SASTRA

ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION

DEEMED TO BE UNIVERSITY

(U/S 3 OF THE UGC ACT, 1956)

THINK MERIT | THINK TRANSPARENCY | THINK SASTRA

SCHOOL OF ARTS, SCIENCE & HUMANITIES
THANJAVUR, TAMIL NADU – 613401

Stock Market prediction using XG Boost and Fb Prophet algorithms

Report submitted to the SASTRA Deemed to be University as the requirement for the course

INT424: ALGORITHMIC TRADING

Submitted by

SENTHOORAN E

(Reg. No.: 123150034)

July-2022

SASTRA

ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION

DEEMED TO BE UNIVERSITY

(U/S 3 OF THE UGC ACT, 1956)

THINK MERIT | THINK TRANSPARENCY | THINK SASTRA

SCHOOL OF ARTS, SCIENCE & HUMANITIES
THANJAVUR, TAMIL NADU – 613401

Certificate

This is to certify that the report titled **"Stock Market prediction using XG Boost and Fb Prophet algorithms"** was submitted as a requirement for the course, **INT424: Algorithmic Trading** for M.sc Data Science program, is a record of analysis submitted by Mr. Senthooran E (Reg. No.123150034) during the academic year 2programn the School of Computing.

Signature of Faculty in charge :

Name with Affiliation : **DR. Ashok Palaniappan PhD**

Date : **10-07-2022**

Project *Viva-voce* held on 11-07-2022

11/07/2022

Examiner 1 **Examiner 2**

SASTRA

ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION

DEEMED TO BE UNIVERSITY

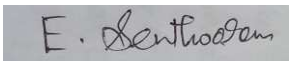
(U/S 3 OF THE UGC ACT, 1956)

THINK MERIT | THINK TRANSPARENCY | THINK SASTRA

SCHOOL OF ARTS, SCIENCE & HUMANITIES
THANJAVUR, TAMIL NADU – 613401

Declaration

I declare that the report titled **"Stock Market prediction using XG Boost and Fb Prophet algorithms"** submitted by me is an analysis done by me during the even semester of the academic year 2021-22, in the School of Arts & Science. The work is original and wherever I have used materials from other sources, I have given due credit and cited them in the text of the report.

Signature of the candidate : 

Name of the candidate : Senthooran E

Date :10-07-2022

TABLE OF CONTENT

List of Figures

ABSTRACT

In this project, I try to put a machine learning strategy for stock price prediction into practice. Stock price forecasting uses machine learning effectively. In order to make wiser and more accurate financial decisions, the goal is to forecast stock prices. In order to improve stock forecast accuracy and generate lucrative trades, we provide a stock price prediction system that combines mathematical functions, machine learning, and other external aspects.

Keywords:

Stock market prediction, forecasting, time series analysis, XG Boosting, Fb Prophet, and machine learning.

INTRODUCTION

People can buy and sell currencies, stocks, equities, and derivatives through virtual platforms supported by brokers in the financial market, which is a dynamic and composite system. Investors can purchase shares of publicly traded corporations on the stock market through trading on exchanges or off-exchange venues. When compared to the danger of starting a new business or the requirement for high-paying employment, this market offers investors the potential to make money and live affluent lives by investing small initial sums of money. Numerous factors have an impact on stock markets, creating unpredictability and significant market volatility.

In many practical applications, including weather forecasting and financial market prediction, time-series prediction is an extensively utilized technique. To forecast the outcome in the following time unit, it employs continuous data collected over a period of time. Numerous time series prediction algorithms have proven to be effective in real-world settings. The stock market is a typical setting for time-series data, and numerous researchers have examined it and suggested numerous models. In this project, the stock price is predicted using the Fb Prophet model and XG Boosting.

1.1 Problem statement

In data analysis, time series forecasting and modeling are crucial. Time series analysis is a specific area of statistics that is often employed in areas like operation research and econometrics. Time Series is a frequently used tool in data science and analytics. Stock prices are erratic and subject to a variety of influences. Utilizing XGBOOSTING and Fb Prophet, the primary objective of this project is to forecast stock values.

1.2 Stock market

The stock market is a marketplace where buyers and sellers can transact on publicly traded shares at particular times of the day. Share market and stock market are frequently used interchangeably.

The National Stock Exchange (NSE) and the Bombay Stock Exchange are the two main stock markets in India (BSE).

1.3 Dataset

	Date	Open	High	Low	Close	Adj Close	Volume
0	2000-01-03	122.070313	122.071877	122.070313	122.071877	84.002754	657536.0
1	2000-01-04	131.250000	131.837891	126.976563	131.686722	90.619118	7191808.0
2	2000-01-05	122.664063	126.953125	121.151955	121.151955	83.369713	21069440.0
3	2000-01-06	119.192581	119.531250	111.459763	111.459763	76.700089	10495616.0
4	2000-01-07	102.542969	102.542969	102.542969	102.542969	70.564087	971392.0

Fig 1: Dataset of stock market

TIME SERIES ANALYSIS

A particular method of examining a set of data points gathered over a period of time is called a "time series analysis." Instead of only capturing the data points irregularly or arbitrarily, time series analysis records the data points at regular intervals over a period of time. To maintain consistency and reliability, time series analysis often needs a lot of data. A large data collection guarantees that your analysis can sift through erratic data and that your sample size is representative. Additionally, it guarantees that any trends or patterns are not outliers and can take seasonal variation into consideration. Time series data can be employed for forecasting, which is the process of anticipating future data based on the past.

Example of time series data:

- Stock prices data
- Weather data

MODELS

3.1 Introduction of Machine learning

Artificial intelligence, which is widely defined as a machine's ability to mimic intelligent human behavior, includes the subfield of machine learning. Artificial intelligence (AI) systems are used to carry out complicated tasks in a manner akin to how people solve issues. A method of using AI is machine learning. AI pioneer Arthur Samuel referred to it as "the branch of study that offers computers the power to learn without explicitly being programmed" in the 1950s. Data—numbers, images, or text—is the foundation of machine learning. Examples of this include bank transactions, repair records, and time-series data from sales reports.

A machine learning model will be trained on the data, which has been collected and prepared for use as training data. The application performs better with additional data. The next step is for programmers to choose a machine learning model to employ, provide the data, and then sit back and watch as the computer model learns to spot patterns or forecast outcomes. The model can be adjusted over time by the human programmer, including modifying its parameters, to assist it to produce more accurate results.

3.1.1 Subcategories of Machine Learning

They are three subcategories of Machine Learning.

1. **Supervised machine learning**

   Labeled data sets are used to train the models, which helps them develop over time and become more accurate.

2. **Unsupervised machine learning**

   The models are trained on an unlabeled dataset and then given free rein to operate on it.

3. **Reinforcement machine learning**

   By doing actions and observing the effects of those actions, an agent learns how to behave in a given environment.

3.2 Extreme Gradient Boosting (XG Boosting)

"XG Boost is a gradient boosting framework-based ensemble machine learning technique based on decision trees." It is a decision tree-based method variation and is regarded as one of the top gradient boosting techniques for dealing with the variance/bias tradeoff issue. As a result, it performs better and is more effective than other gradient boosting techniques. I can look at the below figure to understand the gradient boosting technique and the benefits of XG Boost in comparison to other bagging methods.

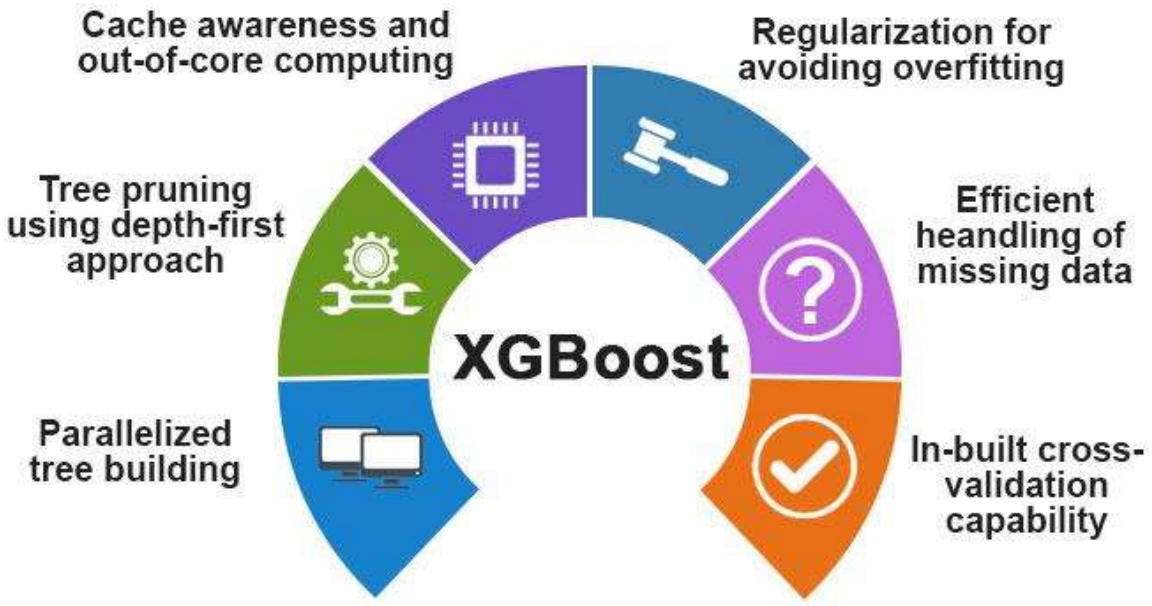


Fig 2: Characteristics of XG Boost

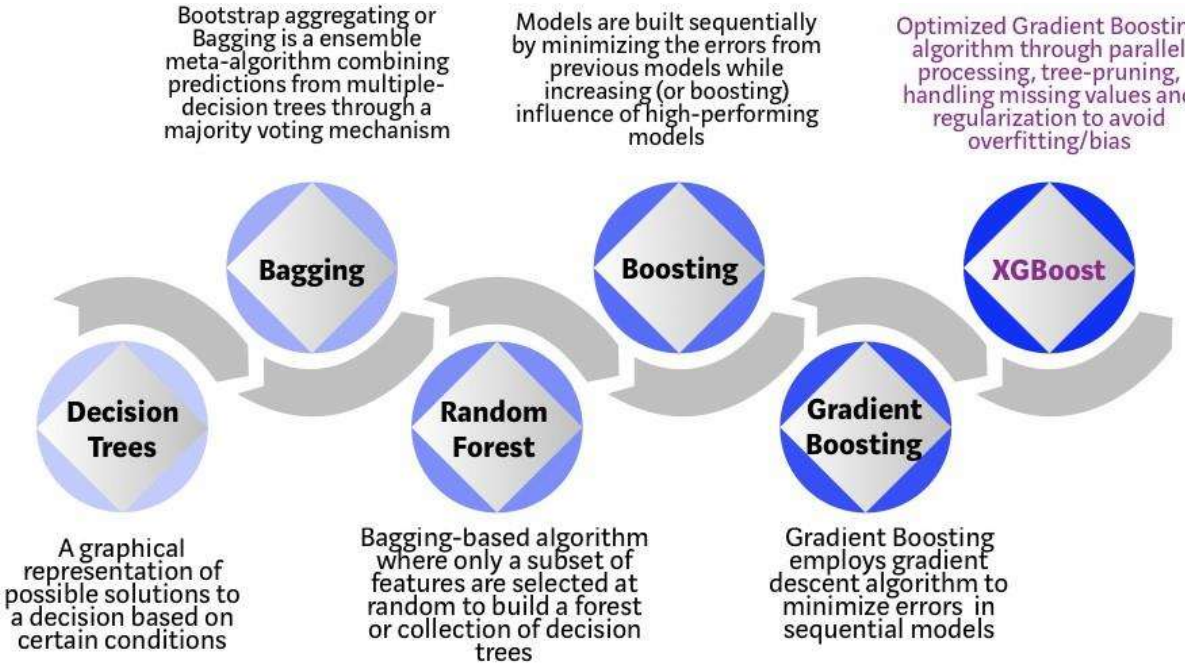


Fig 3: Evaluation of XG Boost

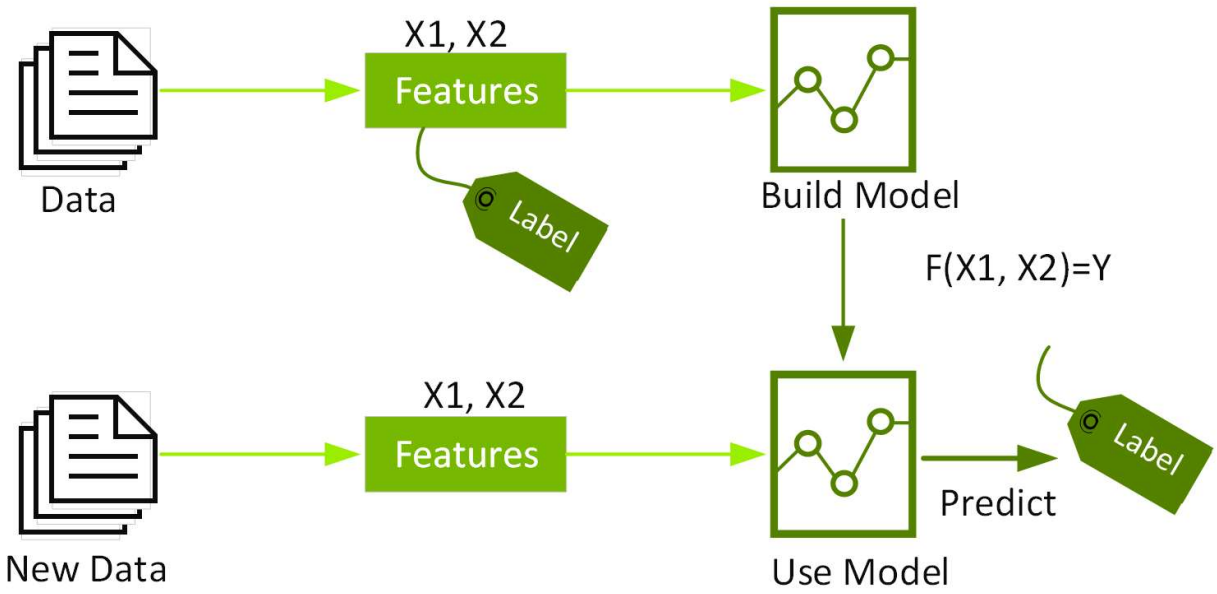


Fig 4: Process of XG Boost

3.3 Facebook (Fb-prophet)

The Facebook Prophet is a Python and R-based open-source forecasting technique. Forecasts are generated automatically. Numerous applications using time series data and the collection of sample time forecast data involve the employment of prophets. With these models, it is impossible to obtain precise future data, but we can determine the trend.

Facebook Prophet using Additive Regressive models using the following four components:

$$y(t) = g(t) + s(t) + h(t) + epsilon\_t$$

- **g(t)**: A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting change points from the data.
- **s(t)**: A yearly seasonal component modeled using the Fourier series and a weekly seasonal component using a dummy variable.
- **h(t)**: A user-provided list of important holidays.
- **epsilon_t**: Error term used by the prophet.

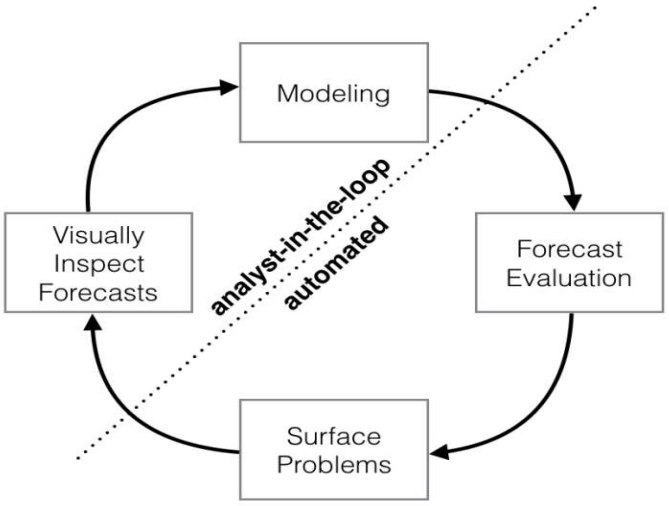


Fig 5: Process of Facebook Prophet

CODES

A computer programmer will write text in a particular language's protocol, and this text is referred to as code (short for source code). Programming languages like C, C#, C++, Java, Perl, and PHP are examples. Text written for markup or styling languages like HTML and CSS can alternatively be referred to as code in a less formal context (Cascading Style Sheets).

4.1 Codes for XG Boosting

```
!pip install xgboost
import xgboost as xgb
import sklearn
from xgboost import XGBRegressor
from xgboost import plot_importance

xgb = XGBRegressor(n_estimators=1000, learning_rate=0.01)
xgb
xgb.fit(X_train3,
y_train1,eval_set=[(X_train3,y_train1),(X_valid3,y_valid1)],early_stopping_rounds=100,verbos
e=True) # Change verbose to True if you want to see it train

plt.figure(figsize=(12,8))
feature_importance = plot_importance(xgb, height=0.9)
feature_importance

predicted_results_v = xgb.predict(X_valid3)
predicted_results_t = xgb.predict(X_train3)
predicted_results_v= predicted_results_v.reshape(-1, 1)
predicted_results_t= predicted_results_t.reshape(-1, 1)
predicted_results_v = scaler1.inverse_transform(predicted_results_v)
predicted_results_t = scaler.inverse_transform(predicted_results_t)

plt.figure(figsize=(20,15))
```

```
plt.plot(X_train2.index,list(y_trainn),label="actual Price train",color='b')
plt.plot(X_train2.index,list(predicted_results_t),label="predicted Price train",color='r')
plt.plot(X_valid2.index,list(y_validn),label="actual Price valid",color='m')
plt.plot(X_valid2.index,list(predicted_results_v),label="predicted Price valid",color='g')
plt.xlabel('Date')
plt.ylabel('INFOSYS closing Price')
plt.legend()
plt.show()
```

4.2 Codes for Facebook Prophet

```
! pip install fbprophet

from fbprophet import Prophet
m = Prophet(daily_seasonality = True)     # the Prophet class (model)
m.fit(df)    # fit the model using all data

future = m.make_future_dataframe(periods=365)      # we need to specify the number of days in future
prediction = m.predict(future)
m.plot(prediction)
plt.title("Prediction of the INFOSYS Stock Price using the Prophet")
plt.xlabel("Date")
plt.ylabel("Close Stock Price")
plt.show()

m.plot_components(prediction)
plt.show()
```

RESULTS

Based on historical data, the proposed XG Boosting and Fb Prophet implementation in Python forecasts the share prices of three firms (TCS, INFOSYS, and WIPRO). In this study, we put into practice an algorithm that forecasts the share price over a specific time period.

The XG Boosting and Facebook Prophet prediction results are plotted. The stock's forecasted closing price values and actual values differ similarly. The two approaches can be viewed as promising approaches for stock price prediction.

5.1 Predicted results of stocks prices by XG Boosting

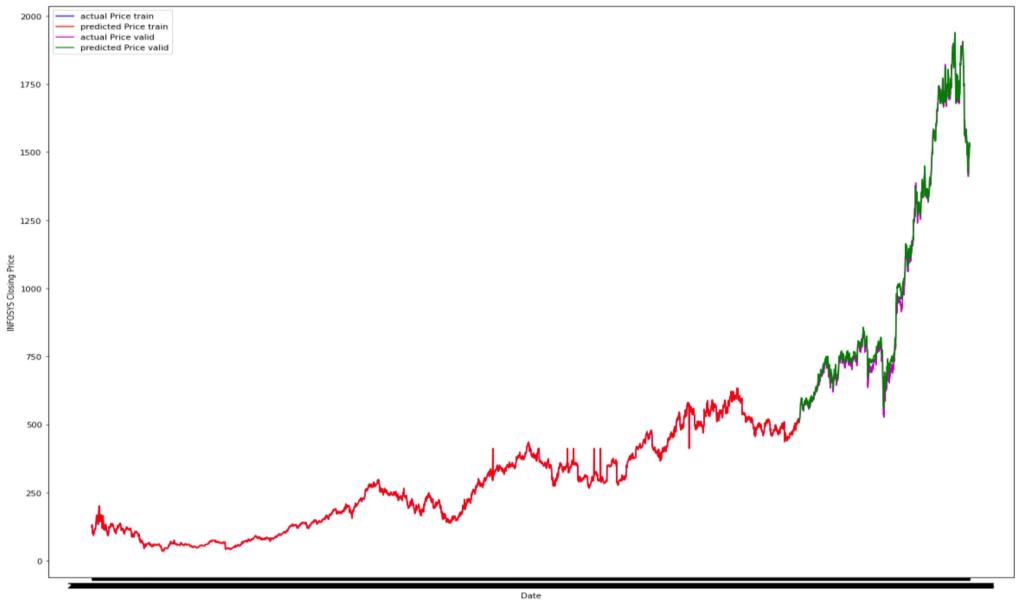


Fig 6: INFOSYS closing price

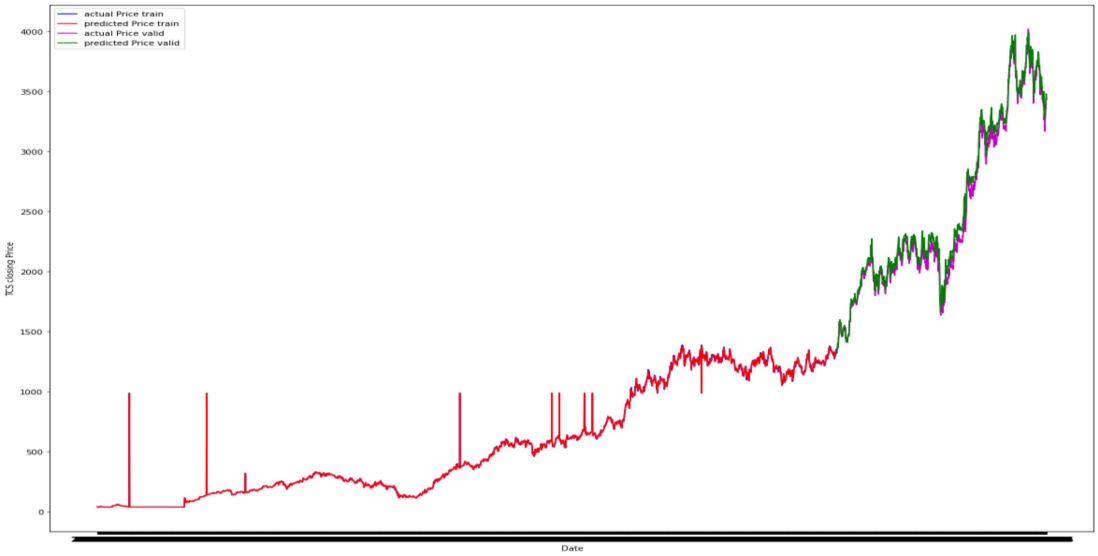


Fig 7: TCS closing price

Fig 8: WIPRO closing price

5.2 Predicted results for Facebook Prophet

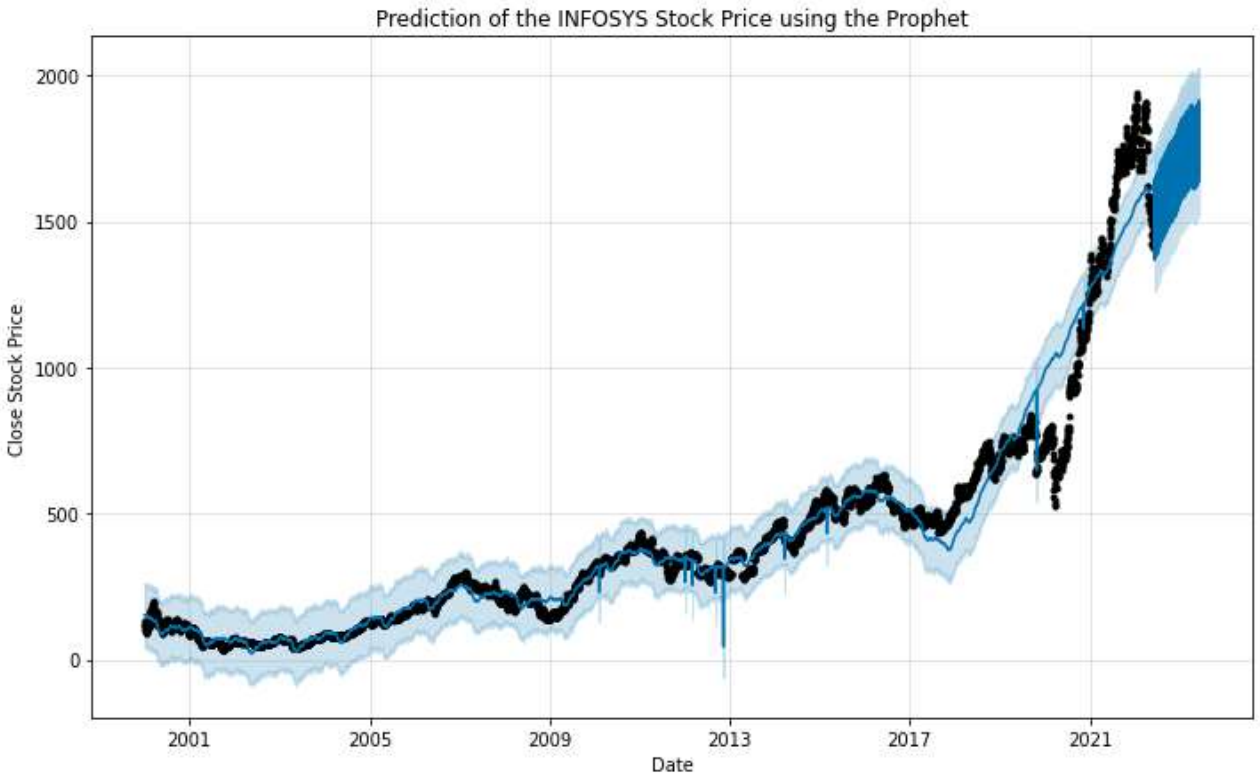


Fig 9: INFOSYS stock price

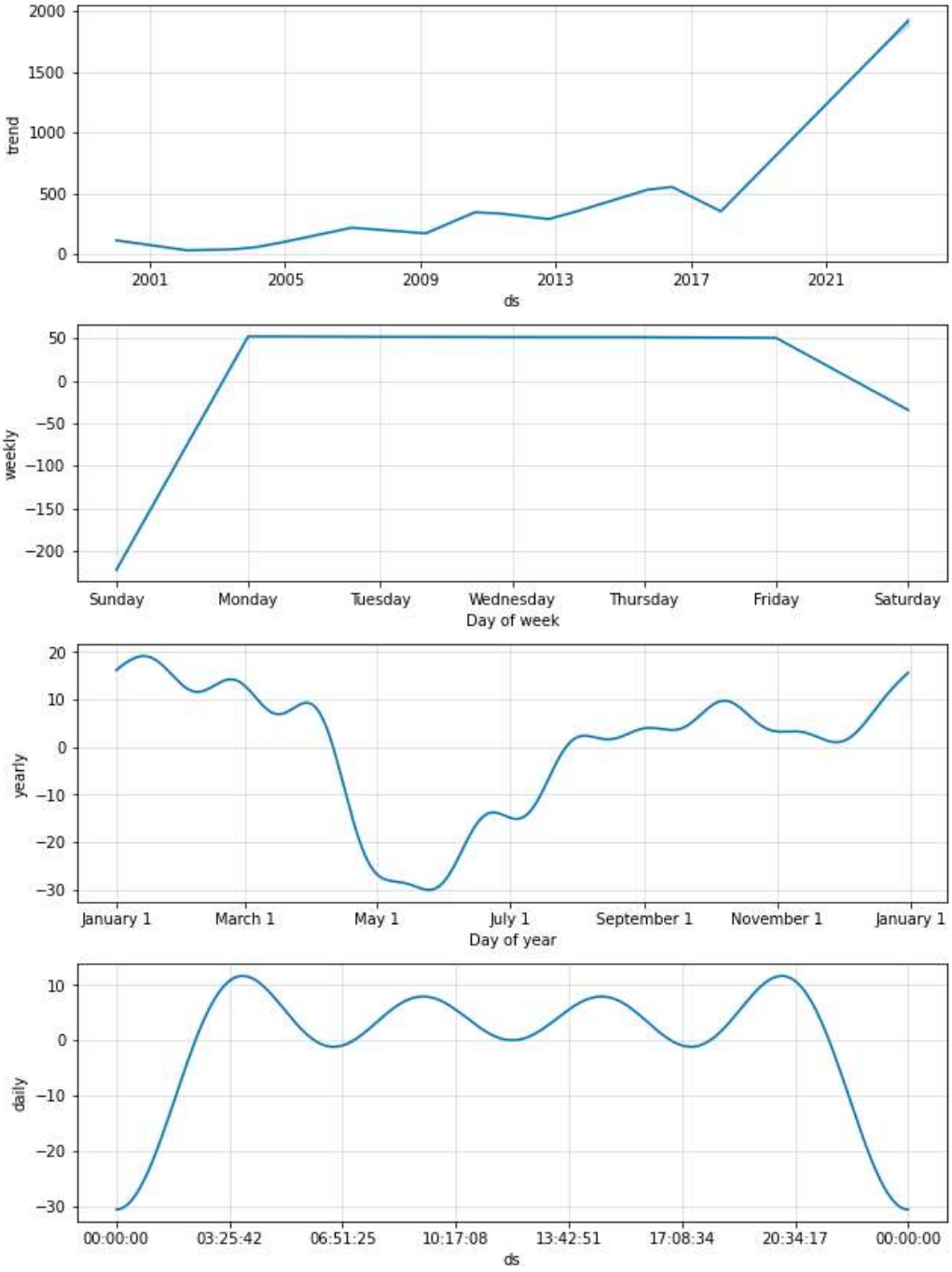


Fig 10: INFOSYS stock Trend lines for daily, monthly, and yearly stock price

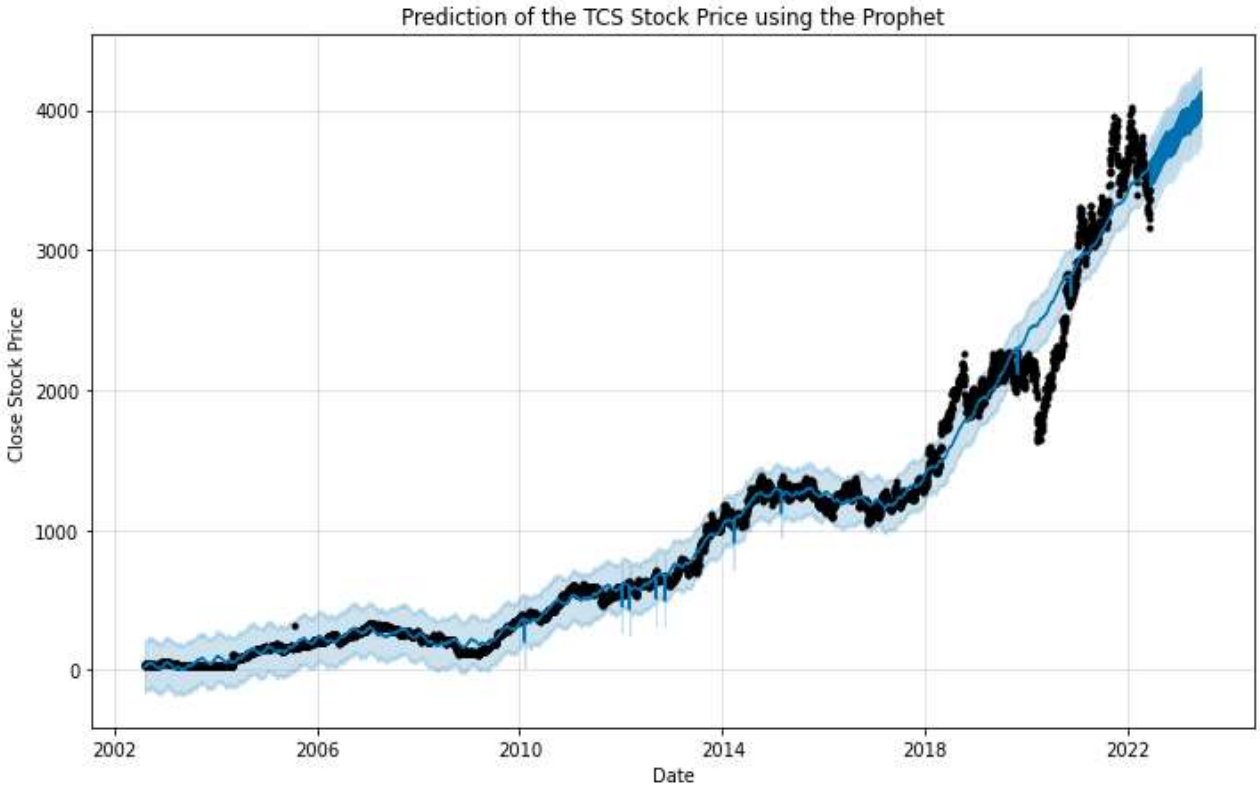


Fig 11: TCS stock price

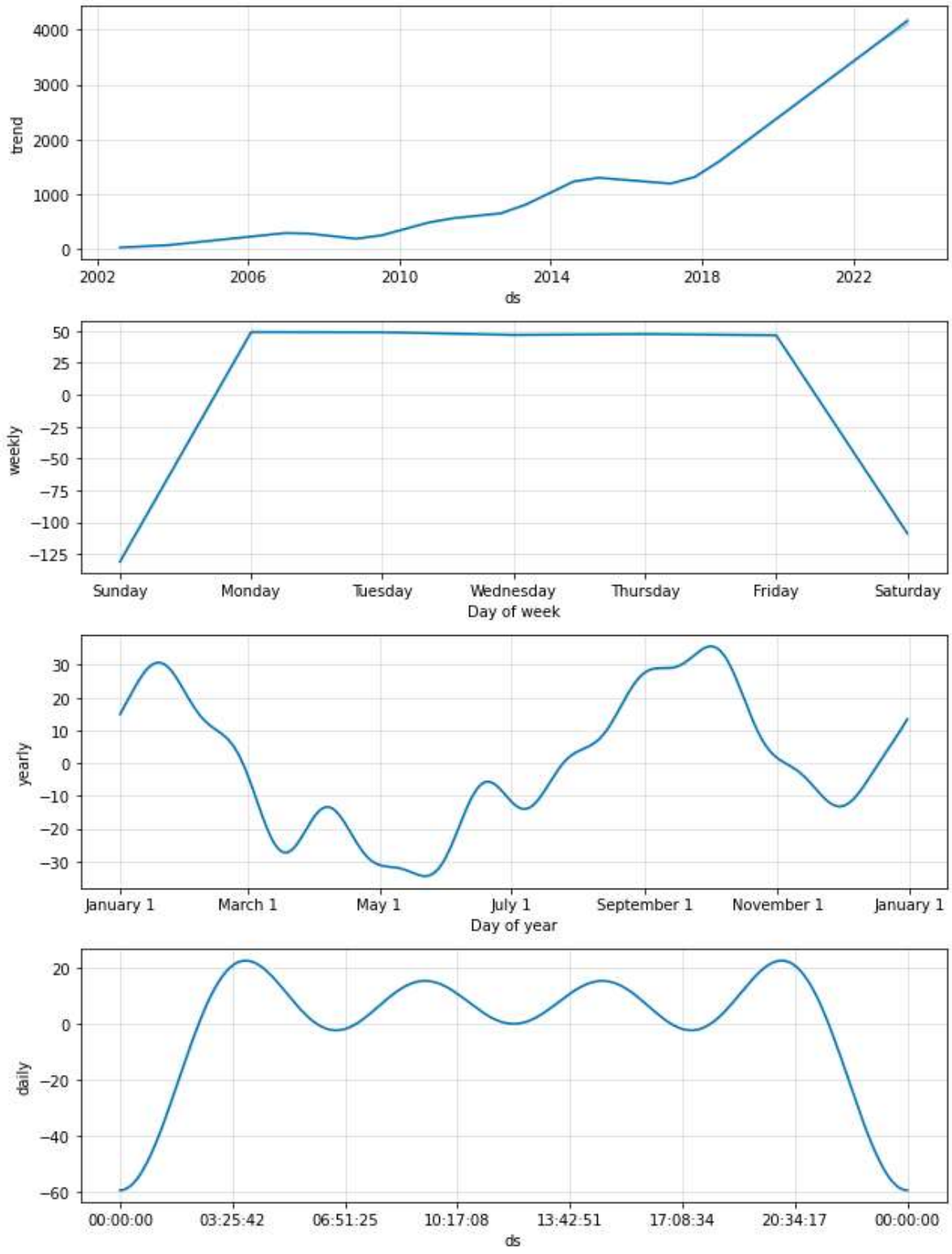


Fig 12: TCS stock Trend lines for daily, monthly, and yearly stock price

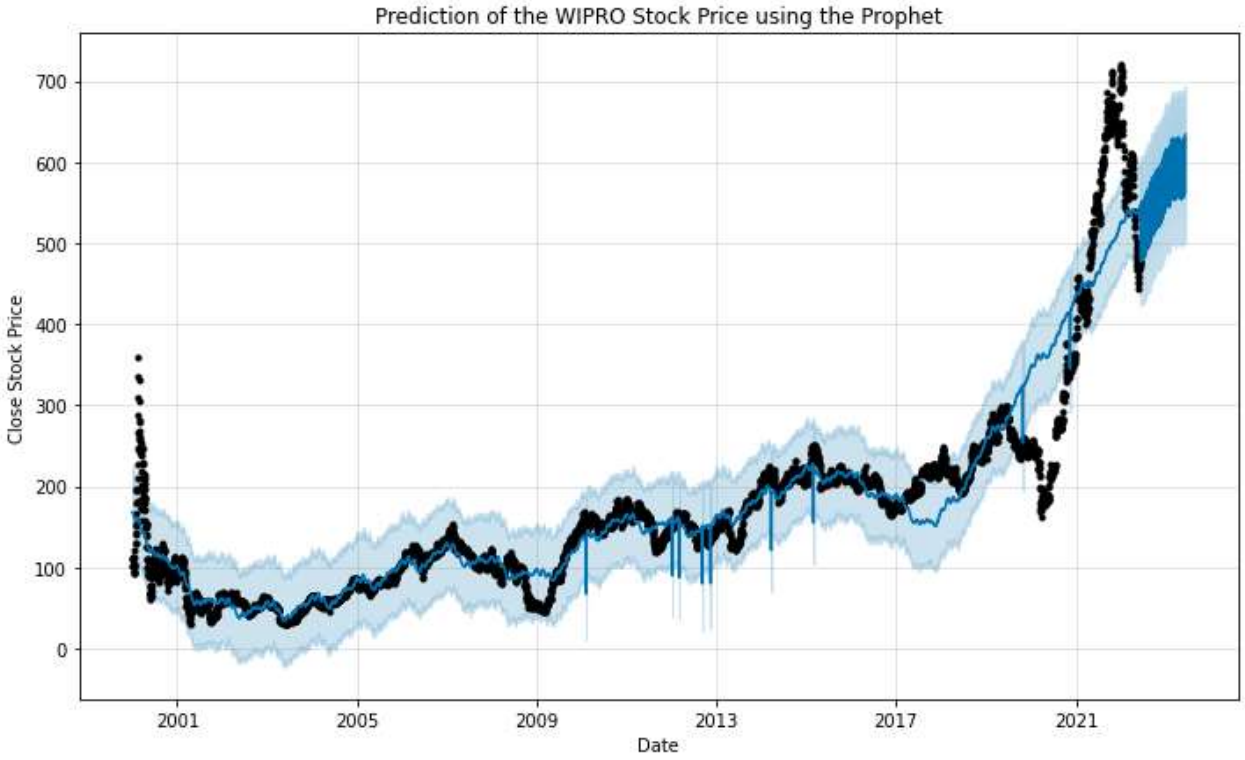


Fig 13: WIPRO stock price

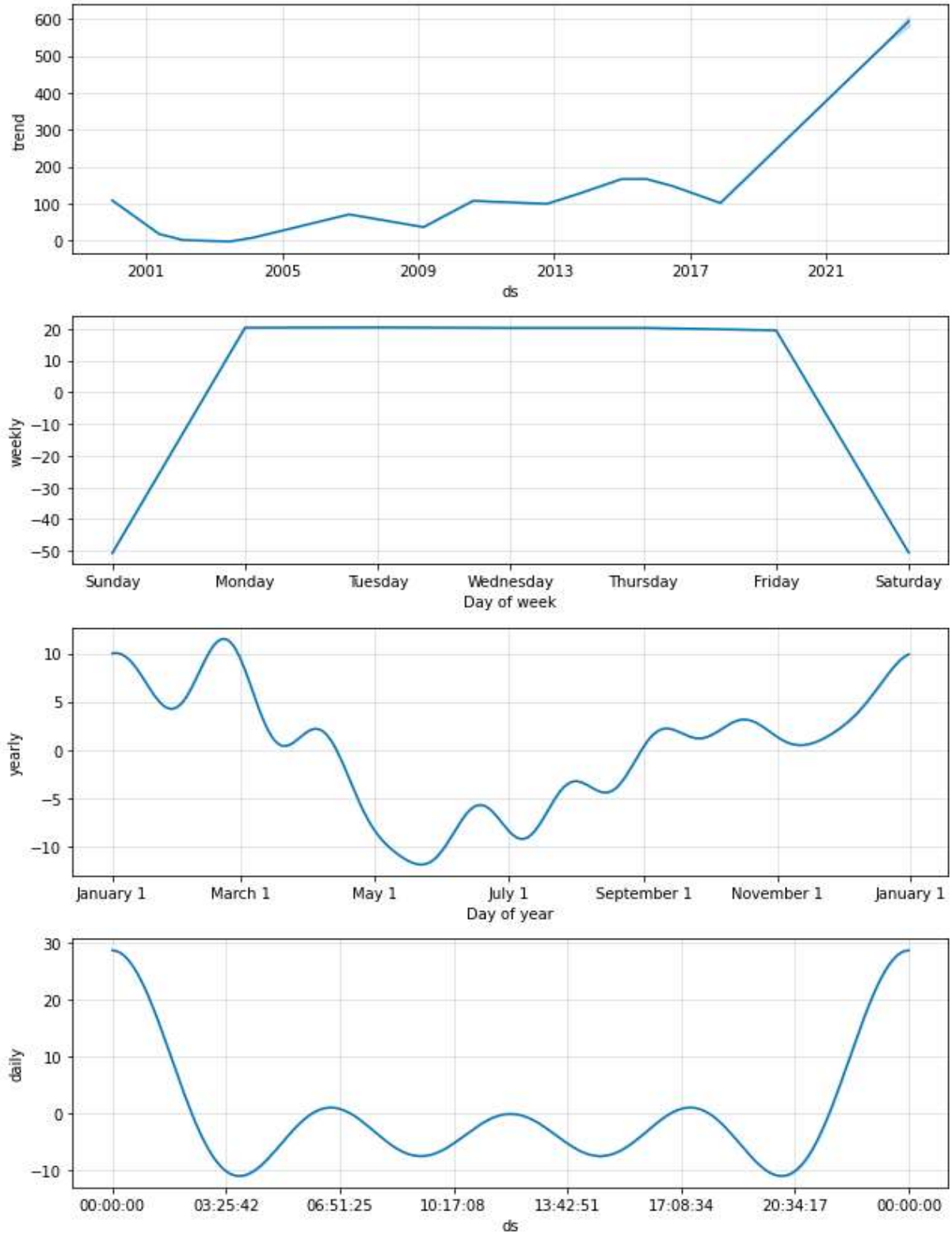


Fig 14: WIPRO stock Trend lines for daily, monthly, and yearly stock price

DISCUSSIONS

6.1 XG Boosting

Speed and performance:

It is comparatively quicker than previous ensemble classifiers and was originally designed in C++.

A core algorithm is parallelizable:

The fundamental XG Boost algorithm may take advantage of multi-core processors' capabilities because it is parallelizable. Additionally, it can be parallelized onto GPUs and across computer networks, making it possible to train on very huge datasets.

Consistently outperforms other algorithm methods:

On numerous machine learning benchmark datasets, it has demonstrated superior performance.

Wide variety of tuning parameters:

Cross-validation, regularization, user-defined goal functions, missing values, tree parameters, Scikit-Learn compatible API, and other parameters are all intrinsic to XG Boost.

6.2 Facebook Prophet

1. Facebook Prophet can produce results in a matter of seconds and is as accurate as a seasoned analyst.
2. Facebook Prophet can handle several outliers and null values with a minimum amount of data processing.
3. Users can manually input numbers for seasonality and holidays, which can make it easier to integrate the specific subject knowledge.

CONCLUSIONS

In this study, we create a model for predicting close stock prices utilizing XG Boosting and Fb Prophet algorithms for prediction. We are predicting the closing stock price of firms. With these two models, we applied datasets from Infosys, TCS, and WIPRO and attained accuracy levels of over 90%.

FUTURE WORK

1. To extend these models for predicting cryptocurrency trading.
2. To add sentiment analysis for better prediction.

REFERENCES

[1] Mao Y-y, Zhang Q-y. Research on Stock Prediction Based on Machine Learning[J]. Modern Computer, 2020(23):45-48.

[2] Zhang Q-q, Lin T-h, Qi X-y, et al. A review of stock forecasting based on machine learning[J]. Journal of the HeBei Academy of Sciences. 2020(37):15-21.

[3] Diakoulakis I E, Koulouriotis D E, Emiris D M. A Review of Stock Market Prediction Using Computational Methods[M]. Springer US, 2002.

[4] Black F. Studies of stock market volatility changes[J]. Proceedings of the American Statistical Association Business& Economic Statistics.

[5] Cui X, Shang W, Jiang F, et al. Stock Index Forecasting by Hidden Markov Models with Trends Recognition[C]// 2019 IEEE International Conference on Big Data (Big Data). IEEE.

[6] Huang T-y, Chen F-f. Application of Kernel Function of Stock Price Forecasting Based on SVM[J]. Journal of ChongQing University of Technology (Nature Science). 2016(30):89-94.

[7] Zhang L, Xu Q Y, Chen C, et al. Forecasting Stock Price Based on Back Propagation Neural Network[J]. Advanced Materials Research, 2014, 1006-1007:1031-1034.

[8] W Klösgen, Zytkow J M. Handbook of data mining and knowledge discovery[J]. Kybernetes, 2002, 33(7):809-835.