSTM model” to serve the purpose of technical analysis [14]. The experimental data has been obtained or mined from various news sites, Twitter as well as financial news from Yahoo. The result determined from both the experiments shows an accuracy level of 96 per cent with an RMSE value of 0.023 for predicting the movement of the closing price of the chosen telecommunication company. Thus, the findings revealed that sentiment features which contain both

fundamental, as well as technical indicators form the essential tools to move closer to predicting the stock price.

An approach to Bayesian Regression Analysis to build the time-series models as well as stacking various predictive models in a row for the time-series have been discussed in the study conducted by [17]. It has described how this regression method has modeled time series through nonlinear trends. It has been critically established that different classification models have been used for decades to remodel the time-series and sentiment datasets; however, few have succeeded in achieving the predictive result while considering the stock market price. And while considering those, the hierarchical model indicating time-series through Bayesian regression is one of them. According to information obtained, there are various types of research being conducted with the traditional ARIMA time-series model, however, the result remains uncertain cause. Hence, with this approach, especially stacking of predictive models helps in proper risk assessment in order to predict the price and accordingly support the decision-making process.

2.5 Deep Learning Algorithms in Tesla Stock Price Prediction

Stock market price prediction with deep learning algorithms is an interesting and successive approach that not only support researchers in their studies but also gave hope to successfully predict the price by analyst and thereby supporting the investors [1]. It is because of different external and internal factors that highly affect the stock market price and enhance continuous fluctuation. Thus, a novel approach with “deep reinforcement learning” has been studied by [1] with sentiments from community & knowledge gap. The result obtained from the experiment shows that this proposed model can help in achieving remarkable outcomes in predicting the stock market price for companies like Tesla. Prediction of stock price and the surfacing of the trends it holds in recent times has become a biased topic in the field of financial engineering. Even so, the complexity, as well as non-linearity in the particular area, is tremendously rising due to which the financial analyst, researcher and investors are facing unobliterate challenges. With the advent, the deep learning classification, especially the

neural architecture has become an efficient tool in sentiment analysis; especially in abide with financial time-series analysis [23].

The critical establishment of the results with the LSTM model and deep neural architectures are well-coordinated in this section [23] [19] [22]. It is because of the real-time nature found in collective information; quasi contemporary information is specifically harvested to predict financial trends. While outlining the condition of the information, and the development of the architecture, the “Long-short-term memory” LSTM model has been witnessed to tackle the high nonlinearity issues in forecasting prices of the stock market [24] [6]. The quantification of predictive power based on the sentiments from social media as well as the financial data regarding stock price predicting has been established by various research papers while utilizing comprehensive datasets which are related to both fundamental & technical variables of the stock market [12]. To be specific, sentiment analysis becomes a strategic tool for subsequent stock market price predictions as compared to the previous financial data, although both are considered simultaneously to conduct research on their features through various classification models. In fact, not only those algorithms but the use of other pre-trained models such as finBERT is also used for sentiment classification. It is a transformer that provides both sentiments classification and the softmax scores. Hence, it can be established from here that apart from engaging in utilizing the classification algorithms for predictive analysis based on sentiments, researchers and analysts have equally considered the use of other models for quantifying the impact of the features constituting the so-called “black-box model” [12].

Network analysis has become an interesting path in the prediction of stock prices which is highly volatile and inconsistent [22]. As it is already known stock price prediction is highly perceived with the use of historical data and the current trends in public sentiments [21]. The use of such data in establishing a predictive solution is primarily forecasted by various researchers and analysts based on different statistical, econometric and deep neural approaches. In a specific sense, deep learning models, such as neural network has gained more popularity in recent years than any other method. And the sub classification is based on the experimental approach from the

study of [22] shows that the “back propagation (BP) neural network” promotes consistency and robustness in outperforming other neural network models.

The critical analysis of Twitter sentiments through “Long-short-term memory” (LSTM) is another vital approach that enables proper decision-making in the financial market while serving the purpose of predicting the stock market. On a specific note, sentiment analysis serves the purpose of measuring subjectivity as well as polarity along with the consistent approach to the tweets reflecting a company’s financial growth and trading scenario [24]. It helps to capture the exact mood of the market that has a strategic influence on the stock price. According to the result obtained from the experiment, the root-mean-square error (RMSE) evaluated through sentiment analysis is 0.021. Thus, it can be stated that sentiment analysis through the LSTM model can be a help in predicting the stock market price of Tesla on contrary to other methods and even more can achieve an outcome which is more comprehensive and obvious.

2.6 Other Classification models in Tesla Stock Price Prediction

Determination of price movement in the stock market has been already established as a complication which is constantly faced by researchers even in recent times, even when several approaches have been done [8]. There are several factors such as industrial performance, certain economic variables, and the sentiment of the investors, company information and performance as well as social media sentiments that are constantly raising concern due to the complexity observed in price movement. Apart from the classification algorithms that have been discussed earlier, this section has focused on showcasing other models that can be used in this prediction process. In fact, this section has been viewed as an alternative approach or even an approach with machine learning where different data has been used.

One of the studies conducted by [8], describes the relocation of price movement through machine learning models based on the information from historical or past data, candle-stick-chart data of stock, and the data from social media. The movement has been predicted using only one classifier and the multi-channel collaborative network

structure that incorporates both candle-stick chart data and sentiments from social media to predict the trend in the stock. Even more, the past data collected are successively transformed into candle-stick-chart data for better elucidation of the patterns that indicate stock movement. Based on the experimental result obtained from the proposed models and the whole process has achieved 75.38 Accuracy in predicting the stock price of the chosen company, Apple Inc. Additional information is achieved on the performance level of the prediction which becomes more favorable for longer periods.

As discussed earlier, the use of transformers apart from machine learning or deep learning algorithms shows convenience in analyzing the correlations between social-media sentiments & the behaviour of the stock market [2]. The approach to determining the attitude of the financial market, sentiment analysis has been performed through the BERT transformer [2]. As an optimized condition, the correlation has been observed which shows the sentiment has no impact on this correlation at a significant level. Moreover, due to the nature identified with the financial news, classification through the BERT transformer becomes more negative or even neutral. Even more, the influence of noise in data is another factor that impairs the prediction quality. The result obtained here shows a characterization of multi-model architecture that might be able to weigh more persistently in extracting information and supporting sentiment analysis if being investigated more precisely the branch-specific posts, which in this case, the financial data.

2.7 Summary of Research

The establishment of this chapter has helped in acknowledging the contribution of various authors in their research for predicting the stock market price trend. Price prediction or predicting the movement in the closing price of the stock market is a constant challenge as discussed all over the study. It is notable that several methods and architectural models have been proposed by various researchers to convey a positive outcome in the precise prediction of the stock price. Several controversies have been viewed on such advent in order to understand the better performance level of the models

and their accuracy level. This study has considered determining the stock price movement of Tesla Inc. where the sentimental analysis becomes a prime factor. Accordingly, machine learning and deep learning models are investigated through the literary works of the authors where the price movement and prediction for other companies and even for Tesla Inc. are established. Bestowed on the result, both ML and DL classifiers have shown intriguing performance in prediction. Since a brief idea has been obtained from the information on their capacity and accuracy level, it would be prominent to investigate and experiment further so as to properly identify the model that can be used in Tesla stock price prediction. Therefore, it can be concluded that this chapter has developed an understanding of the models being accompanied by the literary notes of the authors and hence sustained a proliferative outcome bestowed on the findings from this study.

Chapter 3. Methodology

In this research work, the aim is to identify the most suitable deep learning-based algorithm which can be capable of predicting the prices of Tesla stock prices based on the time series analysis and Twitter-based sentiment analysis. A multi-step methodology is utilized to accomplish this, which includes retrieving and handling the data set, pre-processing of data, analysing and exploring the data, determining, and extracting features, training the model, and making predictions. Each phase is addressed in the following section. In order to anticipate the price of Tesla shares, here two distinct approaches are employed. First, we will examine the stock price time series, and second, we will determine how Twitter opinions affect the price of Tesla Stocks. The figure shown below depicts the process of Methodology.

3.1 Time series-based Analysis to predict Tesla Stock Prices

Time series analysis is the study of past trends utilizing the historical data of any entity. Based on the historical patterns, the future predictions can be made using Machine learning and deep learning approaches. In this research, we have utilized the deep learning approaches for model training along with-it evaluation of each model is performed using different set of metrics. The flow diagram representing the framework for Tesla stock price prediction using time-series analysis is shown in Figure 3.

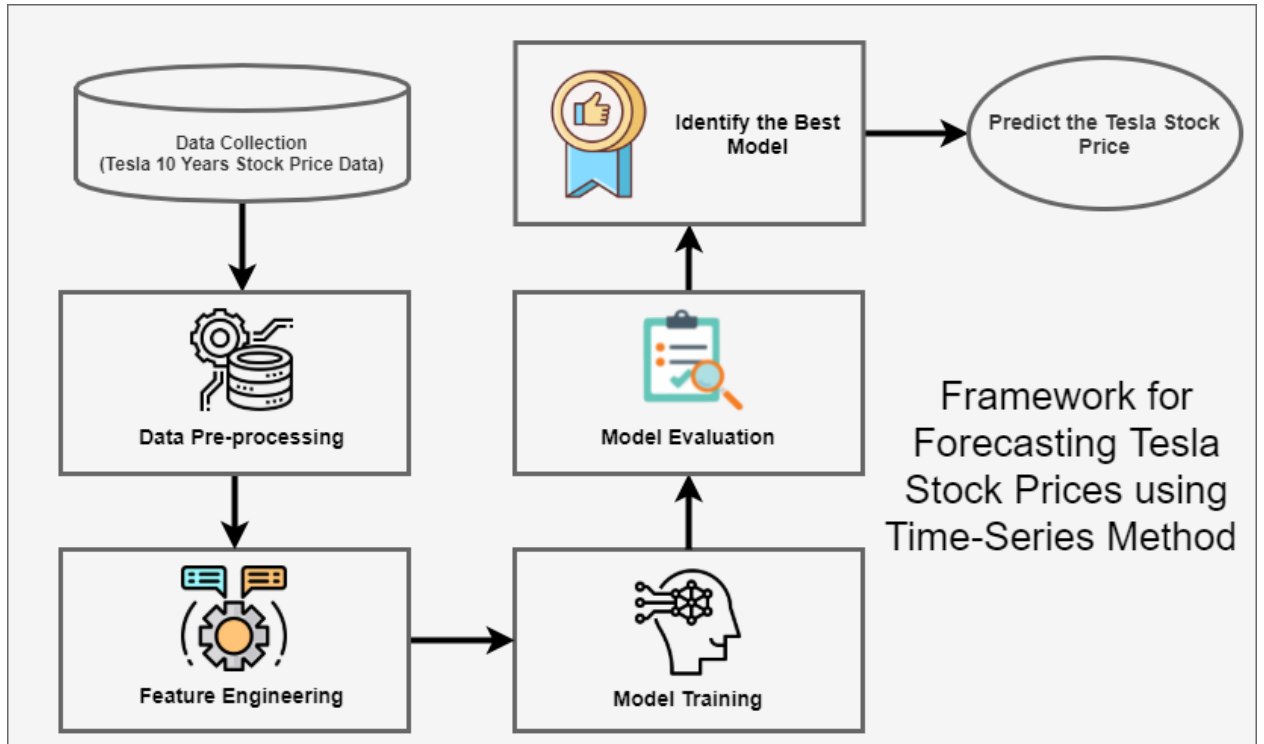


Figure 3. Framework for Forecasting Tesla Stock Prices using Time-Series Method

3.1.1 Data set Description

Due to advancements in technologies around the world, it is observed that investing in the stock market has become simple and straightforward. In order to predict the change in the stock prices of Tesla shares based on historical data, the dataset is collected from the Yahoo Finance website which is the bonafide source. This dataset contains 3058 rows and 7 columns. These columns have the date, Open, High, Low, Close, adjacent close, and volume. These all columns contain the numerical value, but the range varies a lot except for the date column which has a Date Time format. This data contains information from June 2010 to August 2022. The size of the dataset is 5MB.

3.1.2 Data Pre-processing

Pre-processing of the data is important when utilizing machine learning or deep learning algorithms because these algorithms cannot perform well on rough data. In order to clean the data first, the null values from the data are eliminated and the data type of each column is changed as required. The date feature is also changed into the required date time format.

3.1.3 Exploratory Data Analysis

After pre-processing of the data, the data analysis and visualization steps are performed to get better insights from the data because it helps in understanding various entails of the data and also assists in selecting machine learning or deep learning algorithms for the prediction of the Tesla share prices. First, the data distribution is plotted using a pair plot which is shown in figure 4.

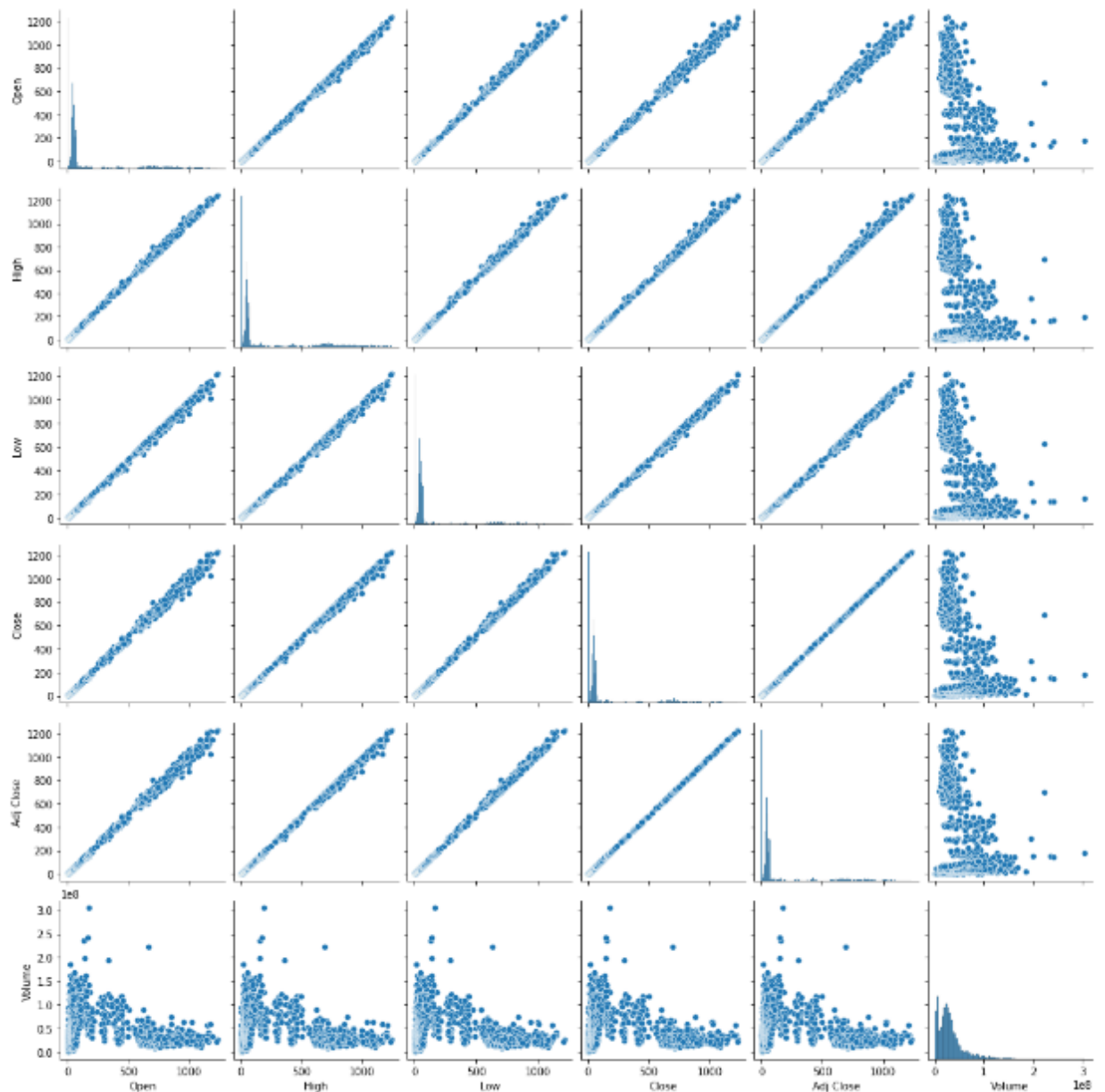


Figure 4 Pair Plot Of all attributes

Further stock values trend from 2010 to 2022 is plotted using a line plot as shown in figure 5. From this figure, it is observed that stock values continuously increase but the highest increment is observed from 2020 to 2022.

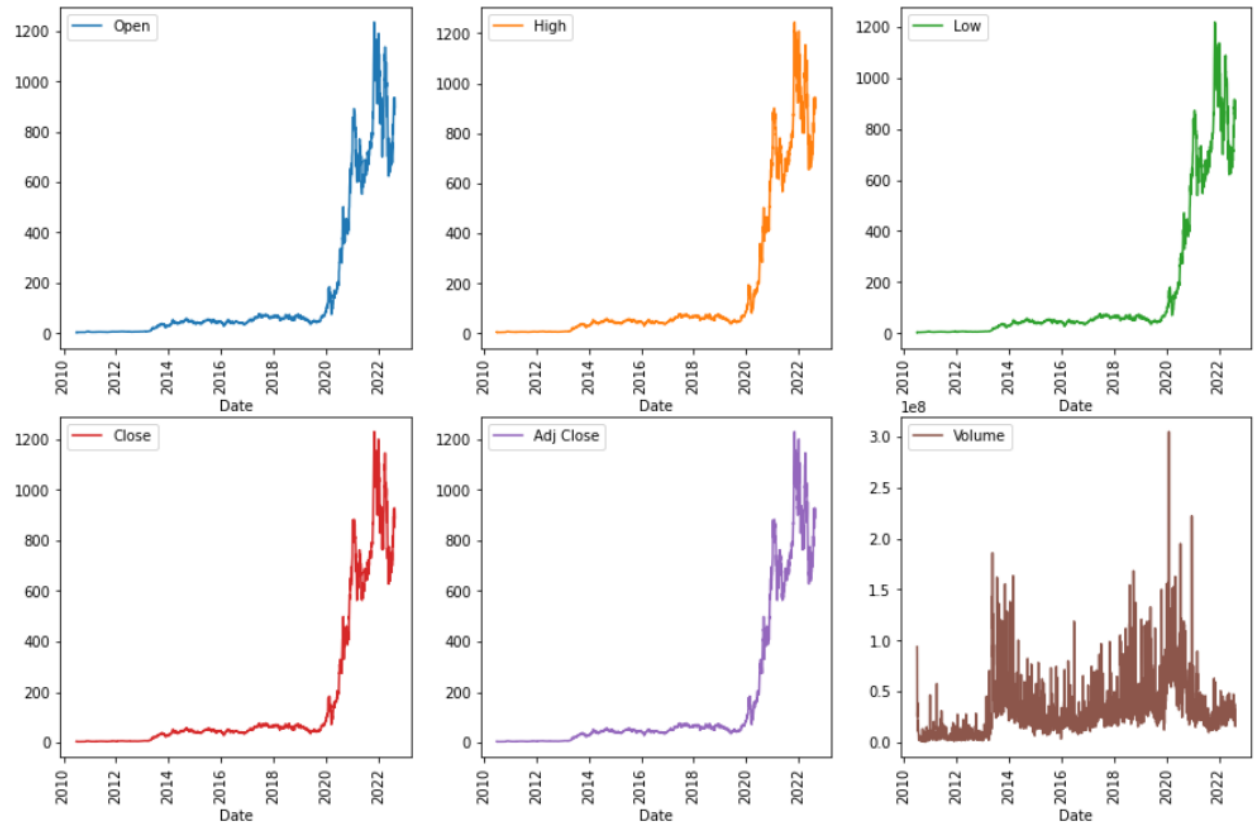


Figure 5 Trend of Tesla Stocks by Years

Outliers play an important role in analysis therefore box plot is plotted for this data as shown in figure 6. Since it is the time series analysis of past data therefore the removal of outlier is not viable because in stocks trend variations up to a greater extent is normal.

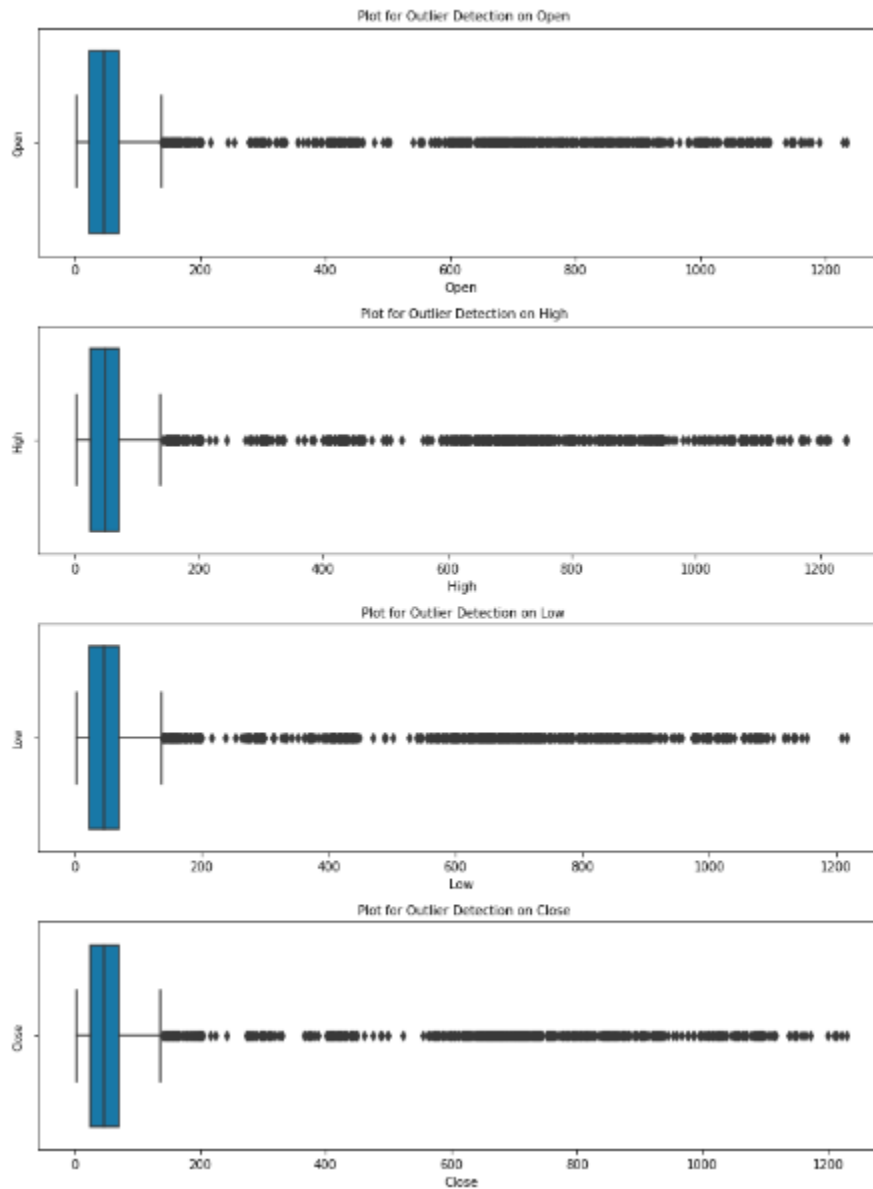


Figure 6 Outliers in the data distribution

The volume of the shares is the number of shares exchanged among the various traders therefore it is essential to analyse the volume of the stocks being exchanged. An analysis is performed on the volumes according to the years and the month as shown in the figure 7 and 8.

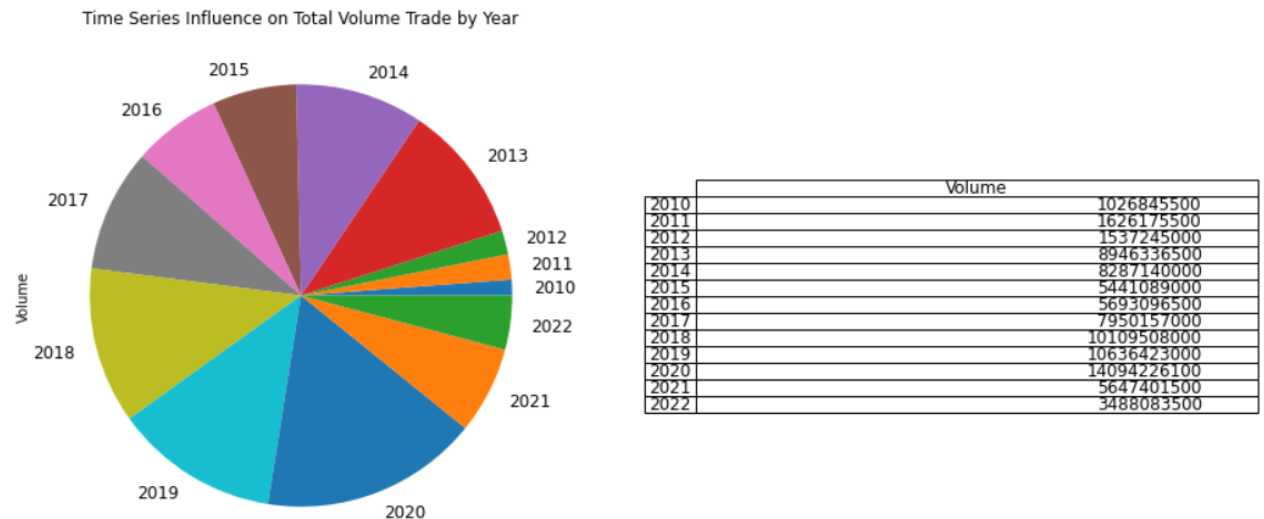


Figure 7 Total Volume analysis by Years

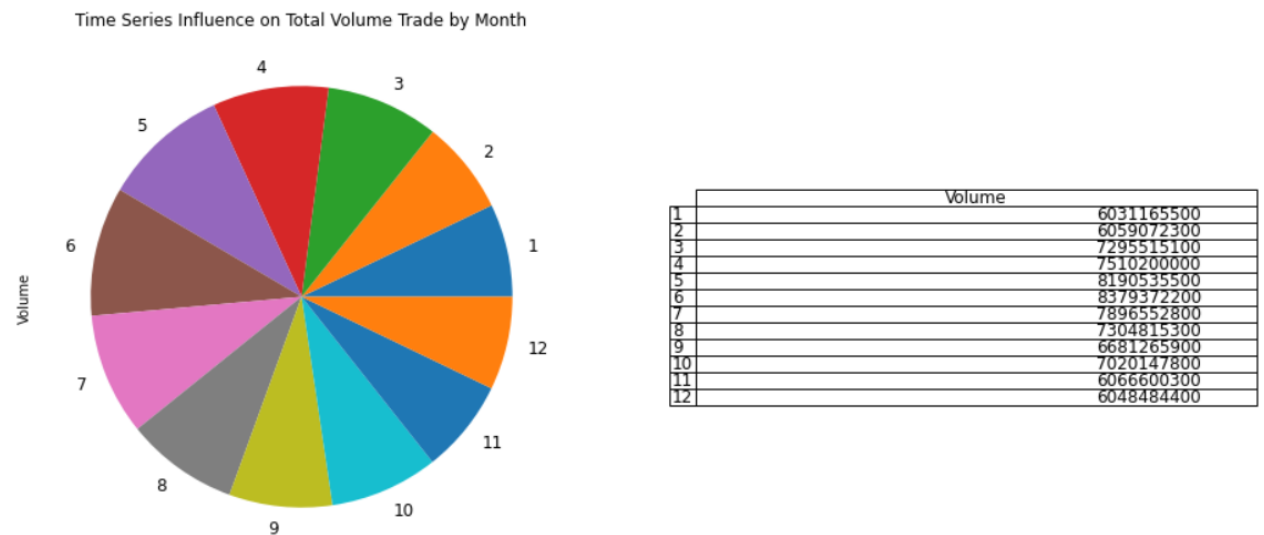


Figure 8 Total Volume analysis by Months

3.1.4 Feature Engineering

Features are the input in the machine learning or deep learning algorithm. Based on these inputs mapping is executed by the algorithm in order to predict the target variable. Therefore, Feature Engineering is an important step where feature analysis, feature extraction, and feature selection are performed. Here the first descriptive statistical analysis is performed on the data and each feature is transformed into float32 type. Since all features are in numerical format but not in the common range, therefore, Minmax Scalar has been implemented to range all the feature values from 0 to 1. This is a multi-class regression task therefore the target variables are open, close, high, and

low columns. The remaining all other and these columns too are the feature variables or the input to the model.

3.1.5 Model Training and Testing

After executing data pre-processing and feature engineering the data is available in the required format. In a time series analysis of the Tesla Stock prices three advanced deep learning algorithms are executed which are the Conv-RNN algorithm, Conv-LSTM algorithm, and Conv-BILSTM algorithms. In order to train these algorithms, the whole data is splitted into the training and test data in the ratio of 90 to 10. This training data is fitted into each model for training purposes and after training each model is tested on the test data. Each model is trained for 100 epochs.

3.1.6 Model Evaluation

Since the target value is the open, close, high, and low columns which have values of continuous nature therefore it is a regression-based task where the predicting column has continuous values. For evaluation of the executed models, the regression-based metric is used which is the mean absolute Error value (MAE). In time series analysis and prediction of the Tesla stock prices the best model is identified that scored the lowest value of MAE. for better understanding and in-depth analysis, a line plot is plotted between the real stock price and the predicted stock price.

3.2 Sentiment Analysis to predict Tesla Stock Prices

It is observed that the market trends in the share market are highly affected by the news and Twitter posts because it integrates the different views of all investors. Therefore, in order to predict the share price of Tesla impacted by the Twitter tweets is also executed. Sentimental analysis of the Twitter posts is executed using Natural language Processing, and machine learning, and deep learning algorithms are applied further to predict the change in share prices due to posted tweets. The framework for forecasting the tesla stock prices using sentiment analysis is shown with the help of flow diagram in Figure 9.

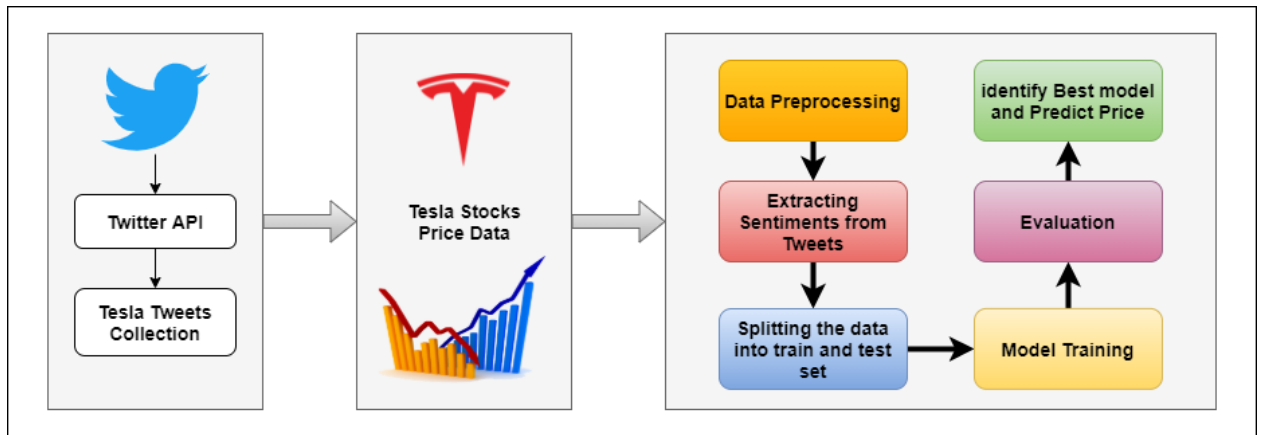


Figure 9. Framework for Forecasting Tesla Stock Prices using Sentiment Analysis from tweets

3.2.1 Data Set Description

Stock prices of any entity show high variation in the change which is usually affected by the news and the tweets posted by the private or governmental agencies therefore, the sentimental analysis of these posts really impacts the trends of the stocks prices and can help in making more better predictions on the prices of the stocks. The 15 days' worth of tweets concerning Tesla was retained for this project work using the Twitter API which is dated from 6 June 2022 to 20 June 2022. The keyword, content, ConversationId, date, retweet-Count, and tweet-URL are just a few of the six attributes that make up this data collection. The price of Tesla stocks for the exact same days has also been gathered and retrieved from Investopedia.com. Date, Price, Open, High, Low, Volume, and Percentage Change are just a few of the six columns this data collection contains.

3.2.2 Data Pre-processing

Data pre-processing is indeed a crucial component in NLP and machine or deep learning implementation, in order to remove noise from the dataset and enable better analysis. The steps involved in pre-processing data include cleanliness, generalizations, removing null values, etc. Such attributes like ConversationId, date, retweet-Count, and tweet-URL are removed from Twitter data because these are not really relevant. Consequently, the first step is to remove any unneeded columns before changing the names of the columns appropriately. Additionally, we have looked for and removed

null values. After completing all of these processes, the Date column's data type is changed to date-time format, allowing for the extraction of only dates out of each row. We have combined all tweets from the same day into a single row for that specific day. After pre-processing the Data from Twitter, we processed the stock price data by eliminating noise, converting the dates to Date Time format, and then giving a price value based on the exact dates of the Tweets. The third step entails adding columns for emotional assessment to the processed dataset and filling in adjacent missing values of share price in accordance with the pattern in the stock price.

3.2.3 Feature Engineering

Attributes are transformed from original data to meaningful data for supervised methods. We just employed the close attribute as the target variable from the share price data in order to do a better prediction. Other attributes that must be deleted include keywords, ConversationId, Retweet-Count, Tweet-URL, and others. We combined both datasets that relate to the same day and concatenated all tweets from that date. We also designed a new functionality in accordance with the requirements. We applied the NLTK package (Vader Lexicon) to identify the emotions from the Twitter posts, and as a result, three new columns (Negative, Positive, and Compound) are introduced to the dataset to contain the emotions according to the same timeframes.

3.2.4 Exploratory Data Analysis

Since the total sample only contains 15 rows of data covering 15 days and 5 characteristics, we also looked at the frequency of each variable in the data. This analysis is crucial because it will stop models from producing conclusions that are slanted if a particular value occurs frequently. Bar plots have been created to show the features as shown in Figures 10, and 11.

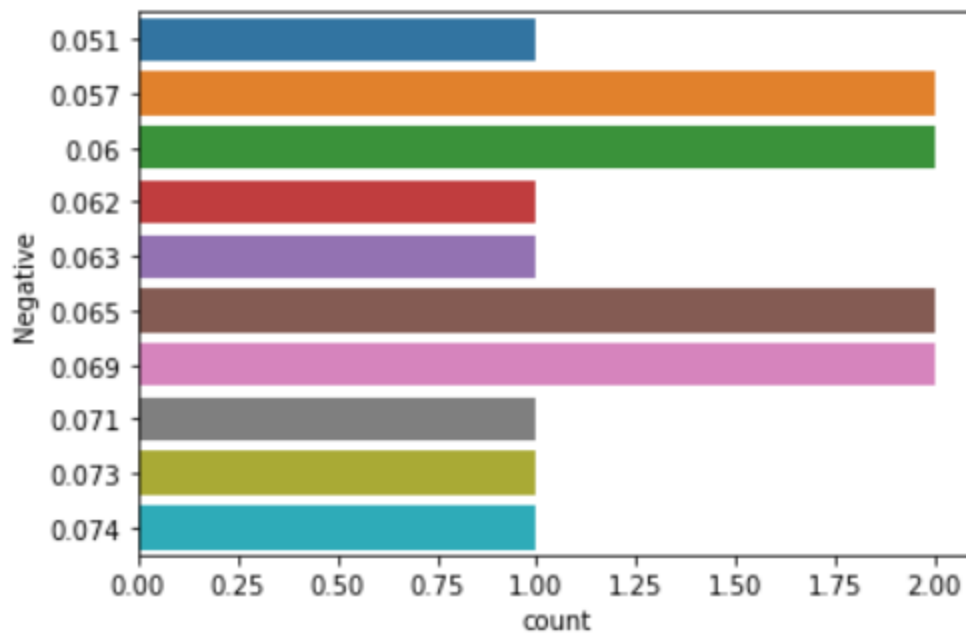


Figure 10 Negative sentiments for Training data

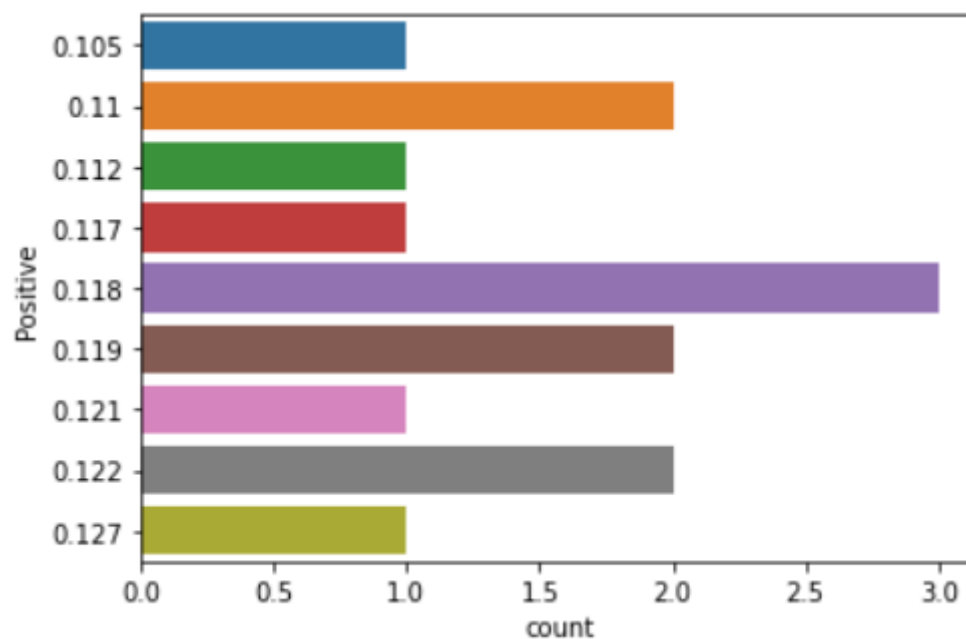


Figure 11 Positive sentiments for Training data

3.2.5 Model Training and Testing

In order to predict, the price of the Tesla Stocks two machine learning and one deep learning algorithm is implemented. Since this data contains the date as the feature and based on the past data predictions are to be made on future dates therefore first ten dates are selected as the training data and the remained dates are taken as the test data. In the machine learning domain, random forest and XgBoost algorithms are executed while

in the deep learning domain the deep neural network is implemented. These all algorithms are trained on train data and then tested on the test data.

3.2.6 Model Evaluation

Every algorithm has been assessed in order to choose the most effective one for Twitter sentiment analysis-based Tesla stock price prediction. This is a regression challenge since it involves predicting the stock price using sentiment analysis of Twitter posts. We have employed MAE (Mean Absolute Error), which quantifies the overall size of the divergence between an observation's real value and its forecast. All deployed models are analysed using the test MAE score.

Chapter 4. System Design and Specifications

In this research task, the time series-based analysis and prediction of the Tesla stocks are executed by implementing three advanced deep learning-based algorithms which are Conv-RNN, Conv-LSTM, and Conv-BiLSTM while the forecasting of Share prices based on the Sentimental analysis of the Twitter tweet posts is executed by implementing two machine learning and one deep learning algorithm which are Random Forest, Catboost and Deep neural Network algorithms. All implemented algorithms are discussed in brief in further sections.

4.1 Convolutional-Recurrent Neural Network Model

The Convolution-Recurrent Neural Network (Conv-RNN) represents a combination that contains two most specific and prominent “neural networks”. This Con-RNN typically involves a convolutional neural network (CNN) and recurrent neural network (RNN). Integrating knowledge of its composition, the network model consists of 2-dimensional CNN architecture with batch normalization, “ELU activation”, max-pooling & dropout, where the dropout rate is 50%. This can be identified as a proposed model where 3 significant convolutional layers have been placed by sequence to the corresponding activations. The layers certainly follow permute & reshape layer, which is essential for Conv-RNN. This is because of the features that differ between the two neural networks (CNN & RNN). Thus, the CNN architecture finally developed into 3D feature vectors while the RNN layers were typically designed into 2D feature vectors. Conv-RNN architecture is shown in Figure 12.

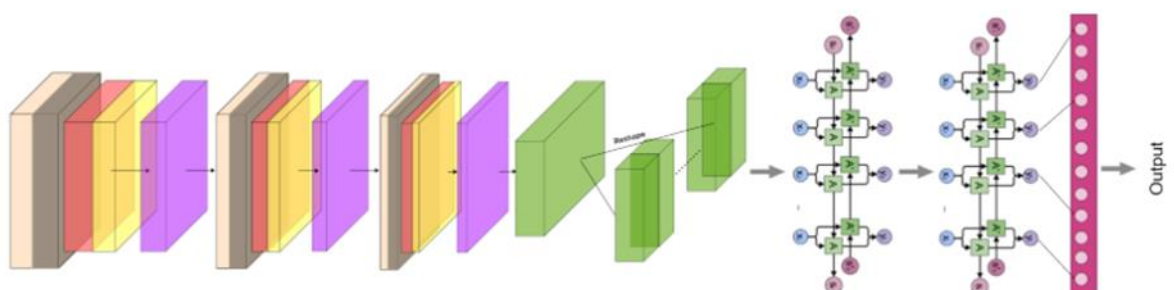


Figure 12 Conv-RNN Algorithm

4.2 Convolutional-Long Short-Term Memory Model

The Convolutional- Long Short-Term Memory Network (Conv-LSTM) represents an LSTM architecture specifically designed to solve sequence prediction issues through spatial input arrangements, such as video and images. This architecture generally involves two network models, i.e., Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM). The CNN is used to form specific layers to extract features on the input data through a combination of LSTMs that can support sequence predictions. It was developed to establish visual time-series prediction issues as well as the generation of textual descriptions via image sequences. Thus, it can be stated that the network architecture has helped to dive deeper into predicting problems through visual representation and text mining and thereby serves as a real-time solution in different prediction activities. Conv-LSTM Model architecture is shown in Figure 13.

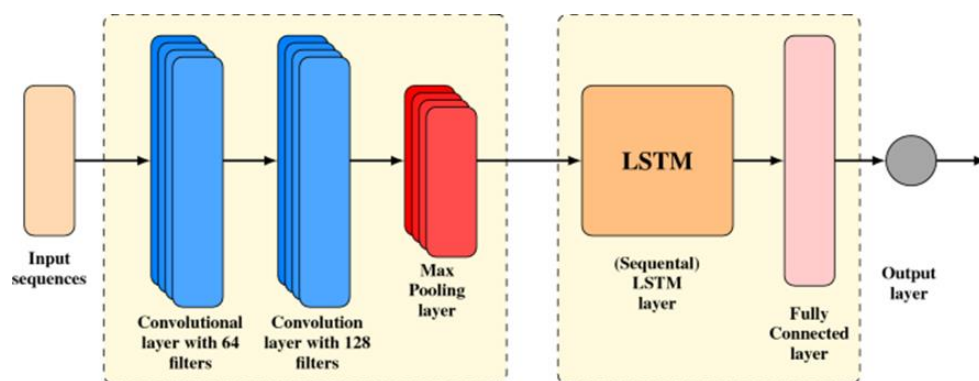


Figure 13 Conv-LSTM Model

4.3 Convolutional- Bidirectional Long Short-Term Memory Model

The Convolutional Neural Network-Bidirectional Long Short-Term Memory (Conv-BiLSTM) represents an architecture where two specific models have been used to serve the purpose of different prediction activities. Various proposed models were used based on this architecture which has gained attention through different studies. For instance, the real-time observation of a driver's emotions has significantly improved traffic safety. Herein, the recognition or prediction of this emotion can be done through a deep learning architecture, specified as Conv-BiLSTM. It has helped in predicting drivers'

emotions while maintaining traffic safety. This recognition depends on their geometric features which are extracted from the facial information as well as heart rate. Herein, the suggestive BiLSTM output is preferably used as the input for the convolutional module for extracting the heart-rate features. Based on the result, the model has proved its efficiency in quick and steady recognition of emotions. Thus, through this illustration, it can be predicted that the model has greatly served its importance through real-time solutions. Conv-BiLSTM working mechanism is shown in Figure 14.

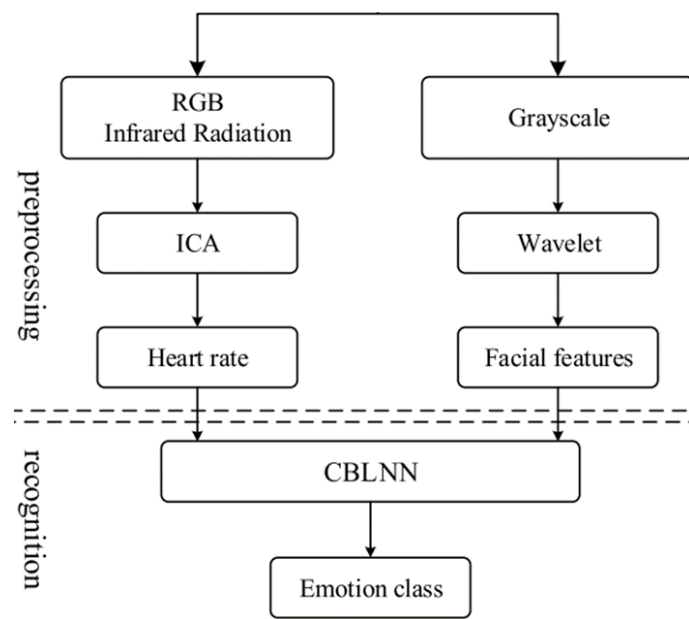


Figure 14 Conv-BiLSTM Algorithm

4.4 Random Forest Regressor

Random Forest (RF) represents a popular algorithm model of machine learning which basically represents a “supervised learning technique”. The well-known machine learning model can be used for both the purpose of Classification & Regression problems. The concept of the model is highly based on ensemble learning where a combined classifier is used to solve complicated issues and thus improve the performance criteria. Coming to its composition, the classifier contains a certain number of “decision trees”, categorized as subsets of the test dataset where an average criterion is used for improving the prediction accuracy level of the trained dataset. The main advantage of using a random forest classifier comes with its efficiency for the

following purposes it requires less training time, the high prediction accuracy for both small and large data, and maintains a continuum in the accuracy level even after a large amount of missing data. The architecture of Random forest regressor algorithm is shown in Figure 15.

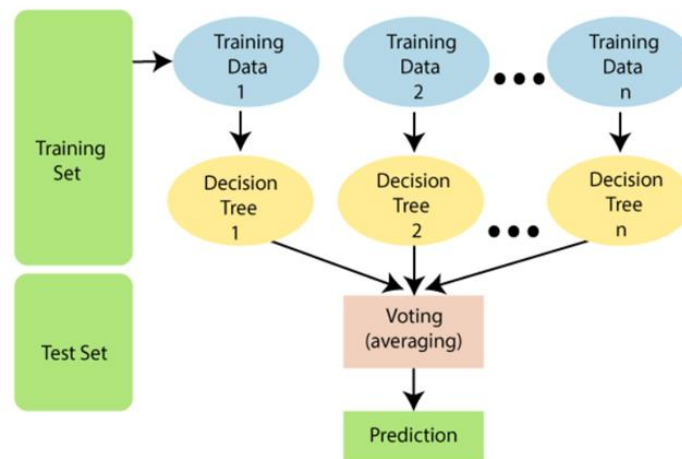


Figure 15 Random Forest Regressor Algorithm

4.5 Catboost Algorithm

Catboost algorithm mainly works with the suitability of Gradient Descent, one of the powerful techniques in classification & regression problems. It was developed by renowned Yandex researchers & engineers, which in the later time become a successor to the “MatrixNet algorithm”. It is one of the well-recognized algorithms which have served its purposes in solving problems related to categorical data. As it is mentioned previously, the classification, as well as regression problems, is suitably handled with this approach; the model seems to be a well-known algorithm in the advent of Kaggle competitions. The main working principle is based on using various decision trees which are further trained to serve the purpose. The most advantageous benefit served by this algorithm is by reducing the training time, thereby indicating its faster prediction rate with benchmarked results compared to other algorithms. Architectural diagram of catboost is shown in Figure 16.

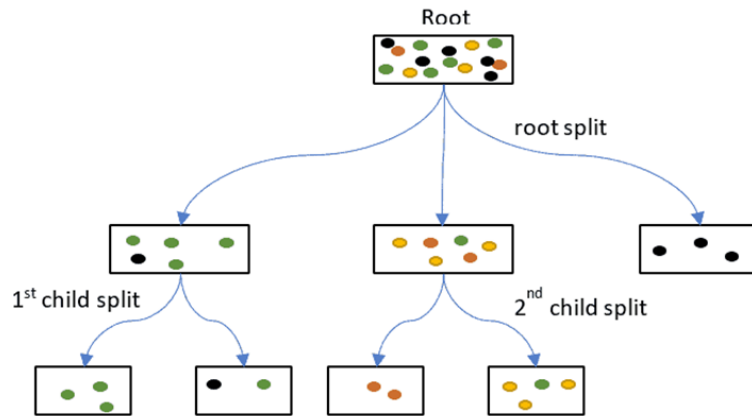


Figure 16 Catboost Algorithm

4.6 Deep Neural Network

Deep Neural Network (DNN) represents the artificial neural network architecture with excess depth in layers, preferably increased hidden layers. The layers are typically found between input & output layers. It has well-served its importance with promising solutions injected into artificial intelligence (AI) that meet the purpose of different daily life events. The DNN framework generally accelerates through server-equipped and enhanced computing engines, such as graphics processing units (GPU); however recent advancement in technology significantly requires the “energy-efficient acceleration” of a deep neural network to serve the modernized applications. These applications have suggestively forwarded to mobile computation nodes. Deep learning has successfully served a wide range of purposes with its emergence in recent times. It, in fact, becomes an integral part of the digital world in different sectors. Layered diagram of Deep neural network algorithm is shown in Figure 17.

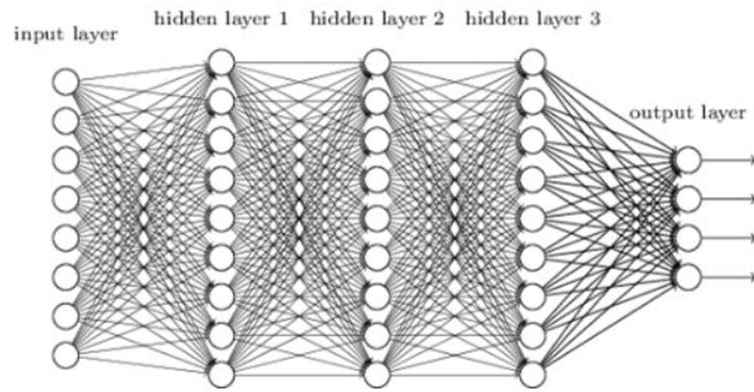


Figure 17 Deep Neural Network Algorithm

Chapter 5. Implementation

In order to make a time series analysis of the Tesla stocks prices, three advanced deep learning-based algorithms are incorporated in this research which are RNN, Conv-LSTM, and Conv-BiLSTM. These models are implemented by using Keras and TensorFlow. All models are trained on training data for 100 epochs and then tested on test data. Each algorithm is optimized using Adam optimizer and mean absolute error (MAE) as the loss function. The superlative model can identify which has scored the minimum value of MAE on the test data among all other models implemented. For the prediction of stock prices based on the sentimental analysis of the Twitter tweet posts two machine learning and one deep learning algorithm are executed. These are Random Forest and Catboost from the machine learning domain while a Deep neural network is a deep learning domain algorithm. Each algorithm is first trained on the training data and then tested on the test data. Each algorithm is evaluated by calculating the MAE score on test data. The best model capable of predicting the Tesla stock prices based on sentimental analysis will be the model which scored the lowest MAE score among all other algorithms. In order to obtain the sentiments from the Twitter tweets texts, Natural Language Processing is utilized. There are numerous libraries that have been employed to implement the research work which include NumPy, matplotlib, pandas, seaborn, Sklearn, TensorFlow, Keras, NLTK, etc. The whole research work is performed in Jupyter Notebooks for training purposes with python as the programming language. The following libraries and specifications are required in order to implement the algorithms.

- Operating System: windows 10
- Random Access Memory (RAM): 16GB
- Hard disk: 1 TB
- Language: Python
- Platform: Anaconda
- Framework: Jupyter Notebook
- Libraries: NumPy, Pandas, matplotlib, Sklearn, NLTK, TensorFlow, NumPy, seaborn, and Keras.

Chapter 6. Testing and Evaluation

Evaluation of the executed models is important because it allows identifying the best performing model so that it can be deployed in the real world for making accurate predictions. For the time series-based analysis MAE, a regression-based metric is employed for the evaluation of all three deep learning models while in sentimental analysis based on Tesla share price prediction which is a regression-based task therefore MAE metric is used to evaluate the implemented machine learning and deep learning models. After evaluation, a model is selected which can predict the Tesla stock prices in real-world tasks.

6.1 Experimentation for time series analysis and prediction of Tesla Stocks

6.1.1 Experiment 1 / Evaluation of the Conv-RNN model

In the first experiment, a simple RNN model is constructed from the scratch in order to predict the time series-based output of the Tesla stock prices. This model is trained on the training data which consists of 2752 samples. The learning rate for this model is also evaluated using a learning rate scheduler. The optimal learning rate for stochastic gradient optimizer is $10e-6$ with Huber loss and MAE as the evaluation metric to make predictions on the open, high, low, and close values of the Tesla stock price. Further, this model is tested on the test data which contains 242 samples and MAE is calculated on this test data, and values scored by this model for high, low, open, and close stock prices are 34.03, 33.99, 48.78 and 33.50 respectively. A line plot is also plotted between real and predicted prices as shown in Figure 18, Figure 19, Figure 20, and Figure 21.

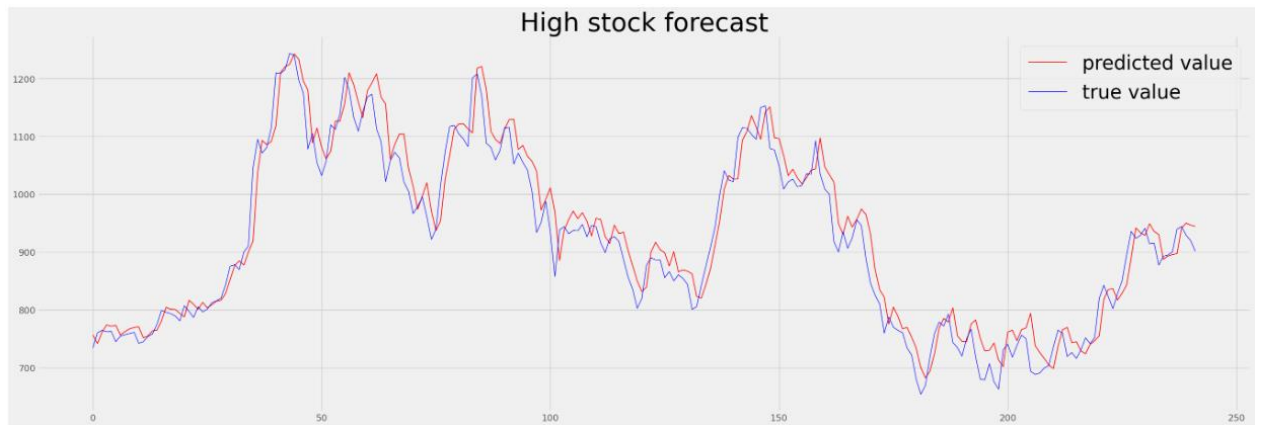


Figure 18 High Stocks Value Forecast by Conv-RNN

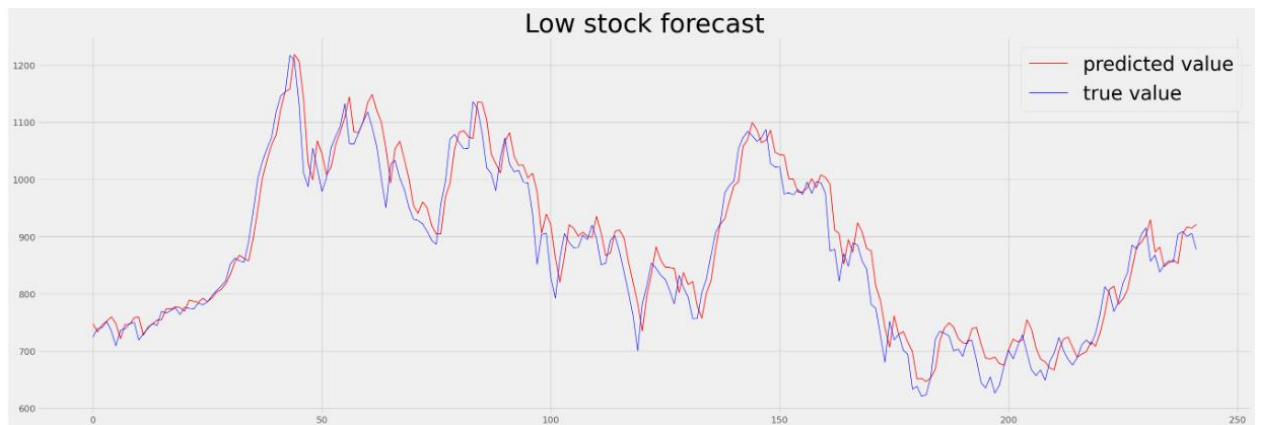


Figure 19 Low Value Stocks Value Forecast by Conv-RNN

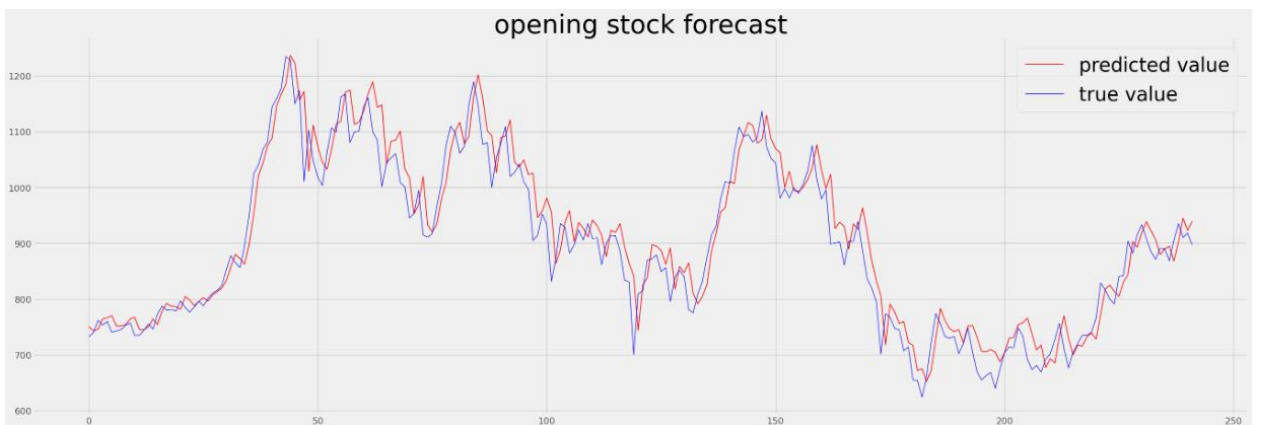


Figure 20 Open Value Stocks Value Forecast by Conv-RNN



Figure 21 Closing Value Stocks Value Forecast by Conv-RNN

6.1.2 Experiment 2 / Evaluation of the LSTM model

In the second experiment, a Conv-LSTM model is built from the scratch using convolutional and LSTM layers. This model is trained on the training data which consists of 2752 samples. The learning rate for this model is also evaluated using a learning rate scheduler. The optimal learning rate for stochastic gradient optimizer is $10e-7$ with Huber loss and MAE as the evaluation metric to make predictions on the open, high, low, and close values of the Tesla stock price. Further, this model is tested on the test data which contains 242 samples, and MAE is calculated on this test data and values scored by this model for high, low, open, and close stock prices are 33.52, 33.98, 34.47, and 34.25 respectively. A line plot is also plotted between real and predicted prices as shown in Figure 22, Figure 23, Figure 24, and Figure 25.

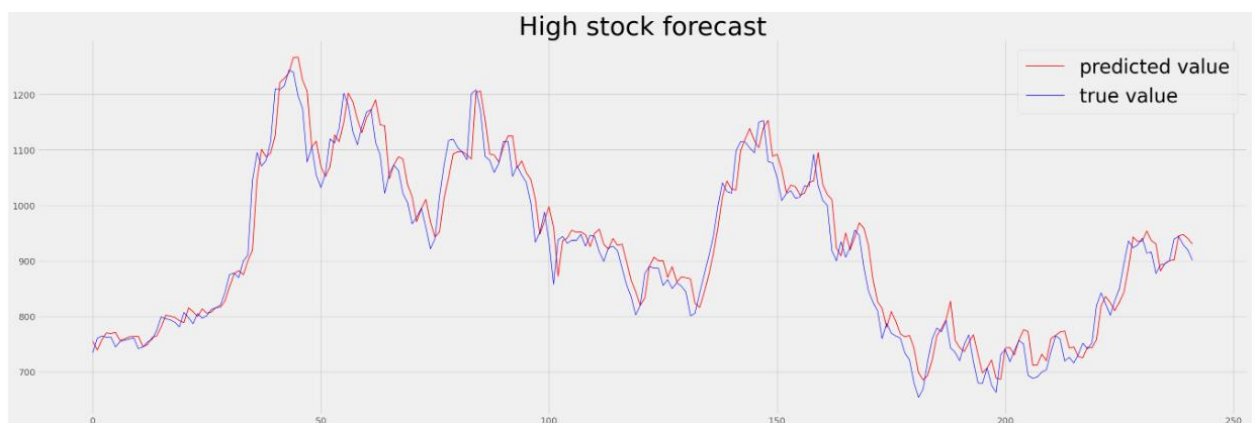


Figure 22 High Value Stocks Value Forecast by Conv-LSTM

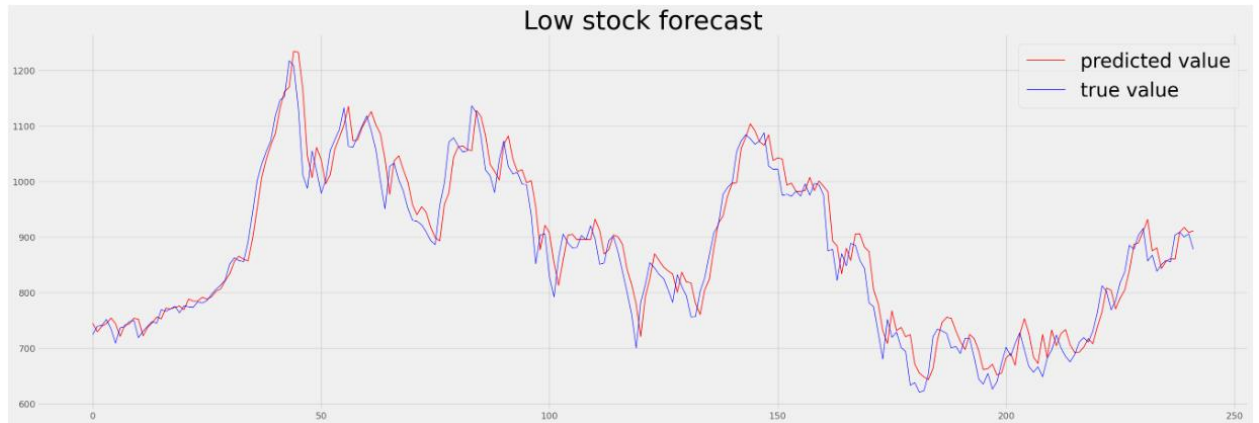


Figure 23 Low Value Stocks Value Forecast by Conv-LSTM



Figure 24 Opening Value Stocks Value Forecast by Conv-LSTM



Figure 25 Closing Value Stocks Value Forecast by Conv-LSTM

6.1.3 Experiment 3 / Evaluation of Conv-BILSTM model

In the third experiment, in order to predict the prices of the stocks regarding high, low, close, and open a Conv-Bilstm model is implemented which is trained over training data which contains 2752 samples. After complete training, the model is tested on the test data, and the MAE value is calculated for high, low, open, and close which come

out to be 37.44, 38.15, 34.47, and 42.45 respectively. For better training of the model the learning rate scheduler is executed with hubber loss and MAE as evaluation metrics. The best learning rate for this model is 1e-6. A line plot is also plotted between real and predicted values as shown in Figures 26, Figure 27, Figure 28, and Figure 29.

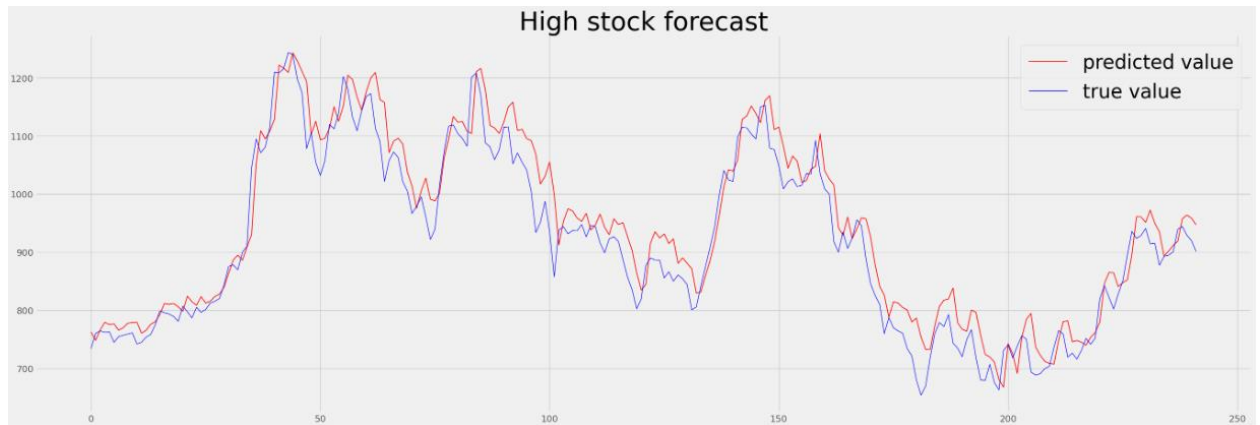


Figure 26 High Value Stocks Value Forecast by Conv-BILSTM

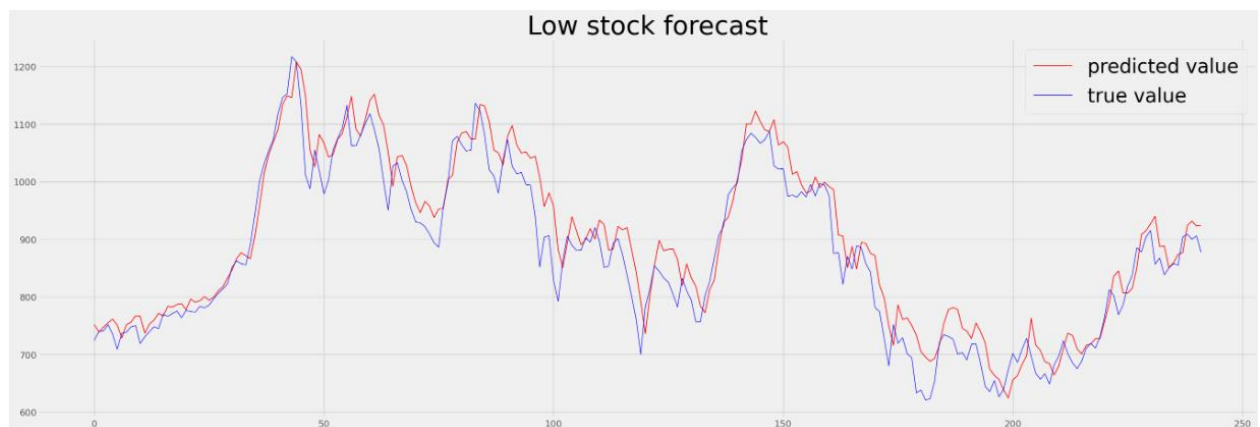


Figure 27 Low Value Stocks Value Forecast by Conv-BILSTM



Figure 28 Opening Value Stocks Value Forecast by Conv-BILSTM



Figure 29 Closing Value Stocks Value Forecast by Conv-BILSTM

6.2 Experimentation for sentimental analysis of tweets and prediction of Tesla Stocks

6.2.1 Experiment 1 / Evaluation of the Random Forest model

In the first experiment, to predict the stock prices based on the sentimental analysis of the Twitter posts, which generally underlays the trends of the uprising or downfall of the market, the random forest algorithm is executed. This algorithm is first trained on the training data which contains the positive and negative sentiments as the features for the first 10 days and the last 5 days' data is used for testing purposes. The Close value of the stocks is taken as the target column and the MAE value is calculated for the evaluation. The MAE value scored by this algorithm is 116.19. A bar plot is plotted to make the comparison between the actual closing price and the predicted closing price as shown in Figure 30.

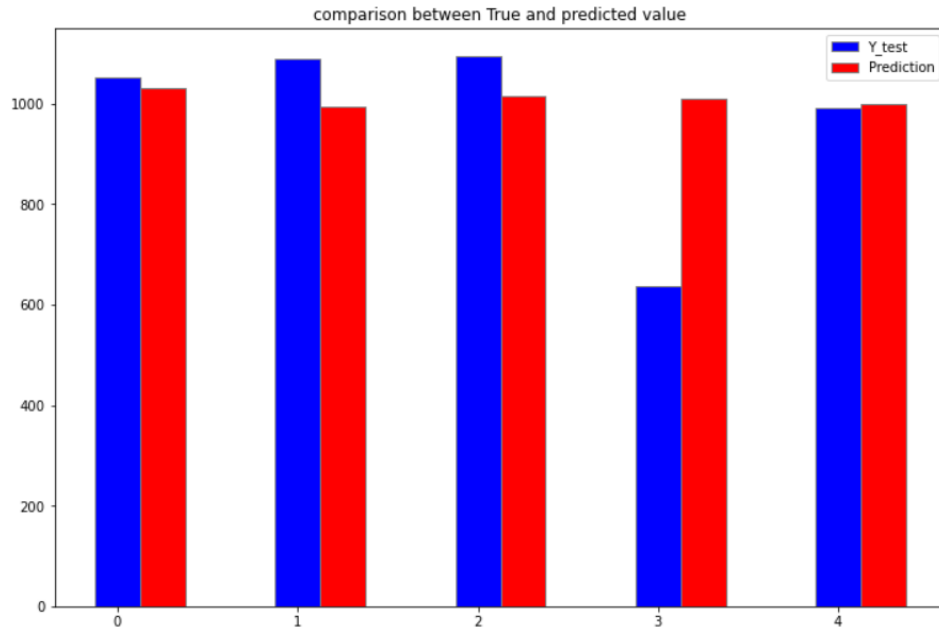


Figure 30 Comparison between the actual closing price and the predicted closing price using Random Forest Algorithm

6.2.2 Experiment 2 / Evaluation of the Catboost model

In the second experiment, the Catboost algorithm is executed to predict the Tesla stock prices based on the sentimental analysis of the Twitter posts about tesla. This model is first trained on the training data which contains the data for the first initial 10 days and the remaining 5 days of data are used for the testing data. This test data is used for the evaluation of the MAE score evaluation. The MAE value scored by this algorithm is 111.20 which least among all other implemented algorithms. A bar plot is plotted to make a comparison between actual closing prices and predicted closing prices as shown in Figure 31.

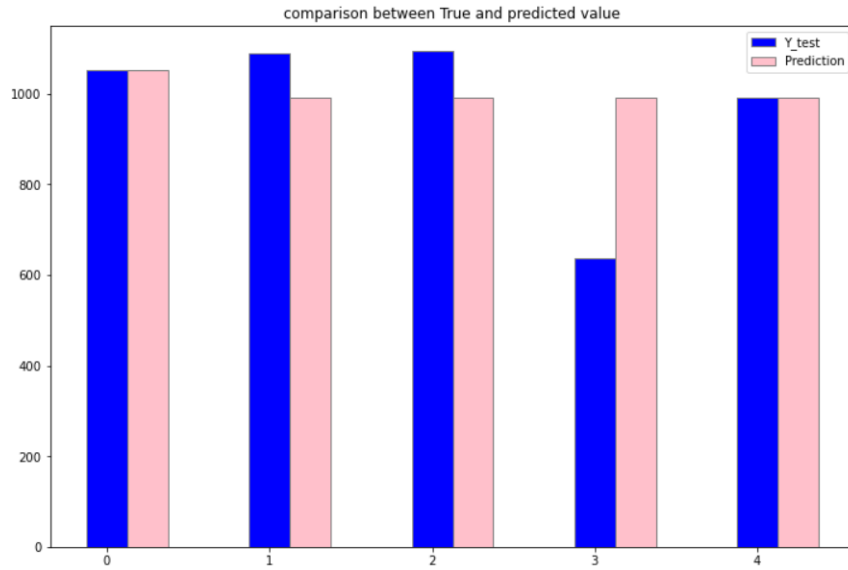


Figure 31 Comparison between the actual closing price and the predicted closing price using CatBoost Algorithm

6.2.3 Experiment 3 / Evaluation of the DNN model

In the third experiment, the deep learning-based deep neural network is built using dense layers and dropout layers to avoid overfitting. This algorithm is trained on the training data which contains the first 10 days' data and the last 5 days' data is used for the testing of the model. This algorithm is trained for 100 epochs but due to less data the model shows overfitting and in evaluation, the MAE value scored by this algorithm is 177.54 which is the highest among all other algorithms. The reason for overfitting and poor performance than other models is due to the small size of the data. To make a comparison between predicted and real share prices a bar plot is plotted which is shown in Figure 32.

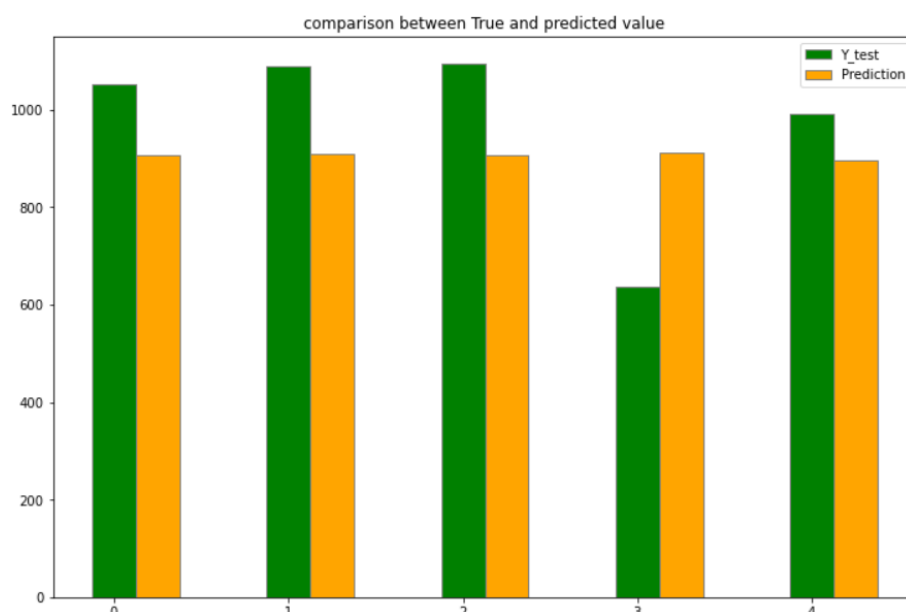


Figure 32 Comparison between the actual closing price and the predicted closing price using Deep Neural Network

6.3 Discussion

Time-series analysis and sentiment analysis on data from Twitter where the two forms of study applied in this research to forecast the value of Tesla stocks. It is observed that the Conv-LSTM algorithm beats all other algorithms with the lowest MAE value after a particular set of experiments for time series analysis. The time series prediction plot of Conv-LSTM demonstrates that our algorithm is performing as expected as it primarily coincides with the curve of actual value pricing. As seen by the displayed lines, the Conv-Bilstm algorithms perform less well than the Conv-RNN model and Conv-Lstm model. MAE value is also calculated for every model in addition to that. The algorithm with the lowest MAE value will be deemed the best algorithm for predicting Tesla stock prices. Conv-LSTM has the lowest MAE value, trailed after Conv-RNN and Conv-Bilstm. As a result, it can be concluded that the Conv-Bilstm algorithm performs inadequately in time-series analysis when compared to the Conv-LSTM and Conv-RNN models. MAE graph is plotted as shown in figure 33 which depicts the comparison of MAE for all implemented models for predicting the close, open, high, and low values of Tesla stocks. After conducting a series of tests & evaluating the MAE value over the testing data, it is concluded that Catboost is the highest performing for forecasting stock price since it produces the lowest MAE value.

However, because our actual data is quite limited but only comprises 15 days of Twitter posts, it is not necessarily essential for Catboost to offer the best results. This is also the reason why deep neural network does not get good prediction in comparison to other models because deep neural networks require large data for training otherwise, they tend to underfit or overfit which is also observed in our analysis. Even more complicated structures can yield better results for huge datasets. The Catboost algorithm surpasses the Random Forest and DNN algorithm with an MAE value of 111.20, and DNN in our scenario performed inadequately on the 15 days of Twitter data for Tesla stocks price prediction, according to a comparison of the MAE scores of the other two approaches. The obtained MAE for Twitter sentiment analysis is shown in Figure 34.

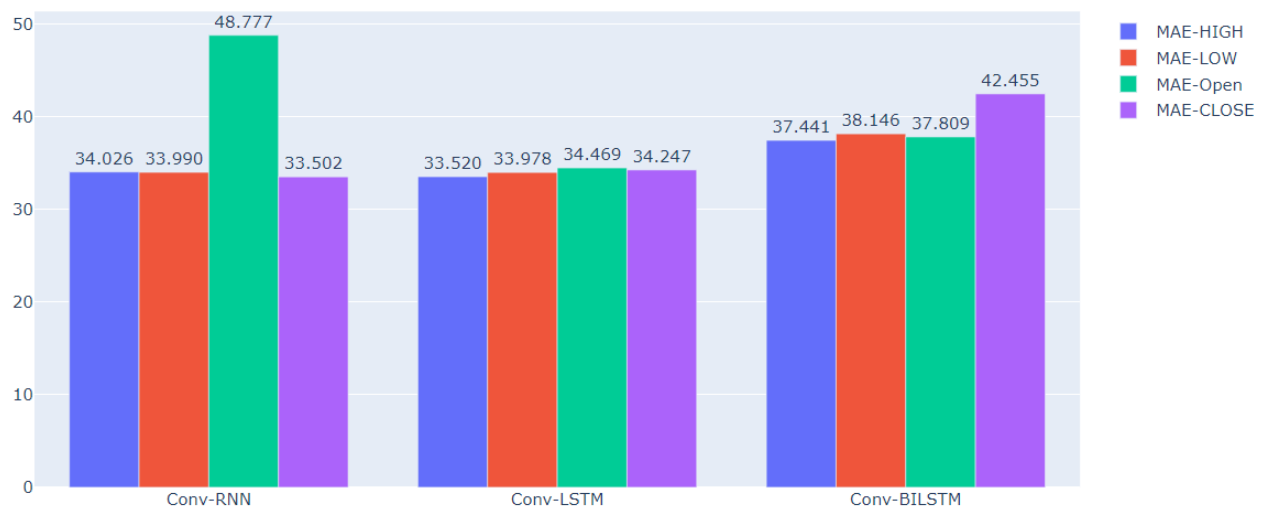


Figure 33 Comparison Of all Models executed for predicting stocks based on Time Series Analysis

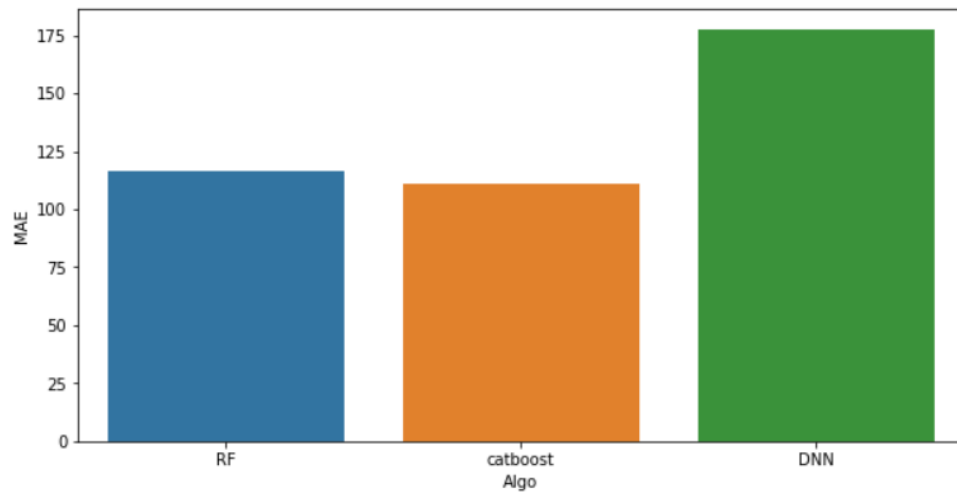


Figure 34. Comparison Of all Models executed for predicting stocks based on sentimental analysis

Chapter 7. Conclusions and Future Work

Since stock markets are often very volatile and completely reliant on the underlying conditions of the many sectors, businesses, and their need, predicting the value of Tesla shares is a tough challenge and yet a relatively unexplored field of study. For successfully predicting the value of stock, several researchers have put forth various models and methodologies. Using the techniques of time series modelling and sentiment classification on Twitter, we tried to forecast the value of the shares in this study. The value of Tesla shares during the previous 12 years has been compiled and forecasted for the time-series study. For the Twitter-based sentiment analysis, the most recent 15 days' worth of posts were gathered, and the sentiments rating was used to forecast the price of Tesla stocks. In this research work, three separate methods that we have employed in both methods. Throughout a battery of experiments, it was discovered that the Conv-LSTM method performs better in terms of forecasting for time-series analysis than the Conv-RNN and Conv-BiLSTM model. For the sentiment-based analysis of Twitter the better outcomes are obtained using the Catboost algorithm than random forest and deep neural network. Only the data from the last 15 days of Twitter posts have been used for the ongoing study. However, leveraging the Twitter API, additional days of data may be gathered for improved prediction. Another difficult issue is gathering data via the Twitter API because of the poor responsiveness and lengthy data collection process. Twitter mostly provides information about people's perspectives about shares and market trends. By obtaining these views, the financial sector may get knowledge about how the market behaves, which can help them make more money. Due to the limited computational capabilities and time constraint, only 15 days of twitter data was collected for sentiment analysis. In future work, a large dataset can be gathered which can further enhance the outcomes of model prediction.

References

- [1]A. Altuner and Z. Kilimci, "A Novel Deep Reinforcement Learning Based Stock Direction Prediction using Knowledge Graph and Community Aware Sentiments.", 2021. Available: 10.48550/arxiv.2107.00931 [Accessed 10 September 2022].
- [2]N. Belo, J. Erker, M. Koehler and J. Niehues, "Financial Decision Making based on Social Media Sentiment Analysis using Transformers.", 2021. Available: <https://www.kaggle.com/cosmos98/twitter-and-reddit-sentimental-analysis-dataset>. [Accessed 10 September 2022].
- [3]S. Bouktif, A. Fiaz and M. Awad, "Augmented Textual Features-Based Stock Market Prediction", *IEEE Access*, vol. 8, pp. 40269-40282, 2020. Available: 10.1109/access.2020.2976725 [Accessed 10 September 2022].
- [4]S. Das, R. Behera, M. kumar and S. Rath, "Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction", *Procedia Computer Science*, vol. 132, pp. 956-964, 2018. Available: 10.1016/j.procs.2018.05.111 [Accessed 10 September 2022].
- [5]G. Deza, C. Rowat and N. Papernot, "On the Robustness of Sentiment Analysis for Stock Price Forecasting.", 2021. [Accessed 10 September 2022].
- [6]B. Erkartal and A. Yılmaz, "Sentiment Analysis of Elon Musk's Twitter Data Using LSTM and ANFIS-SVM", *Lecture Notes in Networks and Systems*, pp. 626-635, 2022. Available: 10.1007/978-3-031-09176-6_70 [Accessed 10 September 2022].
- [7]B. Gustav Edman and M. Weishaupt, "Predicting Tesla Stock Return Using Twitter Data An Intraday View on the Relation between Twitter Dimensions and the Tesla Stock Return.", 2020. [Accessed 10 September 2022].
- [8]T. Ho and Y. Huang, "Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation", *Sensors*, vol. 21, no. 23, p. 7957, 2021. Available: 10.3390/s21237957 [Accessed 10 September 2022].
- [9]K. Hu, D. Grimberg and E. Durdyev, "Twitter Sentiment Analysis for Predicting Stock Price Movements.", 2022. [Accessed 10 September 2022].
- [10]R. Karlemstrand and E. Leckström, "Using Twitter Attribute Information to Predict Stock Prices.", 2021. Available: 10.48550/arxiv.2105.01402 [Accessed 10 September 2022].
- [11]L. Khaidem, S. Saha and S. Dey, "Predicting the direction of stock market prices using random forest", *arXiv.org*, 2016. [Online]. Available: <http://arxiv.org/abs/1605.00003>. [Accessed: 10- Sep- 2022].

- [12]S. Lee, "Feature Investigation for Stock Returns Prediction Using XGBoost and Deep Learning Sentiment Classification.", *CMC Senior Theses.*, 2021. Available: https://scholarship.claremont.edu/cmc_theses/2715. [Accessed 10 September 2022].
- [13]R. Mendoza-Urdiales, J. Núñez-Mora, R. Santillán-Salgado and H. Valencia-Herrera, "Twitter Sentiment Analysis and Influence on Stock Performance Using Transfer Entropy and EGARCH Methods", *Entropy*, vol. 24, no. 7, p. 874, 2022. Available: 10.3390/e24070874 [Accessed 10 September 2022].
- [14]S. Mndawe, B. Paul and W. Doorsamy, "Development of a Stock Price Prediction Framework for Intelligent Media and Technical Analysis", *Applied Sciences*, vol. 12, no. 2, p. 719, 2022. Available: 10.3390/app12020719 [Accessed 10 September 2022].
- [15]V. Pagolu, K. Reddy, G. Panda and B. Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements.", *International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings.*, pp. 1345–1350, 2016. Available: 10.48550/arxiv.1610.09225 [Accessed 10 September 2022].
- [16]M. Parry, "Predicting the direction of stock price movements from historical data and sentiment analysis of Tweets.", 2019. [Accessed 10 September 2022].
- [17]B. Pavlyshenko, "Bayesian Regression Approach for Building and Stacking Predictive Models in Time Series Analytics", *arXiv.org*, 2022. [Online]. Available: <http://arxiv.org/abs/2201.02034>. [Accessed: 10- Sep- 2022].
- [18]B. Pavlyshenko, "Forming Predictive Features of Tweets for Decision-Making Support", *arXiv.org*, 2022. [Online]. Available: <http://arxiv.org/abs/2201.02049>. [Accessed: 10- Sep- 2022].
- [19]K. Pawar, R. Jalem and V. Tiwari, "Stock Market Price Prediction Using LSTM RNN. Advances in Intelligent Systems and Computing", no. 841, pp. 493–503, 2019. Available: 10.1007/978-981-13-2285-3_58/COVER [Accessed 10 September 2022].
- [20]R. Puri, J. Bhandari, R. Gupta and A. Sinha, "Time Series Analysis and Trend Exploration of Stock Market", *Advanced Computational Techniques for Sustainable Computing*, pp. 197-212, 2022. Available: 10.1201/9781003046431-14 [Accessed 10 September 2022].
- [21]H. Singh, P. Bhatt, S. Jacob, A. Kaur, A. Vashist and D. Vij, "Stock Prediction on Historical Data based on SGD and LSTM", *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 2022. Available: 10.1109/iciptm54933.2022.9753946 [Accessed 10 September 2022].
- [22]Y. Song, Y. Zhou and R. Han, "Neural networks for stock price prediction", *arXiv.org*, 2018. [Online]. Available: <http://arxiv.org/abs/1805.11317>. [Accessed: 10- Sep- 2022].

[23]M. Thormann, J. Farchmin, C. Weisser, R. Kruse, B. Säfken and A. Silbersdorff, "Stock Price Predictions with LSTM Neural Networks and Twitter Sentiment", *Statistics, Optimization & Information Computing*, vol. 9, no. 2, pp. 268-287, 2021. Available: 10.19139/soic-2310-5070-1202 [Accessed 10 September 2022].

[24]G. Vargas, L. Silvestre, L. Rigo Junior and H. Rocha, "B3 Stock Price Prediction Using LSTM Neural Networks and Sentiment Analysis", *IEEE Latin America Transactions*, vol. 20, no. 7, pp. 1067-1074, 2022. Available: 10.1109/tla.2021.9827469 [Accessed 10 September 2022].

[25]M. Yu, "Incorporation of Potential Sentiment Analysis Variable from Social Media in Stock Price Prediction", *ProQuest*, 2022. Available: <https://www.proquest.com/openview/dab64403ffc4b12a271208d8f38d3783/1?pq-origsite=gscholar&cbl=18750&diss=y>. [Accessed 10 September 2022].

