

STOCK MARKET FORECAST WEBAPP

Aniket Wasnik*¹, Anshul Chaudharir*², Abhishek Khasre*³, Chetan Kaware*⁴,

Abhishek Kondalkar*⁵, Rajesh Nakhate*⁶

*^{1,2,3,4,5}UG Student, Department Of Information Technology, GH Rasoni College Of Engineering, Nagpur, India.

*⁶Professor, Department Of Information Technology, GH Rasoni College Of Engineering, Nagpur, India.

ABSTRACT

Stock trading is one of the most essential operations in the financial world. Stock market prediction is the attempt to estimate the future value of a stock or other financial instrument listed on a stock exchange. This article explores how Machine Learning can be used to forecast the price of a stock. The majority of stockbrokers utilise technical, fundamental, and time series analysis when making stock forecasts. Python is the programming language used for stock market forecasting. The goal of this paper is to use Facebook Prophet and yfin to examine the flaws in the current system and develop a time-series model that would implement more effective algorithms to largely mitigate them. Anyone can monitor the preferred stock they want to invest in using this model, and they can maximise profit by buying lots of it at the cheapest price and selling it at the highest.

Keywords: Stock, Machine Learning, Prediction, Yahoo Finance, Streamlit.

I. INTRODUCTION

The successful market hypothesis suggests that stock finances are influenced by knowledge in addition to intellectual speculation and that recently revealed information about the business environment is almost immediately taken into account in the current stock valuation. As a result, stock market movements allude to informational propaganda.

Future stock price forecasting involves a variety of variables. such as illogical, physical, and rational elements. All of these factors combine to make share prices exuberant and incredibly difficult to predict with a high degree of accuracy. New oriented algorithms like Facebook Prophet and Y-fin for predicting stock market valuation and tendency have been more well-known than ever before in the age of big data. In the past, traders used to look at the stock measure and retail valuation in addition to the recognition attached to those stocks in order to analyse the retail progression. Commercial retail is a sophisticated and intricate arrangement where the general public can exchange money, goods, and inherited rights according to practical norms supported by outsiders. The retail market enables traders to utilise the natural separation of social groups over trading, as well as via interchange or over contradictory information. The retail market is a name for important regions that traders are committed to, thus the development of the retail market's valuation is constantly a current concern for analysts in the commercial and professional spheres. This retailer has offered stockholders the chance to amass wealth and live a comfortable life by investing a small sum of money while being accompanied by a stunted sense of threat prior to the dread of the perforation of modern business. Due to the expansive and tenacious quality of the commercial stock exchange, systematic forecasting of stock market yields is an exceedingly difficult responsibility. Techniques related to forecast have been shown to be more effective in predicting stock valuation in light of the development of time series and improved ciphering tendencies.

II. LITERATURE REVIEW

During the literature review, we learned about the most recent techniques employed in the field of stock market forecasting. As a mix of multiple stock-selection techniques, [10][4] proposed a solution. They utilised the INDIAN Economic Journal database (TEJ). Their research used information from 2000 to 2019. As their prediction model, they employed a Multi Layer Perceptron (MLP)-based sliding window technique and integrated it with background multi-layer neural networks. In their study, they have employed principal component analysis (PCA), genetic algorithms (GA), and a graded and regressed tree (CART) to select crucial traits. They were depending on more than just technical indexes. In their investigation, fundamental and

macroeconomic indices were utilised. In addition, the authors provided a comparison of strategies for feature selection. The validation component was accomplished by combining model outputs with statistical analysis.

A nation's stock market is seen to be a reliable indicator of its level of economic strength. Prices on the stock market are always fluctuating because they are significantly influenced by the ebb and flow of money in various economic sectors. A basic tenet of economics is that stock prices are always influenced by the relationship between the demand and supply curves. A stock's price will rise if there is a greater demand for it, while a stock's price will fall if there is less demand for it. [11] Even while this variation in stock prices is required to profit from investments made, it is crucial that we estimate the future value of a stock in order to reduce investment losses. [13] By doing this assessment, we want to forecast stock prices and help you make better investment decisions.

A variety of machine learning classifiers and regression algorithms are used to analyse recently observed market patterns. Results from previous studies and various approaches[Fig. 1] are weighted based on several parametrics[Table 1] and then presented in a graphical way. The investigation reveals various established traditional stock market forecast methods that have been developed. Additionally, it examines how machine learning techniques have recently been applied as well as the advantages and disadvantages of each technique for future stock price prediction. [10] The stock price may be influenced by a variety of factors in the modern stock market and industry. Two important components are as follows:

1. The influence of stock prices on other companies, such as how the total stock prices of other companies affect the stock price of a certain company.
2. Outdated production and documentation pertaining to business stock price forecasts

Here, a variety of strategies are used to improve the historical pattern of stock market transaction and predict future outcomes in line with it. The stock market is noted for being active, unpredictable, and non-linear.

III. METHODOLOGY AND IMPLEMENTATION

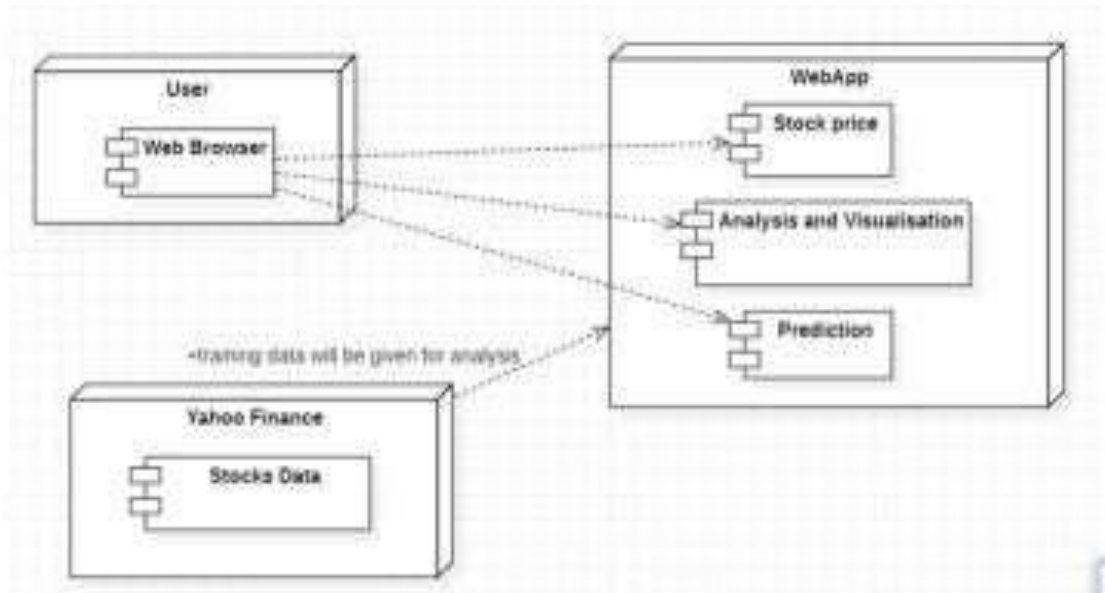


Figure 1: Methodology

In this project we are predicting stock price trends for a daily period, mainly for the development of models incorporating various data sources such as news articles, twitter data, Google data and Wikipedia data. The effect on stock price movements has been shown by all these external factors in combination with stock prices and technical stock indicators. The study made use of the yahoo finance module, which is basically a Python scraper that extracts finance data from the Yahoo Finance site. In Fig 1 Facebook prophet a web based app that predicts stock market prices, are used in this project to forecast stock market prices and building a Recurring Neural Network with LSTM cells as the latest state of the art in time series forecasting. The main aim of this research is to help stock market investors predict the behaviour and/or direction of stock market prices in order to participate in a long-term profit trend.

TABLE 1: Correlation output

	Open	Low	High	Close	Volume
Open	1.000	0.999	0.999	0.858	-0.027
Low	0.999	1.000	0.999	0.854	-0.037
High	0.999	0.999	1.000	0.859	-0.020
Close	0.858	0.859	0.859	1.000	-0.036
Volume	-0.027	-0.037	-0.020	-0.03	1.000

Some techniques and algorithms are used to produce more accurate final forecasts of data. RMSE, moving averages, and correlation plots can all be used to evaluate data more effectively. Techniques data must be tested to determine whether they are stationary before implementing the algorithm. A correlation plot is used to examine how various variables interact with the data. assists us in understanding how one variable is influenced by the other

The correlation values for each column are shown in the TABLE3 above, indicating how each column relates to the other column. The table demonstrates the strong correlation between almost all parameters. Correl's heat map

Heatmap for Correlation:

data shown graphically, with colours used to represent values. It is simple to imagine complex data as well. It illustrates the relationships between each variable and each individual variable. If the correlation value between two variables is high, this indicates that the two variables have a significant correlation with one another.

Moving Averages:

These are a straightforward and typical smoothing prototype used in time series analysis. It entails creating a new series whose values are formed from the sum of the initial time series' raw measurements. These are the fundamental ideas behind time series. These comprise the time series analysis baseline model. The time series is stationary and has a slow ranging mean, which is one of the fundamental assumptions of moving averages. Moving average may not be the best forecasting technique if data is highly erratic and has a non-stationary mean. Moving average can be employed when the data is stationary and has a constant mean. It can be applied to feature engineering, data preparation, and even prediction guidance. The moving averages listed below are of a certain type. Simple

Moving Average (SMA) : It estimates the midpoint of the last 'n' data points. It is a practical measure that can assist in influence if an asset price will continue or it will reverse a trend.

SMA: $(T+(T-1) +(T-2) +... (T-N))/N$ It determines the proportion of an elected assortment of prices, regularly closing prices, by the number of periods in that range.

Weighted Moving Average (WMA):It provides a weighted average of the most recent 'n' data points with weights. It is a specialised pointer that gives the majority of maximum contemporary data points a large weighting and gives data points in the distant past a smaller weighting.

WMA: $(T*WEIGHTING FACTOR) + ((T-1) * WEIGHTING FACTOR-1) + ... + ((T-N) * WEIGHTING FACTOR-N)/N$ prevailed by adding together newly discovered values and weighting each number in the data collection. EMA, or exponential moving average: Although we don't provide weights in this case, it is comparable to the weighted average. Instead of using t-1 and t-2 as the following inputs, it uses the prior period's time to calculate EMA. It is an expert chart pointer that tracks an expense's price over time..

EMA: $(CLOSE - PREVIOUS EMA) * (2/(SPAN+1)) + PREVIOUS EMA$

It reacts faster to data point changes. The duration we wish to compute in this case is Span. Explicit Smoothing It anticipates an additional parameter named "." It acts as a smoothing agent. The model is paying attention to the most recent data points if the value of is higher. Smaller here refers to counting the past.

$$Y_{t+1}: \alpha [Y_t + (1-\alpha) Y_{t-1} + (1-\alpha^2) Y_{t-2} + (1-\alpha^3) Y_{t-3} + \dots]$$

It is the rule of precept technique for smoothing time series data utilizing the exponential window function.

RMSE: If RMSE is less then, the Error is less. Here, RMSE is adopted to estimate which moving average is best for data. Here, we examine the 'Close' column and applying RMSE for that.

$$RMSE: \sqrt{\sum (Y_i - \hat{y}_i)^2 / 2}$$

TABLE 2: RMSE results for Moving Averages

Moving Average	21.519
Weighted moving average	19.750
Exponentially weighted average	20.075
Exponential smoothing	17.664

FACEBOOK PROPHET:

It is an algorithm developed by Facebook's Core Data Science team that is utilized in various applications of time series forecasting. It is very much utilized when there is a chance of seasonal effects. Prophet was firstly developed to generate high-quality marketing forecasts. This tries to find points like:

1. Changes in trend due to various products.
2. Outliers
3. Seasonal effects like weekly, monthly, yearly cycles.

$y(t): g(t) + s(t) + h(t) + \epsilon^*t$ - Additive regression model $g(t)$ - trend

$s(t)$ - seasonality ϵ^*t - Error There are numerous techniques for time series forecasting.

MAPE can be utilised to make exact predictions. This gives us the know-how to anticipate time series accurately using straightforward intrinsic parameters and includes a provision for incorporating the impression of inherited seasonality and holidays. Facebook Prophet aims to incorporate both linear and nonlinear time functions as elements. Exponential smoothing uses the same method of modelling seasonality as an additive component. This library is so important that it has the capacity to handle aspects linked to seasonality as well as stationary within the data. However, Facebook Prophet has some restrictions, such as the assumption that there are input columns with the names "ds" and "y," where "ds" stands for Date and "y" for the target variable. In this case, a trend may be increasing or declining and may be positive or negative. The data sample looks like this:

TABLE 3: Facebook Prophet sample data

	ds	Y
0	2012-01-03	663.590
1	2012-01-04	666.450
2	2012-01-05	657.210

In this case, the target column is Y, and the Date column is ds. Facebook Prophet needs this to be the case for the ds and y input commands. Other than these two, it accepts no other variable names. After that, the model is ready to fit for train data. After data is fit for training future predictions are done on data.

TABLE 4: Future data frame

ds	Y hat	Y hat_lower	Y hat_upper
2020-07-26	1338.352	1248.036	1430.649
2020-07-27	1404.407	1306.644	1494.335
2020-07-28	1406.271	1305.401	1494.738

This is sample data for future predictions.

Y hat- Predicted value

Y hat_upper – upper confidence interval

Y hat_lower – lower confidence interval

Figure3 – Forecasting of future predictions From Figure3, y hat is obtained based on y from previous years. Here, y is the actual prediction from data, and y hat is a future prediction based on previous data. Almost our model recognized good performance in predicting future forecasting of stock prices. We can see it better by using a forecast plot using confidence interval also. There is a parameter called interval_width if that is set to 0.95 that indicates setting confidence level to 95%. There, X-axis is the Date and Y-axis is the target variable that is the Close column variable. Here, Black points indicate original data, Blue indicates Predicted values and Light blue indicates Confidence interval. Our model has done a good performance in the prediction of future stocks.

IV. RESULTS AND DISCUSSION

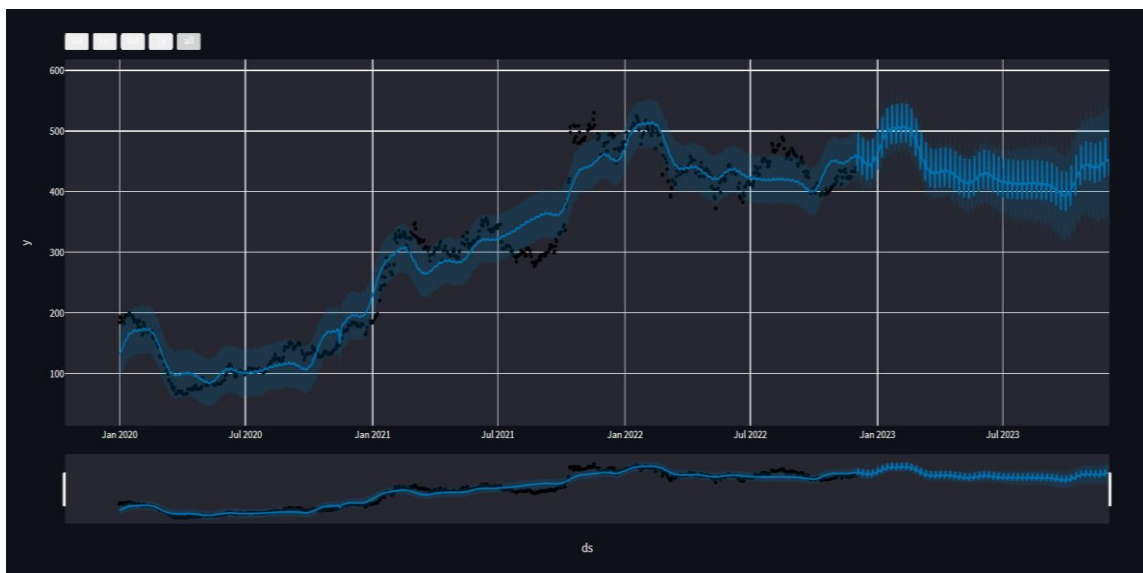


Figure 2: Prediction Graph

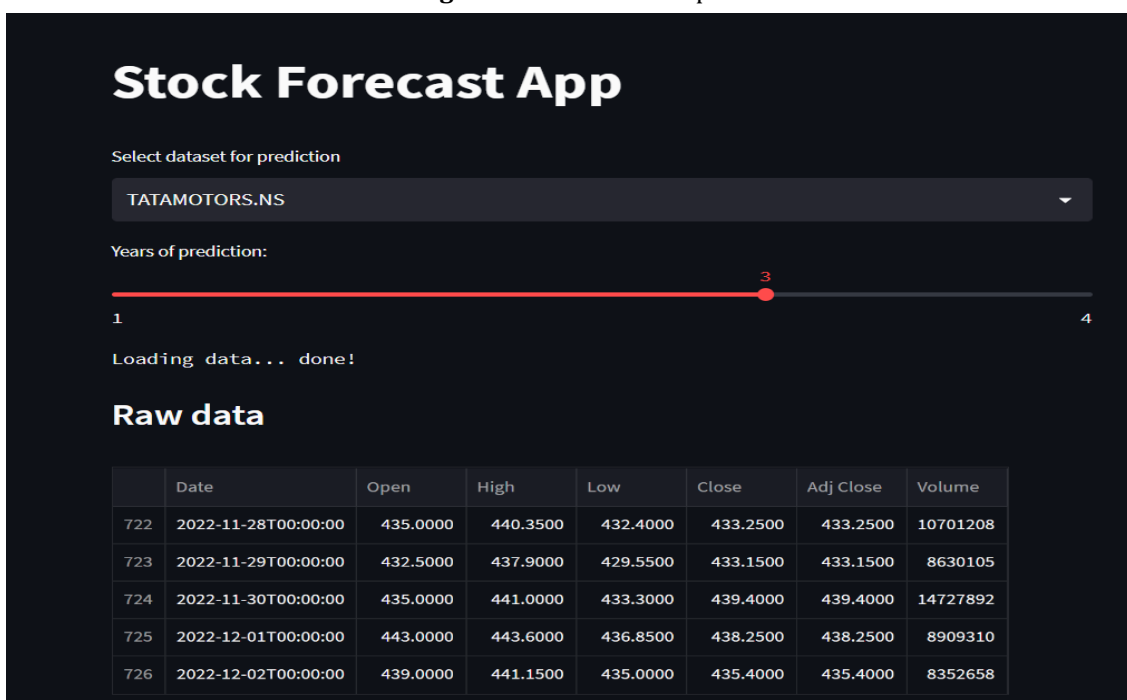


Figure 3: Inputs and data set loading with number of Prediction Year

The system was trained using complicated and raw data obtained from Yahoo Finance. Create an module with a dense output layer . After that, we pulled and trained the model on the improved dataset, which led to findings that were plausible. During the training phase, model output was also visualised using Tensor Board. Following the execution of the code, the model was trained for multipal epochs. Fig. 3: is a graphical representation of model prediction with the blue line indicating prediction curve in Fig. 3. Whilst Fig. 3 shows an almost flat but gradual growth of stock from the years. As such our model at the point of testing and training using Yahoo_Fin, using available metrics in python pandas programming language. we are predicting values for four years .

V. CONCLUSION

It has been observed recently that the majority of people invest in the stock market in order to generate quick cash. An investor also runs a good chance of losing all of their money. As a result, a strong predictive model is required if the consumer is to comprehend future market trends. There are many models that can predict whether the economy is improving or declining, but their predictions are not always accurate. With the aid of Facebook Prophet, an attempt has been made to create a stock market predictive model that accurately predicts the trend for the following day and displays prices that are close to reality.

VI. REFERENCES

- [1] Atsalakis GS, Valavanis KP. (2018). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Syst Appl.* 2009;36(7):10696– 707. *Computer Systems*, vol. 79, pp. 960–972.
- [2] Fischer T, Krauss C. (2018). Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res.* 270(2):654–69.
- [3] Gareja Pradip, Chitrak Bari, J. Shiva Nandhini (2018). Stock market prediction using machine learning. *International Journal of Advanced Research and Development.*
- [4] Khedr, Ayman E., and Nagwa Yaseen. (2017). Predicting stock market behaviour using data mining technique and news sentiment analysis." *International Journal of Intelligent Systems and Applications* 9.7.
- [5] Lei L. (2017). Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. *Appl Soft Comput J.* 2018;62:923–32.
- [6] Lin Z. (2018). Modelling and forecasting the stock market volatility of SSE composite index using garch models, *Future Generation*
- [7] Maini, Sahaj Singh, and K. Govinda. (2017). Stock market prediction using data mining techniques." *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE.
- [8] Nekoeiqachkanloo H, Ghojogh B, Pasand AS, Crowley M. (2019). Artificial counselor system for stock investment. *ArXiv Preprint arXiv :1903.00955.*
- [9] Sun J., Xiao K., Liu C., Zhou W., and Xiong H. (2019). Exploiting intra-day patterns for market shock prediction: a machine learning approach," *Expert Systems With Applications*, vol. 127, pp. 272– 281.
- [10] Tsai CF, Hsiao YC. (2019). Combining multiple feature selection methods for stock prediction: union, intersection, and multiintersection approaches. *Decis Support Syst.*
- [11] Weng B, Lu L, Wang X, Megahed FM, Martinez W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Syst Appl.*
- [12] McNally S, Roche J, Caton S. (2020). Predicting the price of bitcoin using machine learning. In: *Proceedings—26th Euromicro international conference on parallel, distributed, and network based processing*, PDPpp.339–43.