International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Stock Market Prediction- A Comparative Analysis

Hemkar Goswami, Aman Kumar

Students, Dept. of Computer Science and Engineering(SCOPE), VIT University, Tamil Nadu, India.

Abstract - *Stock market prediction is the act of trying to decide the future value of a company stock or other economic mechanism traded on an exchange. Stock market prediction is extremely difficult to conduct especially with high accuracy. People collect huge chunks of data trying to gain knowledge about the stock prices, their increment and decrement, which changes every second making it more and more difficult to foresee. They take into account the historical data, construct graphs and on the basis of this, they try to predict if it will be worth buying or selling the stock. This requires hours and hours of work and people have even considered this to be their full-time job (Market Analysts). Even people with a good understanding of statistics and probabilities cannot guarantee a 100 percent success rate. However, with all of that being said, if one can successfully predict the price of a stock, the profit can be huge. This is not easy to implement or even code for that matter. We need to take into account different risks since even a slight mishap and can make someone lose all their money. Our aim is to create several mathematical models that predict the stock market with high accuracy and then determine the best one amongst them. To obtain this high accuracy, we will be using Machine Learning and Python since recent studies show that if data is fed to the model in a correct manner, the output accuracy will be considerably high since the machines show continuous improvement over the course of time and can easily identify trends and patterns. All in all, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news' sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades.)*

Key Words: **Stock, RMSE, LSTM, Prophet, Linear Regression, Decision Tree**

1.INTRODUCTION

In predicting the stock market performance, the most difficult thing to do is calculation the highs and the lows. The factors involved in the prediction are– Physical Factors versus Psychological, Irrational and Rational behaviour, etc. All these aspects make the share prices volatile and thus make it very difficult to predict with a good degree of accuracy. Prediction of stocks was always a challenging task for Finance and Statistic experts. One of the main reasons behind this prediction is buying stocks that are likely to go up and then selling stocks that are probably to go down. Mostly, there are two ways for stock market prediction. Fundamental analysis is one of them and is based on a company's technique and fundamental information like market position, expenses and annual growth rates. The second one is the Technical analysis method, which concentrates on previous prices of stock and their values. This analysis uses historical charts and patterns to predict future prices of the stock.

However, the predictions could go wrong but this is the only way the invest in stocks. Stock markets are normally predicted by Financial and Statistical previously. But with time, Data Scientists have also started solving prediction problems with the use of learning techniques. Even Machine Learning methods are used to improve the performance of prediction models to get a higher the accuracy of predictions. Using deep learning is the next phase to improve prediction models with better and profitable performance.

Stock market predictions are full of challenges, and Data Scientists usually confront some problems when they try to develop a predictive model. Complexity and nonlinearity are the two main challenges caused due to instability of Stocks and the correlation between market behaviour and investment psychology. It is clear that they cannot be predicted to full accuracy and the factors such as the public image of companies or political situation of countries always affect stock markets trend. So, if the data gained from stock values are efficiently processed and suitable algorithms are used for prediction, the trend of stock values, prices, time and index can be predicted. In stock market prediction systems, machine learning and deep learning can help investors and traders in making their decisions. These methods are used to automatically recognize and learn patterns using a big data set. The algorithms can be effectively self-learning, and can handle the predicting task of price fluctuations in order to help in improving trading strategies.5 Stock market consists of unpredictable non-linear and dynamic nature.

Predicting stock prices is a very challenging problem as it depends on a number of factors which are not only limited to political conditions, global economy, company's financial reports and performance etc. but also to a number of other factors. Therefore, to increase the profits and decrease the losses, the techniques used to predict the stock values in advance by analysing the trend of the last few years, could prove to be of great use for making stock market movements and predictions. Generally, two main approaches are proposed for predicting the stock price of a company or an organization. Technical analysis method uses historical price of stocks like opening and closing price, volume traded, last traded, number of orders waiting to be processed, adjacent

close values etc. of the stock for predicting the stock future price. The next type of analysis is the Qualitative Analysis, which is based on the external factors such as company profile, market situation, political and economic factors, textual information in the form of financial new articles, social media and even blogs by known Economic Analyst. These days, advanced intelligent techniques based on either Technical or Fundamental analysis are used for predicting stock values and prices. Fundamentally, in stock market analysis, the data set is huge and also non-linear. Therefore, to deal with this much amount of data, efficient model is needed that can identify the hidden patterns and complex relations in this large data set.

The top ten technical indicators which uses ten years of historical data are the input values, and two ways are used for employing them. Initially, to calculate the indicators by stock trading values as continuous data, and then converting the indicators to binary data before using them. Every prediction model is evaluated by three measures based on the input ways. The results show that in the binary data evaluation, those deep learning methods are the better than all the others and the difference becomes less because of the improved of models' performance.

The projections of our model will make no attempt to deciding how much money to allocate to each of the stock predictions. More so, the project will also analyse the accuracies of these predictions in between each proposed model.

1.1 Scope

Times are really quite exciting an ever-increasing plethora of events followed the global financial crisis due to the recent event COVID pandemic. Globalization and innovation in the financial markets at its peak its very essential to study the market risks and requirements. Over the time, the stock market has undergone major changes. The stock market is extremely capricious in-deterministic system with a large number of factors influencing the direction of trend on varying scales and multiple layers. Thus, making a prediction for the upward or the downward trend becomes a very challenging task. This project primary scope is to compare multiple existing techniques as a test case for a company such as Tesla and come out with a much more robust prediction model among the existing models which will be capable to manage several scenarios in which investment can be beneficial.

Existing techniques can be too narrow in their approach and can lead to erroneous outcomes for varying scenarios. Stock projections predicted will guide the investor to minimize the risk and reap better returns. This analysis of stock will be beneficial for investors to invest in the stock market based on various factors. The stock values of a company like Tesla depend on many factors, some of them are the demand and supply of shares of a company which is a major reason for price change in stocks. As demand increase and supply decreases, price rises. and vice versa. Corporate results will also have an effect on the profits or progress of the company over a span of time say three months. The popularity of a company can affects buyers. Like if any good news of a company, may result in rising of the stock price. And bad news may break dreams. The stock value depends on many other factors as well.

We utilize data of a particular company its months high and low points of its stock in the prediction. We will analyses the variations in the stock value of the company through various machine learning models and find out which model predicts the stock prices most accurately in real-time. This project will focus exclusively on predicting the trend (price movement) of the individual stock of a company like Tesla,

1.2 Problem Statement

Stock market has become the main source of income for many people and the unpredictability of the stocks makes people lose huge amount of money. Numerous machine learning models have been devised, most of whose accuracy and consistency are poor. To solve this problem, we need to take into account various high accuracy models and run them on a common dataset and perform a comparative analysis to find out the model that gives the best accuracy amongst them and is therefore, reliable.

2. Methodology

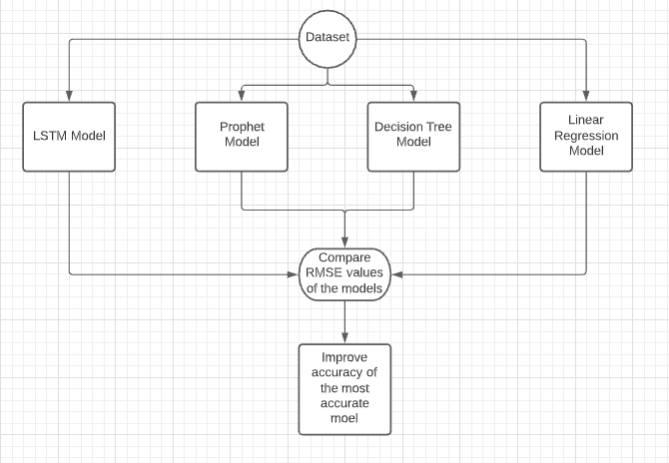

Figure 1- Architecture of Proposed Solution

Comparative analysis of various machine learning models utilized in stock market prediction especially for popular companies like Tesla etc. will be done. All the fissile essential metrics will be used to detect the accuracy of the models in providing suitable stock projections. Various models that are chosen are LSTM Model, Prophet, Decision Tree Model and Linear Regression Model. A common dataset from Kaggle (The Tesla Stock over the past 10 years) is chosen. The

models are run on this dataset and the output values and graphs are compared with the actual values, hence calculating the accuracy. Inference is drawn and noted from the same. The most basic machine learning algorithms like a decision tree and linear regression have been being implemented on the Tesla dataset which is very volatile in nature. The linear regression model provides an equation that determines the relationship between the independent variables and the dependent variable. A decision tree constructs regression or classification models in the shape of a tree structure. It splits down a dataset into smaller and smaller subsets while at the same time a correlated decision tree is incrementally developed. For our problem statement, we don't have any set of independent variables. We only have the dates instead. We will use the date column to extract features like – day, date, month, year and then fit linear regression and decision tree model. Next is sorting the dataset in ascending order and then create a separate dataset so that any new feature created does not affect the original data. Apart from this, we can add a more set of features that we believe would be relevant for the predictions. For instance, there's a hypothesis that the first and last days of the week could potentially affect the closing price of the stock far more than the other days. Finally, we will split the data into train and validation sets using in built python packages of TensorFlow to check the performance of the models. The final outcome is the stock prediction for a period of one month.

Now we have ruminated two more algorithms one is Prophet and the other is LSTM (Long Short-Term Memory). For both of these advance algorithm's predictions are done at a much larger scale computing stock prediction of 5 years considering the stock trends of the last decade. In LSTMs, the knowledge flows through a mechanism known as cell states. In which, LSTMs can particularly remember or forget things. The information of a particular cell state has three distinctive dependencies. We will envision this as follows; the stock price of any particular day will depend upon:

1. The trend that the stock has been charting in the previous days, maybe a downtrend or an uptrend.
2. The price of the stock on the preceding days, because many traders compare the stock's old prices before buying it.
3. The factors that can affect the price of the stock instantaneously. Like a new company policy that is being criticized widely, or a drop in the company's profit, or maybe an unexpected change in the senior leadership of the company.

A linear stack of layers is used in our LSTM implementation. The first layer is an LSTM layer with 80 memory units and it returns sequences. This is done to guarantee that the next LSTM layer receives sequences and not just arbitrarily scattered data. A dropout layer is applied in each subsequent LSTM layer in order to avoid overfitting the model. Finally, we have the last layer as a fully connected layer with an Adam optimizer predicting the next 264 values to an encoded result.

Understanding time-based patterns are critical for any stock market. This is why the time series forecasting of Prophet is employed for stock market prediction. The Prophet provides intuitive parameters which are easy to tune. Prophet () function is used to define a Prophet forecasting model in Python. To fit and forecast the considering effects of seasonality, the Prophet relies on the Fourier series to provide a flexible model. Due to the volatile nature of the stock market daily seasonality will be used in the Prophet. The dataset will be morphed to best use the time series forecasting ability of Prophet.

The GUI implementation is done through Tkinter. Tkinter provides a formidable object-oriented interface to the Tk GUI toolkit. We provide a browse option for uploading the dataset, then there is the radio button menu in order to choose the desired technique. If the correct technique is chosen for the corresponding dataset, the algorithm runs and the RMSE value is displayed. The lesser the RMSE value higher is the accuracy of the model prediction. Other than these proper exception handling parameters have been set to deal with unexpected outcomes. The whole GUI and the four algorithms are tied up together using object-oriented features of Tkinter on python. GUI takes the user input in form of dataset and technique selection and gives the output in the form of RMSE value which is a widely accepted metric for the accuracy of machine learning algorithms, so now users can decide which algorithm suits best for their needs according to their dataset

3. RESULTS AND DISCUSSION

3.1 GUI- Graphical User Interface

Our project aims to compare various different high accuracy stock prediction models and try to increase their accuracy even further. For this purpose, we have successfully developed a GUI which will give us the accuracy of the model based on the input dataset. For a particular input dataset, we shall obtain the RMSE (Root Mean Square Error) value which is inversely proportional to accuracy which suggests that higher the RMSE value, lower is the accuracy. A snippet of the GUI is shown below.

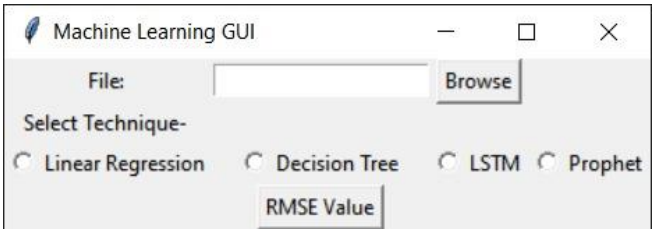

Figure 2- A snippet of the GUI

As seen from the figure, 4 different radio buttons have been given for 4 different stock market prediction models. An input dataset (such as TSLA.csv) can be selected and uploaded by the user and the RMSE value will be displayed for each model upon clicking the "RMSE Value" button. The user can instantly switch between the models, note down the values and easily compare the accuracy between them based on a dataset thus making it user friendly and less complex.

The GUI only displays the RMSE values and the backend is hidden from the user. However, to understand how each and every model works on a dataset, an in-depth analysis of each model is required. We have run each code in the Google Collab IDE and various graphs and outputs have been analyzed to develop a better understanding of the execution of the models. Lastly, an inference is drawn based on the RMSE values of various models and steps are taken to increase the accuracy of the existing models.

3.2 Stock Market Prediction Models

The 4 models used are:

A) Linear Regression Model

B) Decision Tree Model

C) LSTM Model

D) Prophet Model

[A] Linear Regression Model

This model uses two variables (outcome and predictor) to for its implementation. The linear equation obtained is used the predict the values. The graphical representation of the values after adding the dataset and setting the future days as 30 is shown below:

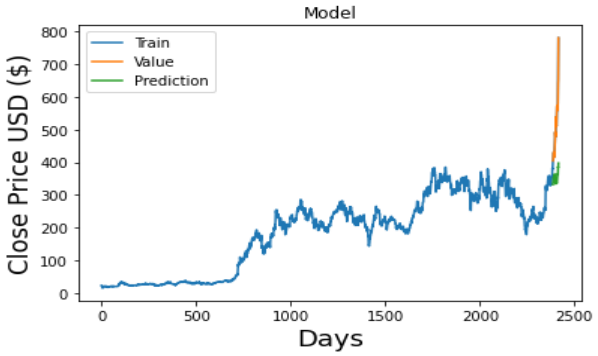

Figure 3- Graphical Representation for the Decision Tree Model

The training, predicted and the actual values is shown in the graph. It is evident that the predicted values have shot up and it is expecting that the TESLA stocks will increase by a tremendous amount and such was the case. It can be seen that the predicted values as well the actual values are overlapping indicating high accuracy of the model. However, to determine the actual accuracy, the RMSE value is calculated as follows. As observed from the implementation, the value of RMSE obtained is approximately 23. Considering the huge input dataset, this value depicts that error while executing this model on the input dataset is quite low. In other terms, the accuracy of this model is quite high and hence, the predicted and the actual values seem to overlap. However, only upon obtaining the RMSE values of the other datasets, one can determine which model has the least accuracy and which model produces the highest.

[B] Decision Tree Model

This model uses takes into account the various predictions in the form of tree and provides its predictive analysis based on the same. Upon uploading our dataset and number of days into the future as 30, the graph obtained is as follows:

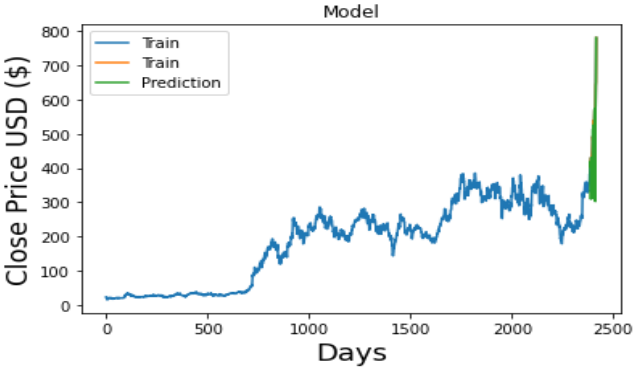

Figure 4 - Graphical representation for Decision Tree Model

In this case as well, the values for the stock are predicted to shoot up. However, the overlapping between predicted and the actual values is not as high as it was for the Linear Regression Model. It predicted that the value will rise up higher than it did indicating that the model is not as accurate. The RMSE value of this model obtained confirms this analysis. The RMSE value of obtained is approximately 150. It is hence evident that the error in predicting values in this model is significantly higher or in other terms, the accuracy is comparatively lower. However, when analyzing in absolute terms, the RMSE value obtained is not very high for this huge input dataset and that the model does generate accurate values when it is compared to rest of the machine learning models in the world. Thus, despite having a significantly higher RMSE value than the Linear Regression Model, this model can be used for predicting results, despite the comparatively lower accuracy.

[C] LSTM (Long Short-Term Memory)

LSTM is a widely used sequence prediction model and has proven to be extremely effective. LSTM is able to store the past information that is important and forgets the information that is not. LSTM has three gates, the input gate adds information to the cell state, the forget gate removes the information that is no longer required by the model and finally the output Gate at LSTM selects the information to be shown as output.

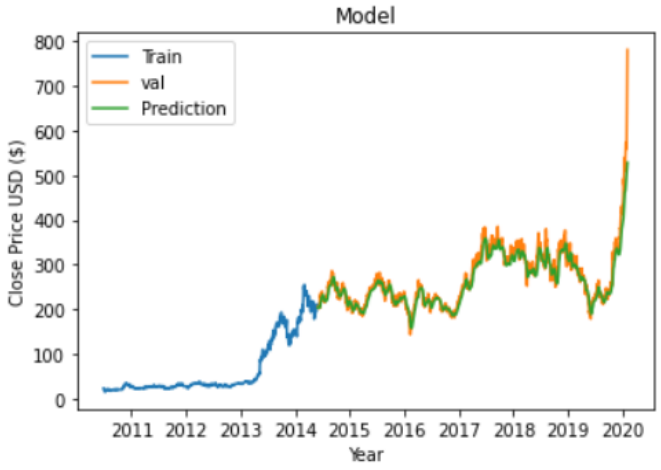

Figure 5- Graphical representation for LSTM

For this situation likewise, the qualities for the stock are predicted to boost up. In any case, the covering among predicted and the actual qualities is a lot higher in contrast with the Linear Regression Model and Decision Tree Model. It predicts that the value will rise up much higher indicating that the model is very accurate. The RMSE value of this model obtained confirms this analysis. The RMSE value of obtained is approximately 11. In different terms, the accuracy of this model is high and thus, the predicted and the actual qualities appear to cover by and large. The LSTM model can be tuned for various parameters such as changing the number of LSTM layers, adding dropout value or increasing the number of epochs.

[D] Prophet

Prophet, a stock prediction model that was engineered and developed by Facebook, is a time series forecasting library that requires no pre – processed data and is extremely easy to execute. The input for Prophet is a data frame with two columns: date and target (ds and y). Prophet tries to capture the seasonality in the past data and works well when the dataset is large. We use a decomposable time series model with three main model components: trend, seasonality, and holidays. Using time as a regressor, Prophet is trying to fit several linear and nonlinear functions of time as components. The following output is obtained:

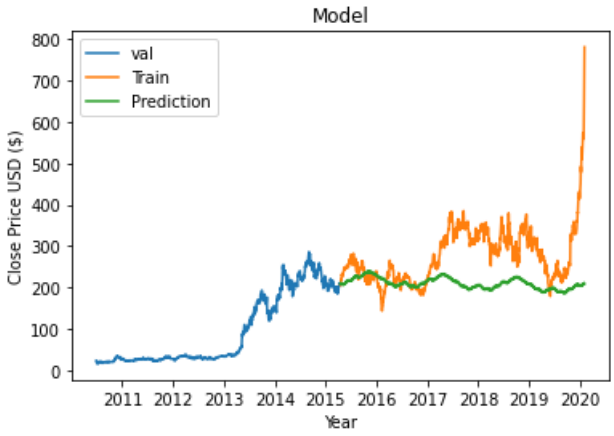

Figure 6- Graphical representation for Prophet

It is evident from the graph that the errors in this model are considerably higher than Linear regression Model and LSTM Model and around the same range as Decision Tree Model. This can be further confirmed by computing the RMSE values as the RMSE value obtained is approximately 96.4. Prophet tries to capture the trend and seasonality from past data. This model usually performs well on time series datasets. As it turns out, stock prices do not have a particular trend or seasonality. It highly depends on what is currently going on in the market and thus the prices rise and fall. Hence forecasting techniques like ARIMA, SARIMA and Prophet would not show good results for this particular problem.

4. CONCLUSIONS

In this project we tested a number of models to predict stock market and after going through various research papers, we came up with 4 models which give promising accuracy, namely LTSM, Linear Regression, Prophet and Decision Tree. The dataset used contains the details of Tesla which has been a very volatile stock for the past decade and predicting this with high accuracy is quite arduous. The Tesla stock doesn't replicate any conventional stock market trend and this is the main reason why we have worked on this dataset. Furthermore, we were successful in creating a GUI which helped us determine the accuracy metric RMSE with minimal complexity (Higher the RMSE value, lower is the accuracy of the model). Although all the models have a different accuracy rate which is better than rest of the models but out of the 4, the best model is LTSM with an exceptional accuracy rate while all the other 3 follows. LSTM is a widely used sequence prediction model and has proven to be extremely effective.

In future, we propose to implement 2 models together to produce better results to provide higher scope and functionality. If we are able to implement 2 models together, that would definitely increase the accuracy rate, hence aiding a better prediction. However, the model will become highly complex and depict higher time and space complexity. Supposedly, we implement two models LTSM and decision tree together, decision tree takes into account the various predictions in the form of tree and provides its predictive analysis based on the same while LTSM uses sequence prediction model and the two together can produce far better results than producing those results alone. Furthermore, Algorithmic Trading has a lot more scope which can implemented after predicting the market with good accuracy, here

REFERENCES

[1] M. Faraz, H. Khaloozadeh and M. Abbasi, "Stock Market Prediction-by-Prediction Based on Autoencoder Long Short-Term Memory Networks," 2020 28th Iranian Conference on Electrical Engineering (ICEE), 2020, pp. 1-5, doi: 10.1109/ICEE50131.2020.9261055.

[2] C. R. Madhuri, M. Chinta and V. V. N. V. P. Kumar, "Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA," 2020 7th International Conference on Smart Structures and Systems (ICSSS), 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9202042.

[3] S. Vazirani, A. Sharma and P. Sharma, "Analysis of various machine learning algorithm and hybrid model for stock market prediction using python," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 203-207, doi: 10.1109/ICSTCEE49637.2020.9276859.

[4] R. Sable, S. Goel and P. Chatterjee, "Empirical Study on Stock Market Prediction Using Machine Learning," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 2019, pp. 1-5, doi: 10.1109/ICAC347590.2019.9036786.

[5] N. Sirimevan, I. G. U. H. Mamalgaha, C. Jayasekara, Y. S. Mayuran and C. Jayawardena, "Stock Market Prediction Using Machine Learning Techniques," 2019 International Conference on Advancements in Computing (ICAC), 2019, pp. 192-197, doi: 10.1109/ICAC49085.2019.9103381.

[6] K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), 2018, pp. 151-156, doi: 10.1109/IACS.2018.8355458.

[7] M. Misra, A. P. Yadav and H. Kaur, "Stock Market Prediction using Machine Learning Algorithms: A Classification Study," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2475-2478, doi: 10.1109/ICRIEECE44171.2018.9009178.

BIOGRAPHIES

I want to utilize essential elements of my learning in developing, evolving, and implementing the best of software programming.

I want to use my skills as a software engineer and make the day-to-day tasks as simplified as possible.