Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)

# Statistical Model Comparison for the National Stock Exchange (NSE) Stock Price Prediction

**[1]Prachi Gupta., [2] Hemanth Kumar Molapata.,[3]G. Madhu Sudan., [4]Nagendra Kumar Kalaparthi., [5]Darvinder Kumar., [6]Sreesaketh Karri**

[1]Department of Statistics, University of Allahabad, Prayagraj, U.P.-211002 India.
[2]Asst Professor, Department of Statistics, Hindu College, UoD, New Delhi India.
[3]Asst Professor, Department of Statistics, University of Allahabad, Prayagraj-211002 India.
[4]Asst Professor, Department of Statistics, S.V.College, UoD, New Delhi India.
[5]Associate Professor, Department of StatisticsPGDAV College, UoD, New Delhi India.
[6]Asst Professor, Department of Statistics, Delhi Public School, Bangalore East India.
Corresponding author gmadhusudan7@gmail.com

***Article History:***

***Received:*** *27-07-2024*

***Revised:*** *12-09-2024*

***Accepted:*** *01-10-2024*

**Abstract:**

The stock market is one of the most versatile sectors in the financial system, and the stock market plays an important role in economic development. It is also a platform for trading various securities and derivatives without barriers. This paper focuses on the comparison of different types of time series models to predict the stock prices of five companies by using machine learning techniques.

**Keywords**: Stock Market, Stock Price Prediction, Time series Models, NSE.

## 1. Introduction

**STOCK:** A stock (Akhilesh Ganti, Mar 2019) also known as share or equity, is a kind of security that implies proportionate in the issuing organization. This qualifies the investor for that extent of the company's advantages and profit. Stocks are purchased and sold dominatingly on stock trades, however there can be private deals also and the establishment of about each portfolio.

### National Stock Change (NSE)

The National Stock Exchange is the leading stock exchange in our country. It was established in 1992 and is headquartered in Mumbai. The NSE offers trading in equities, derivatives, and debt securities. The NSE is known for its modern electronic performance, which allows for fast and most efficient trading. It also provides a wide range of trading products and services, including equity trading, index trading, currency trading, and commodities trading. Overall, the NSE plays an important role in the Indian economy and is a key player in the global financial markets.

**Nifty-50:** The Nifty is the flagship benchmark of the National Stock Exchange (NSE), which is a well-diversified index comprising the top 50 companies in terms of free-float market capitalization that are traded on the bourse. It is supposed to reflect the health of the listed universe of Indian companies, and hence the broader economy, in all market conditions. Officially called the Nifty50, the index is computed using the free float market capitalization method, which is essentially the count of shares in active circulation in the market at any given point in time. The Nifty, just like the BSE benchmark Sensex, is today used for benchmarking portfolios and returns of mutual fund schemes and launching index funds.

### Objective

1. To compare different types of model and algorithms on the National Stock Exchange (NSE) data for five companies.

2. Comparison of neural network and deep learning algorithms with traditional statistical approaches for stock price forecasting by analysing National Stock Exchange (NSE) listed 5 companies stock price.

**Variable Description:**

**Open:** Represents the opening price of the stock at a particular date It is the price at which a stock started trading when the opening bell rang.

**Close:** Represents the closing price of the stock at a particular date. It is the last buy-sell order executed between two traders. The closing price is raw price which is just the cash value of the last transacted price before the market closes.

**High:** The high is the highest price at which a stock is traded during a period. Here the period is a day.

**Low:** The high is the lowest price at which a stock is traded during a period. Here the period is a day.

**Adjusted Close:** Adjusted closing price factors in corporate actions, such as stock splits, dividends and rights offerings.

**Volume:** Volume is the number of shares of security traded during a given period of time. Here security is a stock.

**Volume Weighted Average Price (VWAP):** The historic VWAP is the target variable to predict. VWAP is a trading benchmark is used by traders that give the average price the stock has traded at throughout the day, based on both volume and price.

## 2. Methodology

Autoregressive Integrated Moving Average Model **(ARIMA):**

ARIMA (*p*, *d*, and *q*), where *p* is the autoregressive term, *q* is the number of moving average terms, and *d* is the number of differences made when the time series becomes stationary. The prediction results can be adjusted by adjusting the aforementioned three parameters *d*, *p*, and *q*, so as to draw the optimal model. The Autoregressive Integrated Moving Average Model is as follows:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \ldots\ldots + \phi_p y_{t-p} \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots + \theta_q \varepsilon_{t-q} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where $\phi_i$ and $\varepsilon_j$ are the actual value and random error of the time period *t*, respectively; $(i = 1, 2\ldots, p)$ and $(j = 1, 2,\ldots, q)$ are 0th model parameters; *p* and *q*, the order of the model (*p* and *q* are integers), are also the model parameter mentioned earlier; the random error , whose mean value is 0, is assumed to be independent and obey the same distribution in the model. The variance of constant term is denoted as $\sigma^2$ . Equation (1) involves several important special cases of ARIMA series models. If $q = 0$, then equation (1) can be simplified to an AR model of order *p*. When $p = 0$, the model can be simplified to a *q*-order MA model. Among them, the model order (*p*, *q*) is the key link in ARIMA model construction, which determines the accuracy of model prediction. The parameters of the AR and MA operations are defined as (*p*) and (*q*), respectively. These two parameters need to be determined by the auto-correlation graph (ACF).

**ARIMAX:** ARIMAX or Regression ARIMA is an extension of ARIMA model. The ARIMAX model represents a composition of the output time series into the following parts: the autoregressive (AR) part, moving average (MA) part, integrated (I) component, and the part that belongs to the exogenous inputs. The exogenous part reflects the additional incorporation of the present values and past values of exogenous inputs (dynamic factors in our case) into the ARIMAX model.

**AKAIKE INFORMATION CRITERION (AIC):** AIC is a single number score that can be used to determine which of multiple models is most likely to be the best model for a given data set. It estimates models relatively, meaning that AIC scores are only useful in comparison with other AIC scores for the same data set. A lower AIC score is better. Mathematically it is defined by; AIC = 2k – 2ln(L)

Where k = no. of estimated parameters in the model, L = maximized value of the likelihood function for the model.

**BAYESIAN INFORMATION CRITERION (BIC):** The BIC is an increasing function of the error variance and an increasing function of k, that is unexplained variation in the dependent variable and the no. of explanatory variables increase the value of BIC. However, a lower BIC does not necessarily indicate one model is better than another. Mathematically it is defined by, BIC = k ln(n) – 2ln(L)

Where n = no. of observations K = no. of estimated parameters in the model

**ROOT MEAN SQUARE ERROR (RMSE):** RMSE is a frequently applied measure of the differences between numbers (population values and samples) which is predicted by an estimator or a mode. The RMSE aggregates the magnitudes of the errors in predicting different times into a single measure of predictive power.

**MEAN ABSOLUTE ERROR (MAE):** Mean absolute error is a measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a regression model.

The formula is given by MAE = $\frac{1}{n}\sum_{i=1}^{n}|y_i - y_i\hat{}\ |$

$y_i$ = true value; $y_i$^ = predicted values; n = number of observations

**PROPHET Forecasting:** Prophet (Taylor SJ, Letham B. 2017) is a methodology for forecasting time series data dependent on an added substance model where non-direct patterns are fit with yearly, week by week, and daily trends, in addition to occasion impacts. It works best with time series that have solid occasional impacts and a few periods of authentic data. Prophet is vigorous to missing data and moves in the pattern, and commonly handles anomalies well.

**Capacities:** Experts can use any outer information from any source for the absolute market estimate and then resourceful information as learning can be applied legitimately by determining limits.

**Change points:** Known dates of change points, such as dates of product changes, can easily be specified.

**Holidays and seasonality**: Experts' experience involvement help with which occasions sway development in which districts and they can legitimately use as input to store relevant occasion dates and the relevant time sizes of regularity or seasonality.

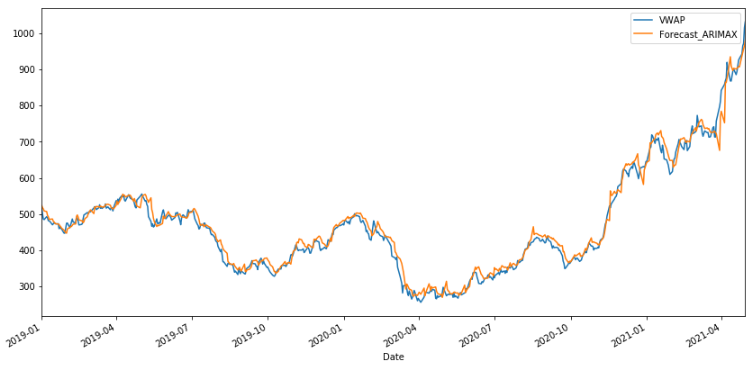
**Smoothing parameters**: Hereby changing the value of τ, this can be chosen from inside scope of increasingly worldwide or locally smooth models. The seasonality and holiday smoothing parameters (σ, ν) help model to amount of the historical seasonal variety which will be expected in the future.

**LIGHT GBM (Light Gradient Boosting Machine):** Light GBM is a gradient boosting framework that uses tree-based learning algorithm. Light GBM grows tree vertically while other algorithm grows trees horizontally meaning that Light GBM grows tree leaf wise while other algorithm grows level wise. It will choose the leaf with max delta loss to grow. When growing the same leaf, leaf wise algorithm can reduce more loss than a level wise algorithm. As with other decision tree based methods, Light GBM can be used for both classification and regression. Light GBM is optimized for high performance with distributed systems.

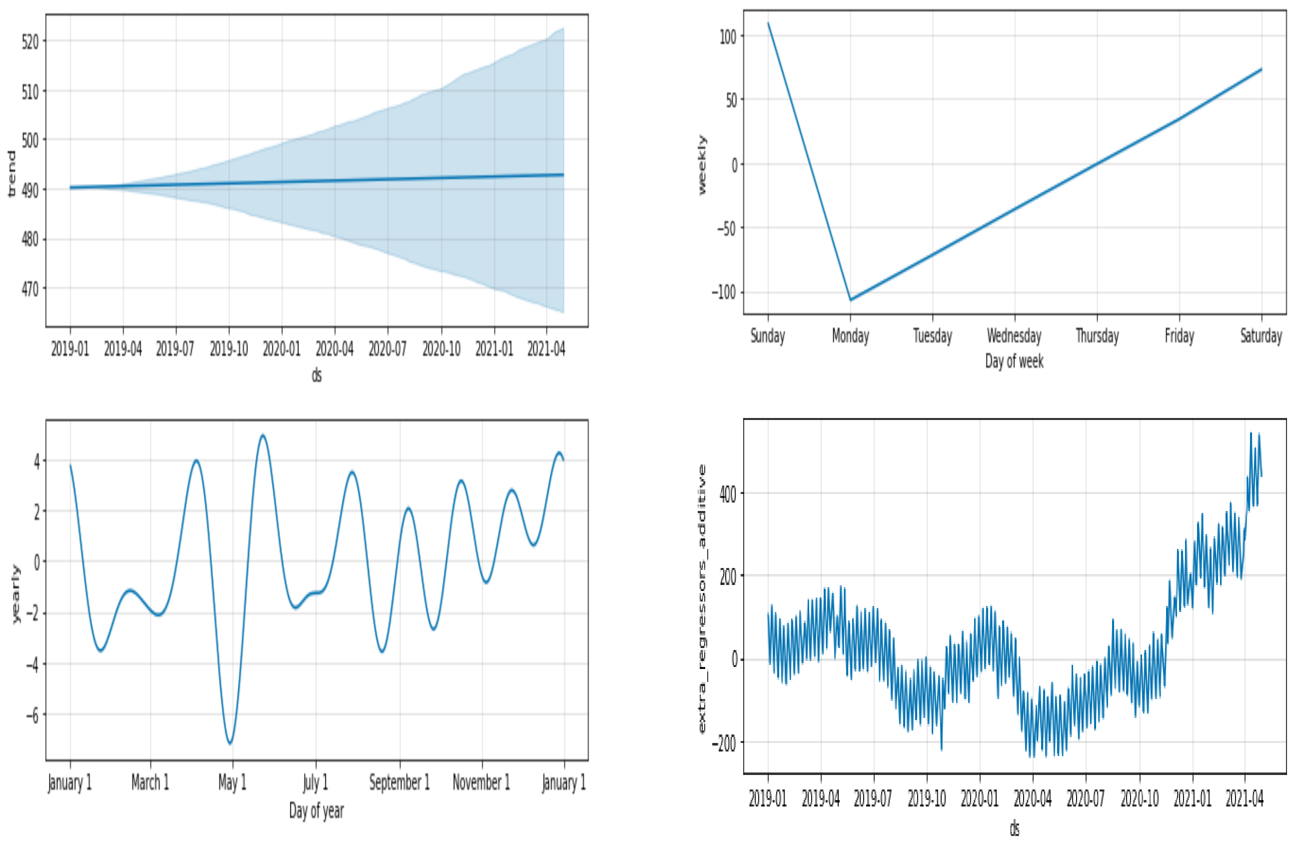
The size of data is increasing day by day and it is becoming difficult for traditional data science algorithms to give faster results. Light GBM is prefixed as ‘Light’ because of its **h**igh **speed.** Light GBM can handle the large size of data and takes lower memory to run. Another reason of why Light GBM is popular is because it focuses on accuracy of results. LGBM also supports GPU learning and thus data scientists are widely using LGBM for data science application development.

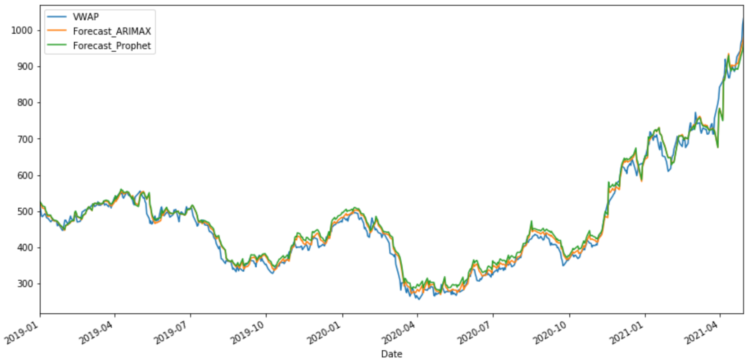
## 3. Empirical Investigations

### Adaniports

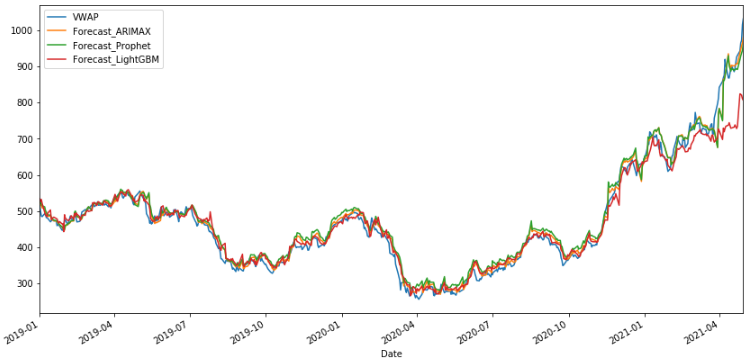



| MODEL NAME | MAE | RMSE |
|---|---|---|
| ***ARIMAX*** | ***11.4641*** | ***22.7331*** |



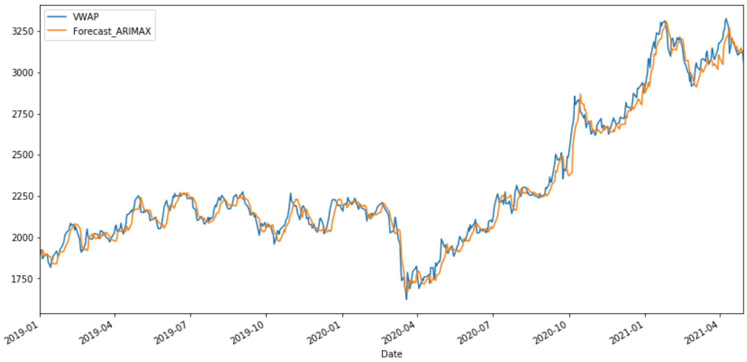



| *MODEL NAME* | *MAE* | *RMSE* |
|---|---|---|
| ***ARIMAX*** | ***11.4641*** | ***22.7331*** |
| ***PROPHET*** | ***17.8051*** | ***40.6612*** |

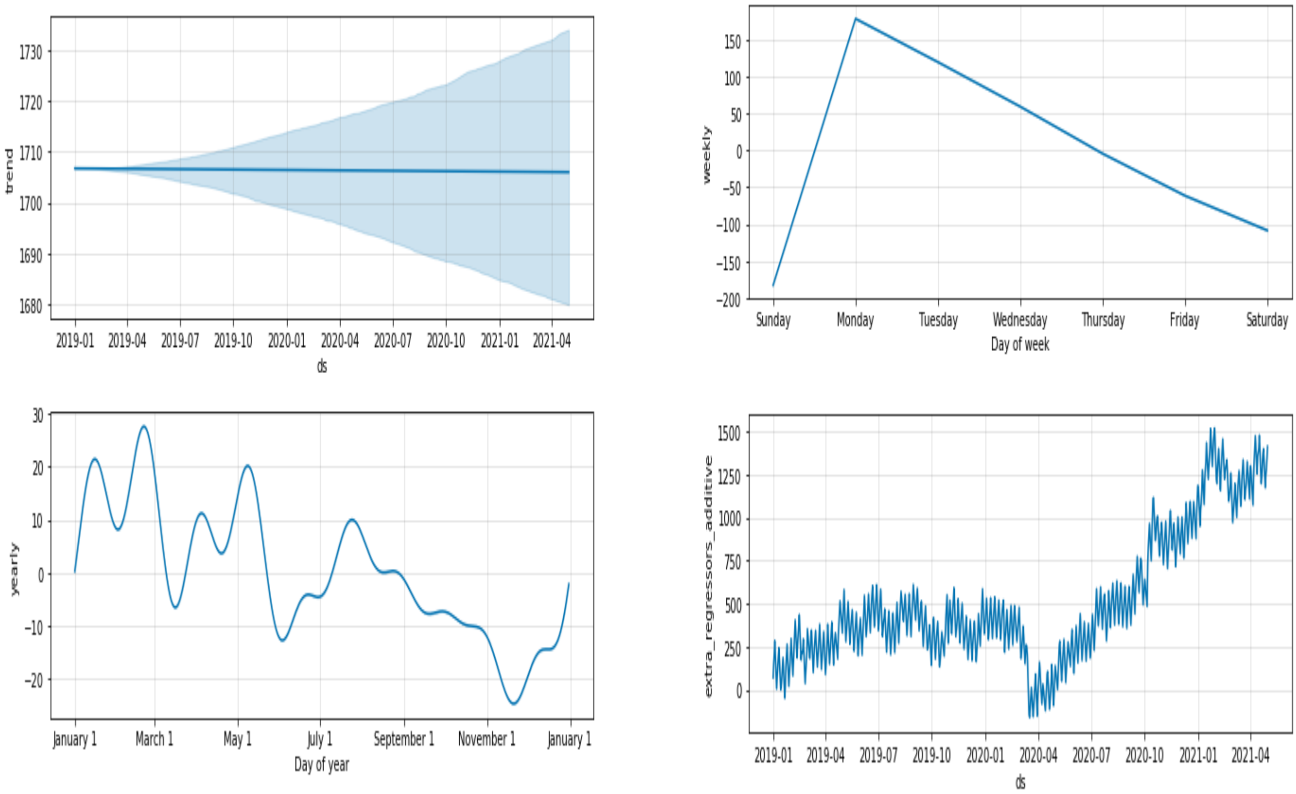


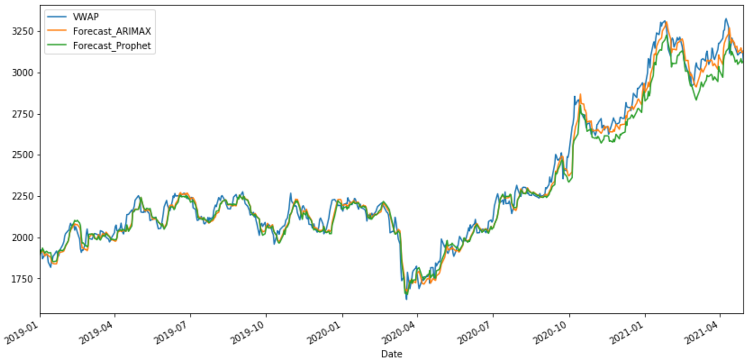

| *MODEL NAME* | *MAE* | *RMSE* |
|---|---|---|
| ***ARIMAX*** | ***11.4641*** | ***22.7331*** |
| ***PROPHET*** | ***17.8051*** | ***40.6612*** |
| ***Light GBM*** | ***12.6912*** | ***20.1033*** |

***TCS :***

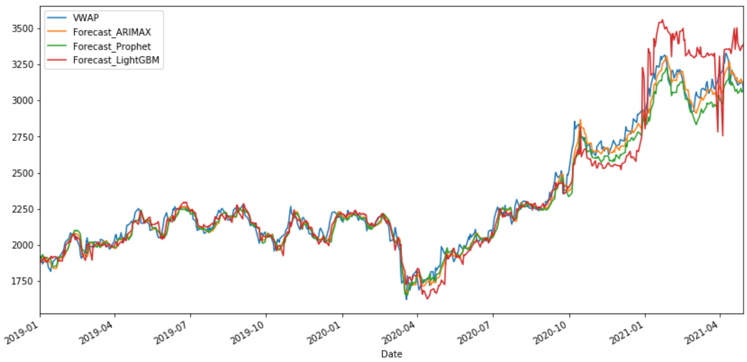



| *MODEL NAME* | *MAE* | *RMSE* |
|---|---|---|
| ***ARIMAX*** | ***43.6138*** | ***57.4385*** |



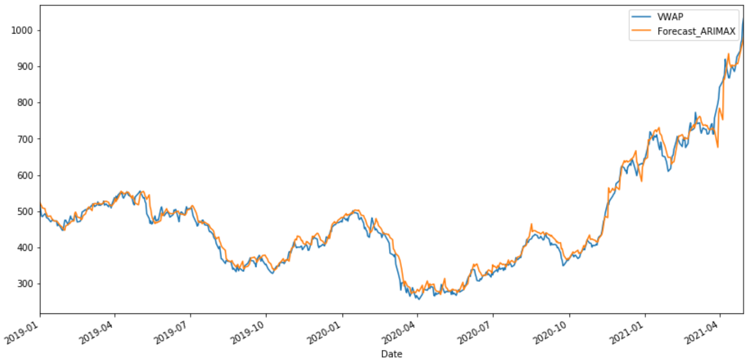

| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 43.6138 | 57.4385 |
| PROPHET | 54.6370 | 73.9456 |



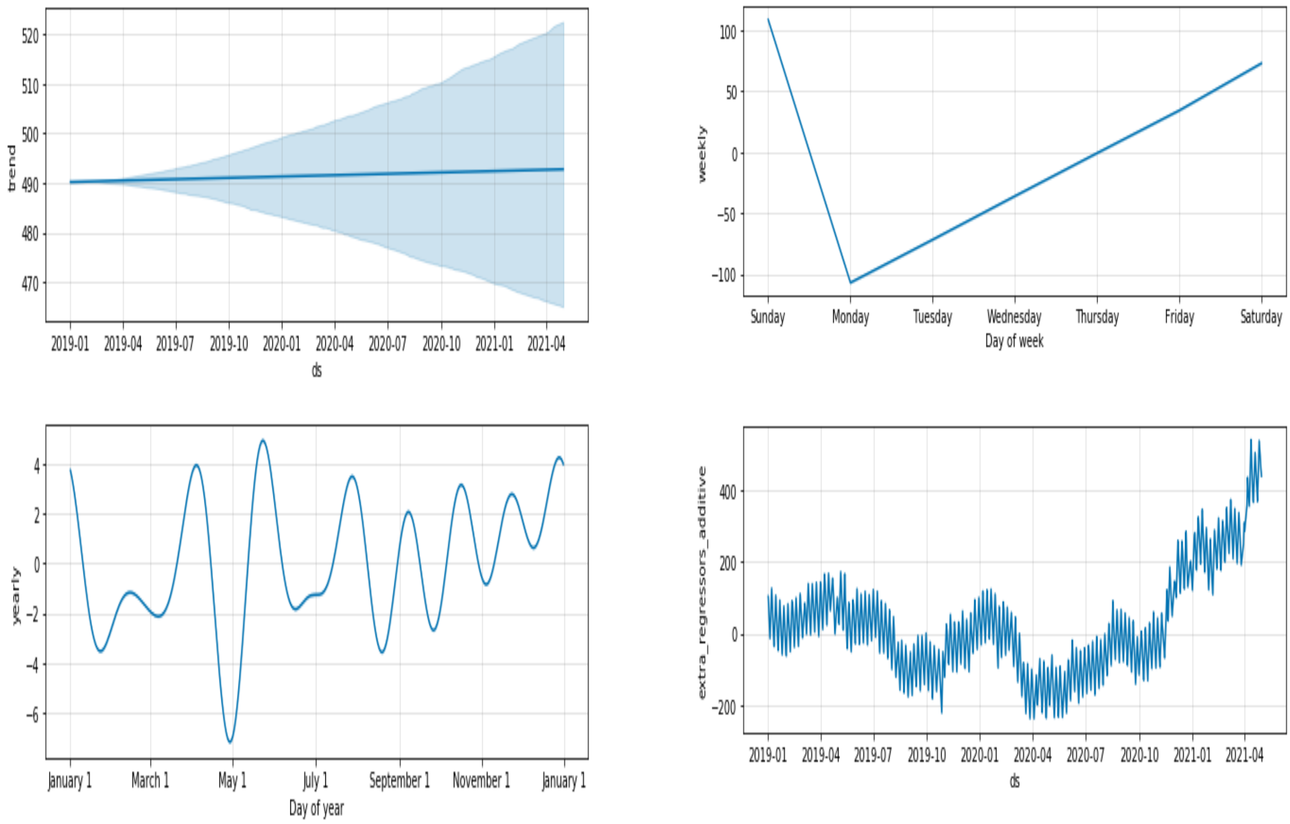

| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 43.6138 | 57.4385 |

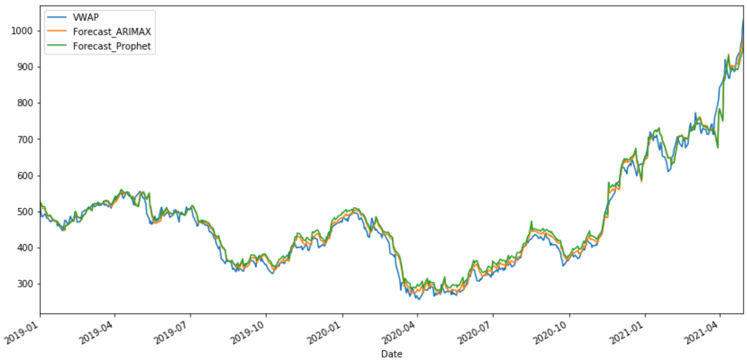
| | | |
|---|---|---|
| *PROPHET* | *54.6370* | *73.9456* |
| *Light GBM* | *81.3152* | *121.2054* |

**Tata Steel**

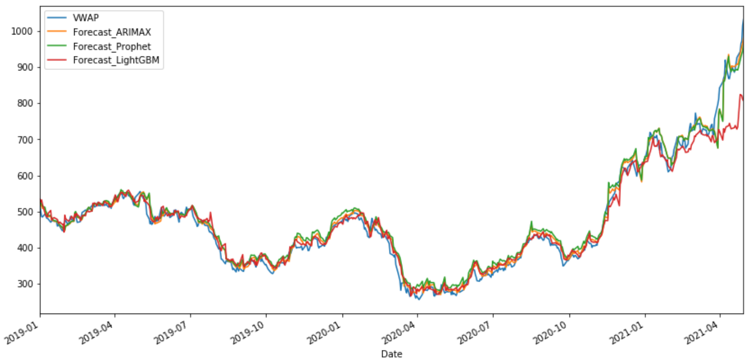



| *MODEL NAME* | *MAE* | *RMSE* |
|---|---|---|
| *AUTO ARIMAX* | *14.1287* | *19.2372* |



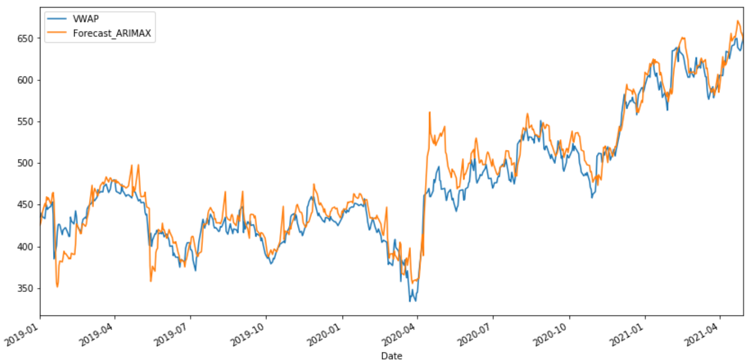



| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 14.1287 | 19.2372 |
| PROPHET | 18.4501 | 23.3080 |



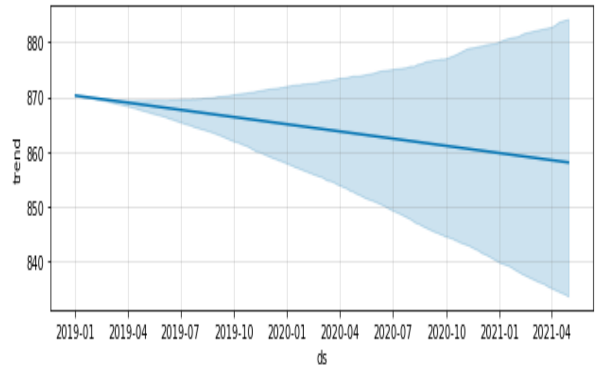

| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 14.1287 | 19.2372 |
| PROPHET | 18.4501 | 23.3080 |
| Light GBM | 18.8233 | 35.1964 |

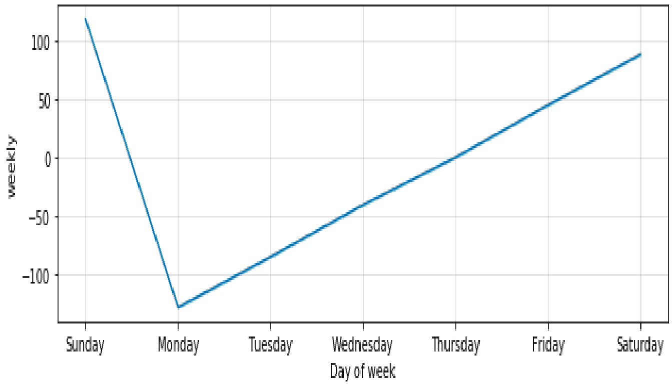
**Sunpharma** :

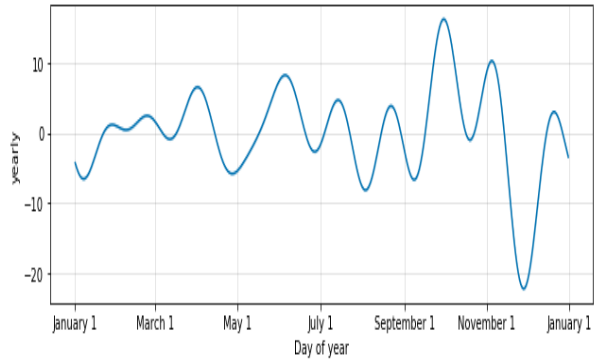


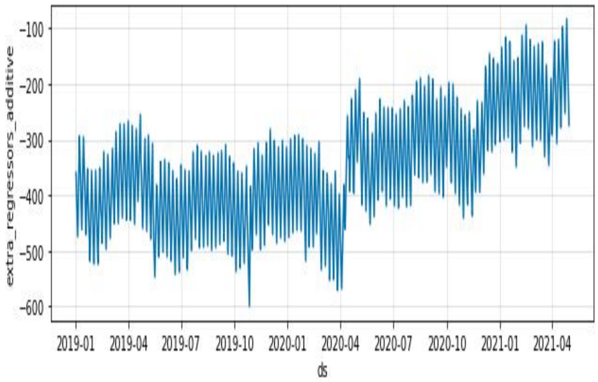

| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 15.3512 | 20.4611 |

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)



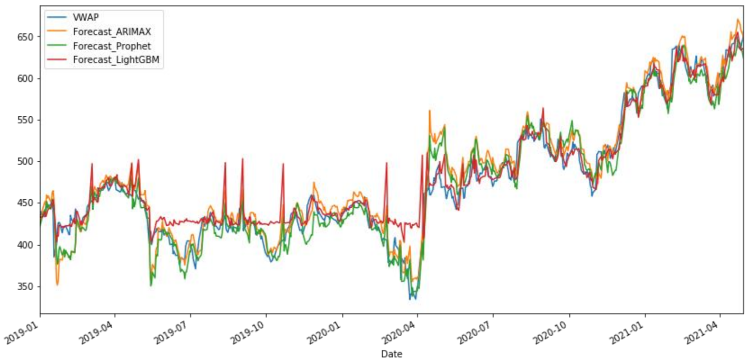



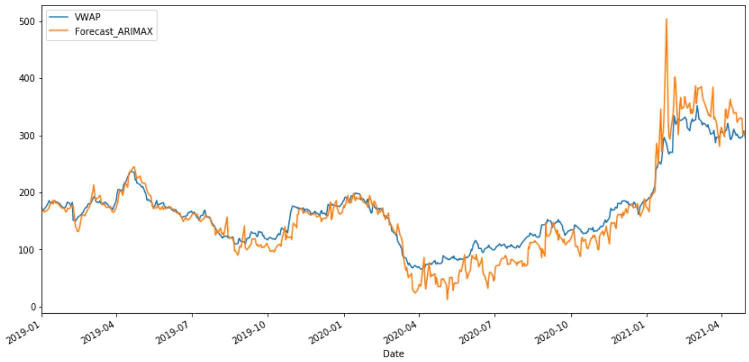


| ***MODEL NAME*** | ***MAE*** | ***RMSE*** |
|---|---|---|
| ***ARIMAX*** | ***15.3512*** | ***20.4611*** |
| ***PROPHET*** | ***14.3392*** | ***19.00905*** |

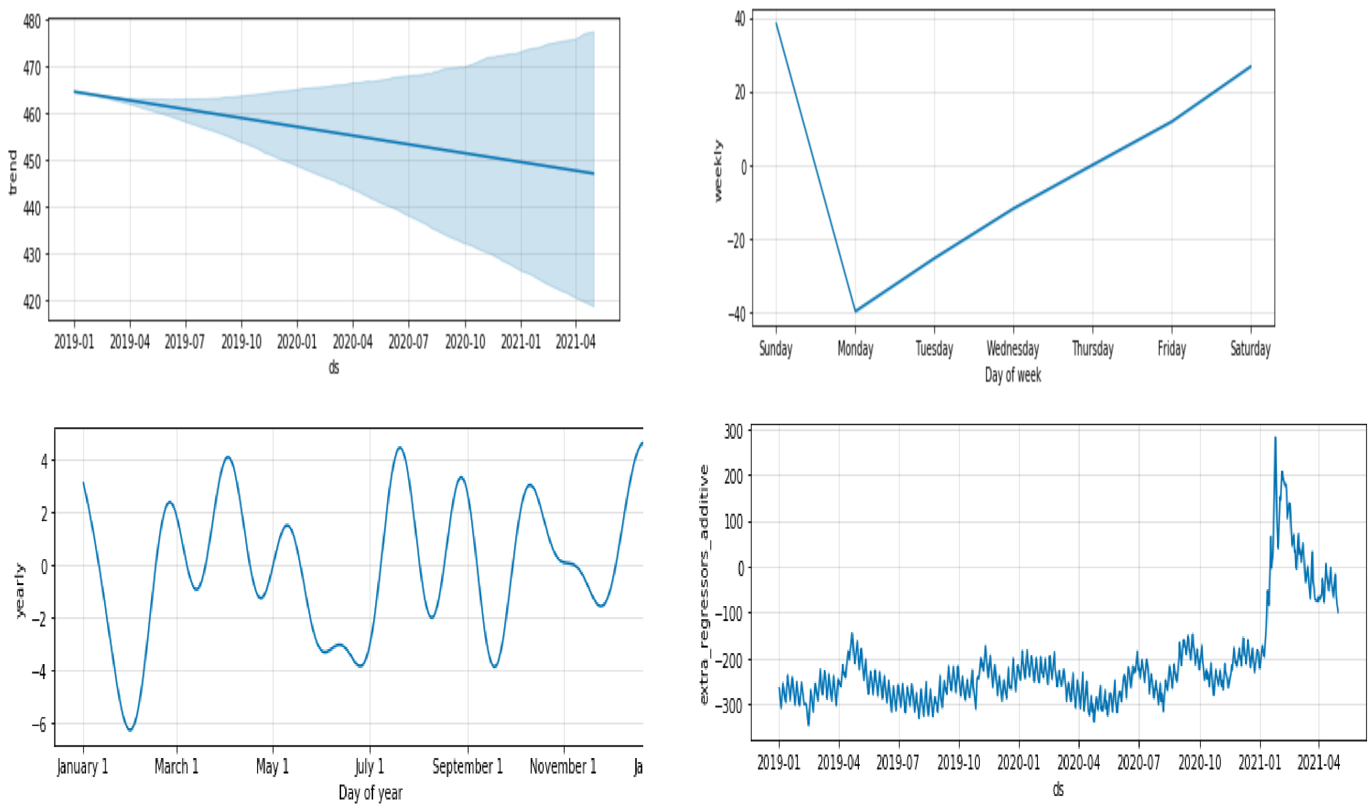


| ***MODEL NAME*** | ***MAE*** | ***RMSE*** |
|---|---|---|
| ***ARIMAX*** | ***15.3512*** | ***20.4611*** |
| ***PROPHET*** | ***14.3392*** | ***19.00905*** |
| ***Light GBM*** | ***13.30007*** | ***20.9007*** |

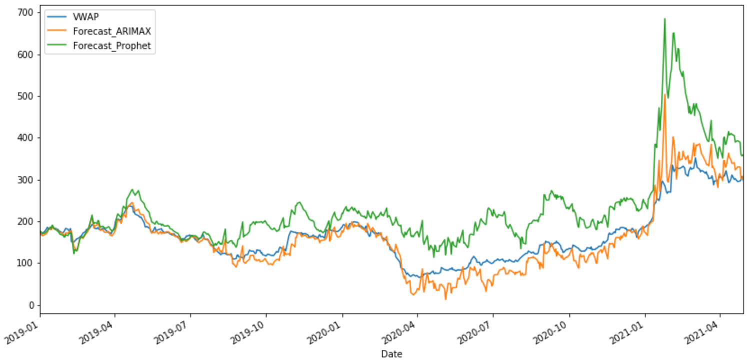
**Tata motors:**



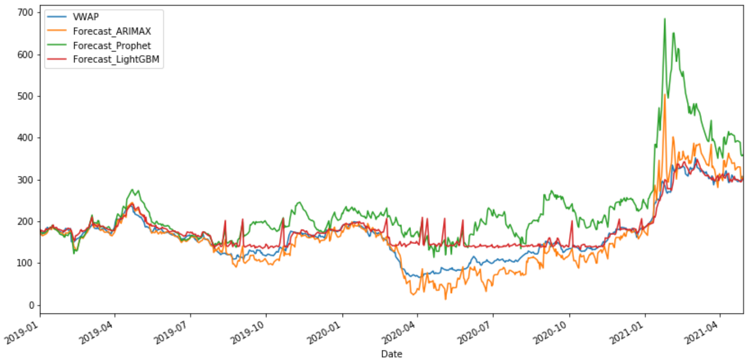

| ***MODEL NAME*** | ***MAE*** | ***RMSE*** |
|---|---|---|
| ***ARIMAX*** | ***18.0883*** | ***25.2581*** |



Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)




| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 18.0883 | 25.2581 |
| PROPHET | 61.8025 | 81.9062 |




| MODEL NAME | MAE | RMSE |
|---|---|---|
| ARIMAX | 18.0883 | 25.2581 |
| PROPHET | 61.8025 | 81.9062 |
| Light GBM | 17.6242 | 27.5185 |

After fitting the models, a forecast of the close price of the testing data is performed. To evaluate the performance of the prediction of the stock price data of the respective models, we have made use of root mean square error (RMSE) and mean absolute error (MAE). Finally, the RMSE and MAE measures of the respective models are compared and visualized.

**Conclusions:** we can conclude that from the emphirical investigation,

**Tata Steel** : The ARIMAX model is recommended for stock price prediction of TATA STEEL . ARIMAX models are known for their ability to incorporate both auto regressive and moving average companies, as well as exogeneous variables, which can be useful for the capturing the dynamics of TATA STEEL's stock price.

**Adani Ports :** The Light GBM ( Gradient Boosting Machine ) model is suggested for stock price prediction of ADANI PORTS . Light GBM is powerful macine learning algorithm that can handle large amounts of dta and capture complex relationships . It may be well – suited for predicting the stock price of ADANI PORTS, potentially considering various features and patterns in the data.

**Sunpharma :** The PROPHET is recommended for stock price prediction of SUNPHARMA. PROPHET is a time series forecasting method developed by Facebook that is designed to handle seasonality, trends, another time – based patterns. It may be particularly effecting for predicting the stock price of SUNPHARMA, considering factors such as seasonality in the pharmaceutical or industry specific events affecting the company.

**Tata Motors**: TATA MOTORS is similar to TATA STEEL , the ARIMAX model is suggested for stock price prediction of TATA MOTORS. This model can help capture the auto correlation, moving average and exogeneous variables that may impact the stock price of TATA MOTORS

**TCS :** The ARIMAX model is also recommended for stock price prediction of TCS . This suggests that considering the auto correlation, moving average and exogeneous varables can be useful for predicting the stock price of TCS

| Company name | Best model |
|---|---|
| Tata steel | Arimax |
| Adaniports | Light gbm |
| Sunpharama | Prophet |
| Tata motors | Arimax |
| TCS | Arimax |

The choice of these models may depend on factors such as historical data availibility, the specific charactristics of each company's stock, and the expertise of the individuals or teams performing the predictions. It's important to note that the performance of these models may vary depending on thde specific dataset, features and pther factors. Therefore, it's recommended to evaluate and compare the performance of these models approprriate metrices before drawing any definitive conclusions about their effectiveness for stock price prediction.

## References

[1] Ding, Y. Zhang, T. Liu, and J. Duan, (2015). Deep learning for event-driven stock prediction. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

[2] Kim, Y. S., Rachev, S. T., Bianchi, M. L., Mitov, I. and Fabozzi, F. J. (2011). Time series analysis for financial market meltdowns, Journal of Banking & Finance 35(8): 1879– 1891.

[3] Madhusudan.G., Sarvana.M., Abbaiah.R(2017)" The Role of Statistical Packages in Academic and Industry", An International Multidisciplinary Research Journal (IJMR), ISSN: 2249-7137 Vol. 7 Issue 10, October 2017 Impact Factor: SJIF =5.099 DOI NUMBER: 10.5958/2249-7137.2017.00097.0

[4] Madhusudan.G., Abbaiah.R, (2018) "Prediction and Analysis of Daucus Carota Yields in Five Districts At Karnataka State", International Journal of Recent Scientific Research (IJRSR) ISSN:0976-3031 Vol. 9 Issue 2(k), pp.24534-24537, February, 2018 Impact Factor 2017: 7.383 Article DOI:http://dx.doi.org/10.21474/ijrsr.2018.0902.1678; UGC Journal no: 49336

[5] Mehtab, S. and Sen, J. (2019). A robust predictive model for stock price prediction using deep learning and natural language processing, Available at SSRN 3502624.

[6] N. Co, S. Huu, H. Hoang, L. Ngo, and N. T. Tran, (2020). "Comparison between ARIMA and LSTM-RNN for VN-index prediction," *Advances in Intelligent Systems and Computing, Intelligent Human Systems Integration 2020*, vol. 1131, no. 10, 2020.

[7] Saud, A. S. and Shakya, S. (2020). Analysis of look back period for stock price prediction with rnn variants: A case study on banking sector of nepse, Procedia Computer Science

[8] Sean J. Taylor & Benjamin Letham (2018) Forecasting at Scale, The American Statistician, 72:1, 37-45, DOI: 10.1080/00031305.2017.1380080

[9] Wen, M., Li, P., Zhang, L. and Chen, Y. (2019). Stock market trend prediction using

[10] high-order information of time series, Ieee Access 7: 28299–28308. 167: 788–798

[11] Zhang, T., Song, S., Li, S., Ma, L., Pan, S. and Han, L. (2019). Research on gas concentration prediction models based on lstm multidimensional time series, Energies 12(1): 161

[12] Madhu Sudan G., et all (2017) "The Role of Statistical Packages in Academic and Industry", An International Multidisciplinary Research Journal (**IJMR**), ISSN: 2249-7137 Vol. 7 Issue 10, October 2017 Impact Factor: **SJIF =5.099** DOI NUMBER: 10.5958/2249-7137.2017.00097.0

[13] M Ramesh, P Vishnu Priya, P Srivyshnavi, G Madhusudan, K Aswini and P Balasiddamuni., (2018) "Estimation of parameters of a multivariate time series var model": International Journal of Statistics and Applied Mathematics(**IJSAM**) **ISSN:** 2456-1452, 2018; 3(1): 442-450