*Nanotechnology Perceptions*
*ISSN 1660-6795*
*www.nano-ntp.com*

# Forecasting And Outlier Detection Using Facebook's Prophet

**K. Senthamarai Kannan, [1] , D. Kabinath [2*] , V. Parimyndhan[3]**

[1]*Professor Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012. Email: senkannan2002@gmail.com,*
[2*] *Research Scholar Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012. Email: kabi130198@gmail.com*
[3]*V. Parimyndhan Research Scholar Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012. Email: parimyndhanvmp@gmail.com*

In this work, detecting outliers with time series forecasting using Facebook's Prophet. Time series data refers to any information that is recorded and organized concerning time intervals, such as daily, hourly, monthly, or yearly. The findings of our study demonstrate that Prophet is superior to traditional statistical methods and machine learning algorithms in detecting anomalies in time series data that exhibit noise and non -stationarity. Additionally, Prophet's computational efficiency and ability to handle large volumes of data were observed. Furthermore, the results indicate that the incorporation of external regressors in Prophet can enhance its accuracy in detecting anomalies. The suggested unsupervised algorithm can identify anomalies in customer orders by utilizing constraints such as the day and date. The algorithm has demonstrated superior performance in terms of RMSE score. These observations lead us to conclude that Prophet is a promising tool for time series outlier detection, particularly in complex and dynamic environments.

**Keywords**: Forecasting, Random Forest, Technical Indicators, Machine learning, Stock. Assay.

## 1.INTRODUCTION

Outliers play a critical role in the identification and management of data points that exhibit substantial deviations from expected patterns within a dataset. These exceptional observations can stem from factors like measurement inaccuracies, system glitches, or extraordinary occurrences. The detection and handling of outliers are indispensable in the realms of data analysis and forecasting as they help guarantee precise insights and dependable predictions.

Prophet, an open-source forecasting library developed by Facebook's Core Data Science team, is a robust tool for predicting time series data[7]. Besides its primary function of forecasting, Prophet also provides useful features for detecting outliers within time series datasets. By utilizing its advanced statistical modeling techniques and adjustable parameters, Prophet enables analysts and data scientists to identify and examine anomalies in their data.

Detecting outliers using Facebook Prophet involves several essential steps[6]. Initially, the time series data is prepared to ensure it adheres to the necessary format and quality criteria. Subsequently, Prophet's integrated outlier detection capability is utilized to identify potential anomalies in the dataset. This is accomplished by fitting a Bayesian structural time series model to the data and evaluating the residuals, which depict the disparities between the observed and predicted values.

Prophet offers various choices for outlier identification, such as additive outliers and temporary changes. Additive outliers denote abrupt shifts in the time series data that do not return to their previous levels, while temporary changes indicate temporary deviations from the underlying trend[4]. By detecting and examining these outliers, analysts acquire valuable insights into unusual patterns present in the data.

After identifying the outliers, analysts can proceed with additional analysis and decision-making procedures. The outliers can be thoroughly examined to comprehend their root causes, evaluate their influence on the overall dataset, and determine the necessity for corrective measures [9]. In certain instances, it may be necessary to remove or adjust the outliers to ensure the accuracy of subsequent analyses or forecasts [2].

In this work, we utilized the State Bank of India Stock price used. The State Bank of India is a financial institution that offers a wide range of banking products and services to various customers, both in India and internationally. It operates through different segments, including Treasury, Corporate/Wholesale Banking, Retail Banking, Insurance Business, and Other Banking Businesses. The bank provides personal banking services such as current accounts, savings accounts, salary accounts, fixed and recurring deposits, and various types of loans, including home, personal, auto, education, and gold loans. It also offers services like overdrafts, mutual funds, insurance, equity trading, portfolio investment schemes, remittance services, and digital banking solutions. In addition, the State Bank of India caters to corporate customers by offering corporate accounts, working capital and project finance, term loans, structured finance, equipment leasing, syndicated loans, cash management, and trade-related products. The bank also engages in activities related to the insurance business and other banking services.

To summarize, Facebook Prophet's outlier detection is a robust method for identifying and analysing abnormal patterns in time series data. By utilizing Prophet's advanced modeling features, analysts can gain valuable insights into outliers, comprehend their impact, and make informed decisions. This process ultimately leads to more accurate forecasting, enhanced data analysis, and improved decision-making capabilities.

The remainder of the paper is organized in the following manner: Section 2 deals with the survey of the literature and introduction of outlier detection and stock market prediction. The research methodology and experimental schedule are presented in Section 3. To review the research findings, Section 4 discusses the findings. The article is concluded in Section 6, which also discusses future work.

## 1. Literature Survey

This section presents Prophet frameworks in particular Stock and discusses the unique ML methods that are frequently applied to stock market outlier detection and forecasting.

Toharudin et al., (2021) built the five-year daily air temperatures in Bandung by using both LSTM and Facebook Prophet. The results indicate that Prophet outperforms LSTM in forecasting maximum air temperature, while LSTM performs better for minimum air temperature. Although the difference in RMSE values is not substantial, it is still noteworthy [10].

Amarbayasgalan et al., (2019) proposed a framework that includes an additional stage for outlier detection, attempting to improve the identification of comparable products that are commonly bought together. By eliminating anomalies, the framework aims to provide more accurate information about these related products. This is achieved by implementing a dimensional model in the Hive data warehouse, which allows for efficient querying and analysis of the data [1].

Harish et al., (2022) Prophet is a forecasting technique that specializes in predicting time series data. It employs an additive model that captures non-linear trends, making it suitable for datasets that exhibit holiday impacts, daily, weekly, and yearly seasonality, as well as pronounced seasonal effects. Prophet is particularly well-suited for time series data characterized by significant seasonal patterns, allowing for accurate forecasting by incorporating these factors into the model [5].

Li et al., (2022) introduce a novel approach for outlier detection in time series data by employing the Nonhomogeneous Poisson Process (NHPP) and multilayer perceptron (MLP). Additionally, the study incorporates three periodic prediction methods, namely AUTO-ARIMA, Prophet, and Trendy and Seasonal Linear Model (TSLM), to perform multi-step time series prediction. The outlier correction process is utilized as the basis for these prediction methods. Overall, this study presents innovative techniques for outlier detection and utilizes various prediction models to improve multi-step forecasting accuracy in time series data [8].

Cicceri et al., (2021) revealed that the machine learning solution, specifically LSTM, outperformed classical statistical approaches such as ARIMA and FP in terms of accuracy, efficiency, and effectiveness in monitoring and controlling wastewater purification processes. The ML solution demonstrated superior performance in accurately predicting and managing the purification processes, highlighting its potential for enhancing wastewater treatment operations[3].

## 2. Material and Methods

### 3.1 Materials

We selected the closing price of Nifty Healthcare Index from National Stock Exchange from 22/11/2021 To 31/10/2024 for our research.

### 3.2 Methods

**Prophet**

Prophet is a type of algorithm used for forecasting future values in a time series. It analyses the time series and separates it into three distinct components, namely seasonal patterns, overall trends, and holiday effects.

$Y(t) = g(t) + s(t) + h(t) + \in t$

In the Prophet algorithm, three main functions capture different types of changes in a time series. The trend function, denoted as g(t), represents non-seasonal changes in the data. The seasonal change function, denoted as s(t), captures the repeating patterns that occur at regular intervals. The holiday function, denoted as h(t), accounts for the effects of holidays and other special events that occur at irregular intervals. The trend function g(t) can be modeled using either a saturating growth model or a piecewise linear model. In the saturating growth model, g(t) is defined using the logistic growth equation.

$g(t) = g(t) = \frac{c}{1+\exp\left(-(k(t-(m))\right)}$

The growth rate of a population can be described by a mathematical formula that takes into account the carrying capacity (C), the growth rate (k), and an offset specification (m). If we want to incorporate trend updates into the model, we can introduce change points ($s_j$) where the growth rate can change at specific times (t). Let's say there are S change points, occurring at times $s_j$ ($j = 1, ..., S$). Where $\delta_j$ represents the change in rate that occurs at time $s_j$. The rate at any given time t can then be calculated by applying these rate adjustments to the original growth rate equation.

$$T = k + a(t)^T\delta$$

where a(t)T δ is the cumulative growth until changepoints sj and $a(t) \in \{0, 1\}^S$ is a the vector can be computed as follows:

$$a_j(t)= \begin{cases} 1, & if\ t\ \geq sj \\ 0, & Otherwise. \end{cases}$$

Then, the prophet modifies the primary logistic growth model to include trend updates for non-linear, saturating growth as follows:

$$g(t) = \frac{c(t)}{1+\exp\left(-(k+a(t)^T\delta)\ (t-(m+a(t)^T\Upsilon))\right)}$$

and the linear growth can be drafted as follows:

$$g(t) = (k + a(t)^T\delta)t + (m + a(t)^T\delta)$$

Let $\delta \in R^S$ such that points in δ are the rate of modifications in g(t). The allocation of change points is made by assigning δ using the Laplacian distribution ($\delta_j \sim$ Laplace $(0, \tau)$), where $\tau$ controls the compliance of growth rate and $\gamma_j$ is set to $-sj\delta_j$ to make the function continuous.

**4.Results and Discussion**

The main objective of our research was to assess the performance of Facebook's Prophet in predicting and identifying exceptional values in time series data. We opted not to divide our Modeling dataset into training and testing sets. This is because our model's aim is not to make predictions on future stock prices but to accurately forecast past prices. As a result, we will employ the entire dataset for both training and forecasting purposes. The Prophet model was initialized by explicitly enabling the yearly seasonality and weekly seasonality, followed by fitting the model to the training set of data. Additionally, the interval width was set to 0.99, indicating that the uncertainty interval would be 99%.

For outlier detection, we have maintained a simple model.

**4.1 Make Predictions Using the Prophet**

Once the model is constructed, we use it to generate predictions for the dataset. The resulting forecast plot indicates that the majority of the predicted values are consistent with the actual values.

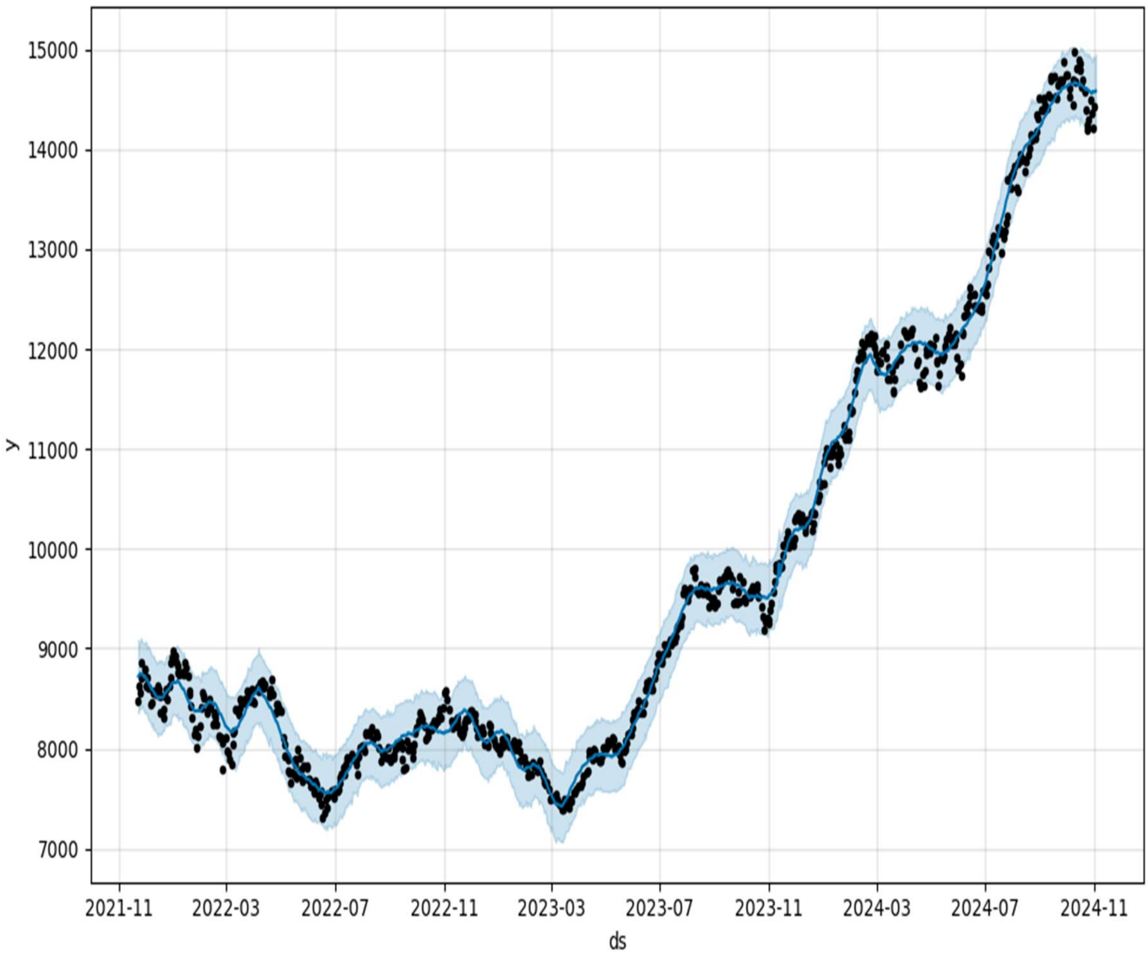


Figure 1: Prophet model prediction

We can also check the components plot for the trend, weekly seasonality, and yearly seasonality.


**Figure 2: Prophet model component**

## 4.2 Check Time Series Model Performance

It's critical to assess the time series model's performance at this point. It is necessary to combine the forecast data frame with the actual data frame in order to compare the predicted values with the actual values because the forecast data frame only contains predicted values. For this evaluation, two performance metrics are employed.

The Mean Absolute Error (MAE), which is calculated by adding the absolute differences between the predicted and actual values and dividing the result by the number of predictions, determines the average absolute difference between the predicted and actual values.

The Mean Absolute Percentage Error (MAPE), on the other hand, determines the typical percentage difference between the predicted and actual values. It can also be calculated by adding up and dividing by the total number of predictions the absolute percentage difference between the predicted and actual values. The advantage of MAPE is that it is not dependent on the scale of the data, allowing for the comparison of forecasts from different datasets.

However, the MAPE cannot be used when the actual value is zero, as the percentage difference becomes undefined.

The mean absolute error (MAE) for the model is 16. To clarify, the baseline model's mean absolute percent error (MAPE) is 4.5%, which indicates that the average difference between the forecasted stock price and the actual stock price is 4.5%.

To detect anomalies in a time series, we need to first establish an uncertainty interval. This interval represents the range of values within which we expect the actual value of the time series to fall with a certain level of confidence. If the actual value of the time series falls outside of this uncertainty interval, we consider it to be an outlier. Specifically, if the actual value is less than the lower bound or greater than the upper bound of the uncertainty interval, we set an outlier indicator to 1. If the actual value is within the bounds of the uncertainty

interval, we set the outlier indicator to 0.
There are 13 outliers out of 733 data points. After printing out the anomalies, we can see that all the outliers are lower than the lower bound of the uncertainty interval.

**Table 1: Time series outlier detection results**

| Sl.no | Date | y | yhat | yhat_lower | yhat_upper | Outlier |
|---|---|---|---|---|---|---|
| 48 | 27-01-2022 | 8005 | 8378 | 8022 | 8711 | 1 |
| 68 | 24-02-2022 | 7801 | 8315 | 7934 | 8662 | 1 |
| 236 | 01-11-2022 | 8564 | 8153 | 7811 | 8506 | 1 |
| 237 | 02-11-2022 | 8580 | 8169 | 7827 | 8484 | 1 |
| 238 | 03-11-2022 | 8582 | 8166 | 7807 | 8557 | 1 |
| 480 | 26-10-2023 | 9198 | 9517 | 9214 | 9871 | 1 |
| 598 | 18-04-2024 | 11672 | 12072 | 11736 | 12404 | 1 |
| 599 | 19-04-2024 | 11625 | 12070 | 11751 | 12436 | 1 |
| 601 | 23-04-2024 | 11639 | 12049 | 11704 | 12427 | 1 |
| 630 | 04-06-2024 | 11737 | 12176 | 11807 | 12481 | 1 |
| 727 | 23-10-2024 | 14189 | 14609 | 14230 | 14972 | 1 |
| 728 | 24-10-2024 | 14241 | 14597 | 14272 | 14889 | 1 |
| 732 | 30-10-2024 | 14208 | 14579 | 14247 | 14940 | 1 |

The visualization includes dots that show actual values, and a black line that shows predicted values. Any dots in orange are considered outliers.

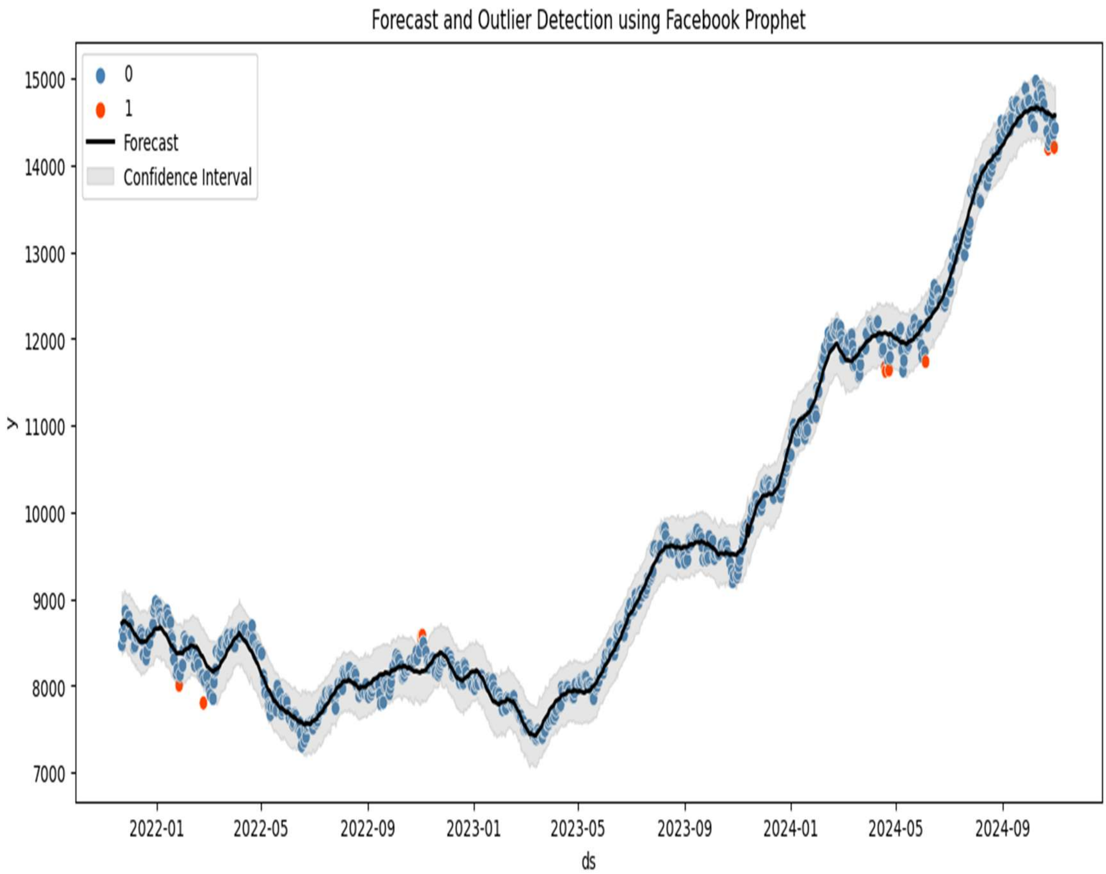


Figure 3: Prophet time series outlier detection results

## 5.Conclusion:

The application of outlier detection and forecasting using the Prophet algorithm provides valuable insights and benefits in various domains. This powerful combination offers robust techniques for identifying and handling outliers in time series data while enabling accurate and reliable forecasting. Through the implementation of outlier detection techniques, Data points that significantly deviate from expected patterns or trends can be found and flagged. By effectively identifying outliers, we can mitigate their potential impact on the forecasting process and ensure the accuracy and reliability of future predictions.

Prophet, a state-of-the-art forecasting algorithm, utilizes sophisticated modeling techniques that account for seasonality, trend changes, and other important factors inherent in time series data. By incorporating these features, Prophet provides reliable forecasts that take historical patterns and potential future changes into consideration, resulting in improved accuracy and better decision-making. The combination of outlier detection and forecasting using Prophet empowers businesses and researchers to make informed decisions based on reliable insights. It enables the identification of anomalies that could impact the forecasting process and facilitates more accurate predictions for future trends and events. This integrated approach can be applied across various industries, such as supply chain management, finance, sales forecasting and more.

## References:


1. Amarbayasgalan, T., Li, M., Namsrai, O., Jargalsaikhan, B., & Ryu, K. H. (2019). Performance improvement of open source based business intelligence system using database modeling and outlier detection. In Proceedings of the Asian Conference on Intelligent Information and Database Systems.

2. Ben-Gal, I. (2005). Outlier detection. In Maimon, O., & Rokach, L. (Eds.), Data mining and knowledge discovery handbook (pp. 131–146). Springer.

3. Cicceri, G., Maisano, R., Morey, N., & Distefano, S. (2021). A machine learning approach for anomaly detection in environmental IoT-driven wastewater purification systems. Environmental Science and Pollution Research, 28(33), 45182–45193.

4. Egan, W. J., & Morgan, S. L. (1998). Outlier detection in multivariate analytical chemical data. Analytical Chemistry, 70(11), 2372–2379.

5. Harish, N., Likith, H. V., Yashwanth, G., Krishnaswamy, N., Eshanya, & Sunil, G. L. (2022). Stock index probability prediction using the FB Prophet model. 2022 International Conference on Futuristic Technologies (INCOFT), 1–5.

6. Kannan, K. S., & Manoj, K. (2015). Outlier detection in multivariate data. Applied Mathematical Sciences, 9(47), 2317–2324.
7. Kannan, K. S., Sekar, P. S., Sathik, M. M., & Arumugam, P. (2010). Financial stock market forecast using data mining techniques. Proceedings of the International Multiconference of Engineers and Computer Scientists, 1, 4.

8. Li, Q., Zhu, J., Yu, Y., & Shi, W. (2022). Anomaly detection and prediction of base station KPI. 2022 7th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 48–52.

9. Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. Journal of the American Statistical Association, 91(435), 1047–1061.

10. Toharudin, T., Pontoh, R. S., Caraka, R. E., Zahroh, S., Lee, Y., & Chen, R. C. (2023). Employing long short-term memory and Facebook Prophet model in air temperature forecasting. Communications in Statistics - Simulation and Computation, 52(2), 279–290.