DOI: 10.28934/ea.2094 First Online: November 10, 2026

ORIGINAL SCIENTIFIC PAPER

Forecasting the Global Aviation Industry Index with Machine Learning Methods

Hadj Moussa Mansouri[1*] | Imane Bouguerra[2]

[1] University of Tamanghasset, Faculty of Economic, Commercial and Management Sciences, Department of Management Sciences, Tamanghasset, Algeria
[2] University of Ghardaia, Faculty of Economic, Commercial and Management Sciences, Department of Department of Economics, Ghardaia, Algeria

ABSTRACT
Forecasting the sectoral financial index has become increasingly important in the post-COVID-19 era, particularly in highly volatile industries such as aviation. Despite the widespread use of machine learning techniques in financial forecasting, empirical studies on their comparative performance in aviation-focused exchange-traded funds during market recovery phases remain limited. This study evaluates the predictive performance of three machine learning models: Random Forest, Facebook Prophet, and XGBoost in forecasting the U.S. Global Jets ETF (JETS). Daily closing price data from January 4, 2021, to January 10, 2025, were used. The models were evaluated using mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The empirical results show that XGBoost achieves the highest predictive accuracy, reducing RMSE by approximately 21.8% compared to Random Forest and by 64.5% compared to Facebook Prophet. While the Facebook Prophet algorithm shows weaknesses in capturing nonlinear market dynamics during structurally sensitive periods, the Random Forest method performs consistently and competitively. These results offer sector-specific evidence of the relative efficacy of machine learning techniques in aircraft market forecasting. They add to the expanding corpus of research on the use of machine learning in financial time series forecasting during periods of economically sensitive recovery.

Keywords: *random forest, Facebook Prophet, XGBoost, forecasting, global aviation industry index*

JEL Classification: C58, C53

INTRODUCTION

In recent years, the aviation industry has encompassed all aspects of air travel, including airlines, aircraft manufacturing, and military aviation. As the industry grows rapidly, it faces challenges in meeting passenger growth, service demand quality, and carbon emissions (Demir & Paksoy, 2024). In addition, the rapid expansion of international aviation has raised environmental concerns. Data-driven analysis is now more accurate and efficient due to developments in big data technology and machine learning techniques, which also promise future research prospects and trends (Gao & Mavris, 2022).Tan, Yan, and Zhu (2019) argue that this phenomenon can be better understood by using machine learning to forecast stock dynamics in emerging countries, especially China. In this case, stocks are categorised according to both short-term and long-term excess returns using the random forest algorithm. By bridging the gap between intricate pricing

* Corresponding author, e-mail: ansouri.hm@univ-tam.dz

mechanisms and portfolio management, this method assists financial practitioners in improving the efficiency of stock selection. According to Meher et al.(2024), accurate share value prediction is crucial in the rapidly evolving market because it enables investors to minimise risk, increase profits, and make well-informed decisions in a highly unstable setting. Demir and Paksoy (2024) explain that stock markets are among the most important means of capital accumulation, attracting individuals interested in doing business and earning money. However, AI systems can assist investors in identifying the best companies to invest in based on factors such as stock price, opening and closing prices, and total stock value. According to Madhuri et al. (2020), one significant aspect that has been identified is that stock market data are time-series in nature and non-stationary, which makes future stock prices challenging to forecast. The Facebook Prophet algorithm is a tool used to analyze time-series data based on daily observations and can be very useful for helping investors determine when it is better to buy or sell stocks. Researchers employ data mining techniques to predict stock prices by constructing models that are then trained on real-time data for assessment purposes. In addition, Zheng et al.(2024) assessed the predictive performance of random forest models and AI algorithms on four stocks to help investors make better decisions by evaluating the accuracy and efficiency of the models. Huang (2023)conducted a study using Google stock price data from January 2, 2018, to March 31, 2023, to predict stock prices using support vector regression, random forest, and K-nearest neighbour classification algorithms. The random forest model predicted an increasing trend for the next ten days.

This study used powerful machine learning algorithms, Random Forest, Facebook Prophet and XGBoost, to predict the global aviation industry index. The prediction problem is discussed in the first section. This emphasizes the importance of algorithms in the prediction process and the use of machine learning algorithms in financial markets. The second section discusses the literature on previous studies. The third section presents the data and methodology of using these algorithms. The fourth section is the results and discussion. The final section discusses the prediction results of each model. The purpose of this study is to illustrate the importance of using these models to predict the market and understand the trend of financial markets.

LITERATURE REVIEW

Much research has examined various approaches to projecting stock market indices in countries' economies. Index forecasting draws more attention from investors because they realize its importance. The literature increasingly includes studies that focus on modelling stock market indicators using Random Forest, XGBoost, and Facebook Prophet.

In their study, Purnomo et al. (2024) compared the accuracy of BiLSTM-based prediction and Random Forest models for predicting Antam's stock price. The results showed that BiLSTM had a lower error rate, while Random Forest performed better in terms of training time and computational efficiency. Zhao (2022) conducted a study comparing ARIMA and Random Forest for stock market forecasting. The results indicated that ARIMA was a useful tool for short-term forecasting, but could only capture linear relationships. Although more accurate, Random Forest was less user-friendly and required substantial computational resources to generate predictions. Comparable findings were reported by Diane and Brijlal (2024), who used Random Forest and Artificial Neural Network (ANN) non-linear machine learning algorithms to predict stock market volatility over a 5- year period. The Random Forest model outperformed the ANN model for both the JFIN and JBIND indexes. Their research also examined the influence of COVID-19, which revealed ambiguous detrimental effects on the models' performance.

Wu (2022) investigated three stock prediction models based on decision tree algorithms: Decision Tree, Random Forest, and XGBoost. The Random Forest model had the highest accuracy and the lowest Mean Squared Error (MSE), while XGBoost had the lowest accuracy. Adetunji et al. (2021)give supporting evidence by using Random Forest machine learning to predict house

prices, overcoming the limitations of the House Price Index. The model assessed using the Boston housing dataset generated reasonably accurate predictions with an error margin of ±5.

Tao (2023) investigated the predictive accuracy of multiple linear regression, LSTM, and Random Forest models regarding BMW's stock trends, concluding that the superior performance of the linear regression model warrants further exploration. In a similar vein, Meher et al. (2024) leveraged the Random Forest algorithm to forecast the stock prices of prominent Indian fintech entities like Niyogin Ltd. and Paytm. Their model exhibited a high degree of reliability, evidenced by a coefficient of determination greater than 95%.

Furthermore, Daori et al. (2022) evaluated the predictive performance of Random Forest models combined with artificial intelligence for four stocks, focusing on both accuracy and efficiency to assist investors in making informed stock decisions. Besides, Talazadeh and Peraković (2024) introduced a new approach to stock market prediction, the "Sentiment-Augmented Random Forest," which integrates the FinGPT generative AI model with the traditional Random Forest, enhancing accuracy by 9.23% and reducing prediction errors. Research conducted by Bin Omar et al. (2022) indicated that machine learning models can forecast stock index prices across three time periods: pre-COVID-19, post-COVID-19, and COVID-19. The results show that AR-DNN(1, 3,10) and AR-RF(1) are the best models for high stock price fluctuations, useful for investors and policy makers.

However, Jha and Pande (2021) proposed a Facebook Prophet (FB) tool for predicting supermarket sales, comparing it to other forecasting models, such as additive models and ARIMA models. The tool demonstrates improved accuracy and fit. Similarly, predictive models for sales, finance, and stock market forecasting are equally crucial for profit maximization. The efficiencies of models were compared and proved to have a low error rate and better accuracy using Facebook's Prophet Model than using traditional models(Sharma, Bhalla, & Ganesan, 2022). Li et al. (2023)elaborated on how advanced Artificial Neural Network and Facebook Prophet models were used to forecast lithium prices in China. The result demonstrates the capability of the Facebook Prophet model to produce accurate predictions, capture short-term patterns in lithium prices, and detect unexpected market changes.

Güleryüz and Özden (2020)compared long-term memory (LSTM) and Facebook Prophet models for forecasting Brent oil prices. The research found that LSTM performed better than Facebook Prophet. In a similar context, Kulshreshtha and Vijayalakshmi (2020) designed innovative models for stock market prediction, including LSTM, RNN, and ARIMA. The models designed to analyze S&P 500 data have a better fit than the Facebook Prophet forecasting library, with a 99% fit rate and lower RMSE. Papacharalampous and Tyralis (2018) used historical and precipitation data to compare Random Forests and Facebook Prophet for forecasting US river stream flow up to seven days ahead. The results indicate that Random Forests provided better overall forecasting performance and captured unexpected fluctuations better than the other methods.

Other models emphasized Malaysia's agriculture industry and its significance to farmers. A model based on ARIMA, SVR, XGBoost, and Facebook Prophet was developed to increase agricultural profitability. Among the algorithms, LSTM was found to be the most efficient and capable of dealing with increasing data and complexity. Future enhancements include investigating detailed relationships between supporting elements and historical prices, as well as undertaking location-specific price assessments to assist farmer trading (Yuan & Ling, 2020). Rahman et al. (2022) conducted a study on the prediction of COVID-19 cases and deaths in Bangladesh with the help of ARIMA and XGBoost models. The results indicated that the relationship driven by nonlinearity and seasonality curves, both of which are symmetric, could lead to potentially useful findings. For both cases and deaths, ARIMA scored higher than XGBoost. The same forecasting method can be used to inform public health plans as well.

On the contrary, different XGBoost models exhibited a remarkable fit and forecast accuracy for COVID-19 cases in the USA when compared to the seasonal ARIMA models, using a 14-day out-of-

sample (Fang, Yang, Lv, An, & Wu, 2022). In addition, the XGBoost algorithm, an ensemble model of a multitude of trees, was able to surpass the current models, which were D-GEX, linear regression, and the KNN methods, when predicting gene expression values on the GEO and RNA-seq datasets (Li, Yin, Quan, & Zhang, 2019). Correspondingly, the research also focused on the analysis of the seasonal ARIMA and XGBoost models for predicting human brucellosis in main land China. It was established that the XGBoost model outperformed the ARIMA due to its speed and feature extraction techniques, which are quite advanced (Alim, Ye, Guan, Huang, Zhou, & Wu, 2020).

Guerra et al.(2022) employed data from the Portuguese banking sector to conduct their experiment. The purpose of their experiment was to assess whether machine-learning models have any application in liquidity risk modelling. The evidence demonstrates that applying the Extreme Gradient Boosting (XGBOOST) technique is the most effective compared to other techniques for detecting timely states of stress testing and developing an early warning system, thus strengthening the Supervisory Review and Evaluation Process (SREP). In addition, Petropoulos and Siakoulis (2021) employed the prediction component of XGBoost and combined it with three machine learning models: Deep Neural Networks (DNN), SVM, and Random Forest. They then compared the performance of these models. They reported that the system using XGBoost outperformed the system using the DNN technique in terms of accuracy. The technique can be applied by central banks to monitor stability in the global financial system.

Furthermore, XGBoost performed better than Random Forest and Decision Trees in categorization and financial performance, indicating that it can effectively respond to unpredictable market conditions. The algorithms were found to have the ability to predict profitable stocks, but there is a need for more analysis using explainable tools (Caparrini, Arroyo, & Mansilla, 2024). The study also provided a data-driven system for fault detection in wind turbines using Random Forests and Extreme Gradient Boosting. The results demonstrated the robustness of the models and their ability to prevent overfitting, with better performance in fault detection than the Support Vector Machine (SVM) technique (ZHANG, QIAN, MAO, HUANG, HUANG, & S, 2018).

Al-Maadid et al. (2022) utilized machine learning models to assess the impact of COVID-19 news on the GCC stock market. The evidence suggests such news impacted stock markets in the UAE, Qatar, Saudi Arabia, and Oman to a larger degree compared to Bahrain. The evidence also found that the XGBoost model is better than DT in precision. These insights re-emphasize portfolio diversification and the application of risk management strategies, given the amount of global news during such crises. Similarly, Mohapatra et al. (2022) utilized ensemble machine-learning models, including XGBoost, Gradient Boosting, AdaBoost, and Random Forest, to predict Indian bank stock returns through technical indicators. The findings indicate that the XGBoost algorithm outperformed the other models, thus providing insights to traders, investors, and portfolio managers into market movements and return forecasting. Han (2023)compared machine learning models in forecasting China's stock market movement by E-Hualu using five models: Linear Regression, SVR, Random Forest, XGBoost, and LSTM. The evidence established LSTM to have performed better than all models in forecasting price movement and market patterns.

A review of the literature presents various studies on using machine learning for forecasting, particularly those employing models such as Random Forest, Facebook Prophet, and XGBoost, which have always yielded better forecasts compared to models including ARIMA and multiple regression. The machine learning models have also outperformed K-Nearest Neighbors (KNN) and Artificial Neural Network (ANN). Although numerous studies have examined stock price forecasting for various companies, none of the previous studies focused on JETS stock price, which represents an important segment of the aviation industry. Closing such a void in studies is crucial for shareholders, investors, and analysts who base their decisions on accurate and reliable forecasts.

DATA AND METHODOLOGY

Research Objective

The main goal of this study is to assess and contrast the prediction capabilities of three machine learning algorithms: XGBoost, Facebook Prophet, and Random Forest (RF) in predicting the U.S. Global Jets ETF (JETS, Ticker: JETS). The study aims to determine which modeling paradigm provides superior out-of-sample forecasting accuracy in a volatile, post-pandemic aviation market.

Although financial price series may exhibit non-stationary behavior, this study focuses on forecasting the observed price levels because the primary objective is to provide direct estimates of future stock prices rather than returns. The selected machine-learning and forecasting approaches can model nonlinear relationships and evolving temporal patterns without requiring the strict stationarity assumptions associated with some conventional time-series models. Forecasting price levels therefore allows the predictive performance of the models to be evaluated directly on the original scale of the stock-price series using MSE, RMSE, and MAPE.

Data Collection

Daily closing prices for JETS from 4 January 2021 to 10 January 2025 were used. The dataset consists of 1,016 trading-day observations and was obtained from the https://sa.investing.com/financial database. This period captures the post-COVID-19 recovery phase, characterized by structural market adjustments and elevated volatility.

Only daily closing prices were employed as the target variable; no additional explanatory variables or technical indicators were included, ensuring a focused univariate forecasting analysis. The price series under study contained no missing values, ensuring data completeness and integrity.

The chronological sequence of the time series was rigorously maintained to prevent look-ahead bias. Rather than using random sampling, the dataset was split into 80% training data and 20% testing data, depending on the time sequence. The testing set served as a genuine out-of-sample evaluation period, while the training set was utilised for model estimates and hyperparameter tuning.

Regarding the forecasting horizon, the tree-based models (Random Forest and XGBoost) were configured for one-step-ahead prediction, using lagged observations to estimate the value for the subsequent period, whereas the Prophet model was used to generate forecasts for 120 future periods. This configuration reflects the forecasting procedures implemented for each model and provides a clear basis for interpreting and comparing their predictive performance.

Model Selection

Three machine learning models were selected to represent different algorithmic paradigms:

Random Forest Model

The Random Forest (RF) is a machine learning model that uses bagging and decision trees to deal with multiple, multifaceted input parameters. The approach can be used for both classification and regression tasks, in addition to providing information on variable importance (Zhao, 2022). As an ensemble approach, Random Forest utilizes decision trees to produce forecasts, avoid overfitting, and maintain stability in models by alternating between sets of independent variables in individual trees (Diane & Brijlal, 2024). The nonlinear approach creates a "forest" of trees, predicting through random sampling and voting to minimize the chances of overfitting (Zheng, Xin, Cheng, Tian, & Yang, 2024). In addition, a classification tree is predicting

the class of an input vector, where each tree in the forest chooses variables by chance in every node. The interaction between trees is vital in determining the rate of error for the model (Mao, 2023).

Facebook Prophet Model

The Facebook Prophet platform, a time-series platform developed by Core Data Science for Facebook, is a platform for producing high-quality sales forecasts accounting for seasonality, outliers, and seasonally influenced outcomes (MADHURI, CHINTA, & KUMAR, 2020). Facebook Prophet is an open-source package developed by data scientists at Meta Corporation, which deals with seasonality in time-series data and is composed of three underlying factors, which include seasonality, growth rate, and error term(Aliu, Hašková, & Šuleř, 2022).

The Facebook Prophet model is made up of three major components, such as seasonality, holidays, and trend, to predict y(t) over time:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon t \quad (1)$$

Where εt is an error term that should be normally distributed, g(t)is a trend, s(t) is a seasonality term, and h(t)denotes the holiday series (Güleryüz & Özden, 2020).

The methodology consists of data collection, data shaping and pre-processing, holiday initialization, forecasting, and evaluation to ensure flexibility and utility (Jabed, 2024).

XGBoost Model

Boosting is a monitored machine learning method run by repetition, with two defined properties. First, it combines several weak models, each with limited predictive power, into a single, more accurate model. The process begins with a basic model that originally provides weak predictions. The difference between these predictions and real values is calculated. To improve accuracy, gradient boosting is used and the model adapts to errors by taking advantage of negative gradients(Goldberg & Mouti, 2022). Among its implementations, XGBoost stands out as an optimized gradient-boosting algorithm for decision trees, effectively handling regression, classification, and ranking problems. It manages missing values efficiently, minimizes overfitting, and reduces bias, making it particularly useful for applications such as forecasting and bond pricing, where data inaccuracies and completeness are critical factors (Katterbauer, Syed, Cleenewerck, & Genc, 2022). As a robust and efficient gradient-boosting machine algorithm, XGBoost is guided by an objective function:

$$Obj^{(t)} = \sum_{i=1}^{n} \boldsymbol{l}\left(\boldsymbol{y_i}, \hat{\boldsymbol{y}}_{\boldsymbol{i}}^{(\boldsymbol{t-1})} + \boldsymbol{f_t}(\boldsymbol{x_i})\right) + \boldsymbol{\Omega}(\boldsymbol{f_t}) + \boldsymbol{constant} \quad (2)$$

where y_i represents the observed value, $\hat{y}_i^{(t-1)}$ denotes the predicted value from the previous iteration, and x_i refers to the feature vector, while n indicates the sample size. The function f_t corresponds to the new function learned by the model at iteration t, and $\Omega(f_t)$ is a regularization term introduced to control model complexity. The loss function L measures the discrepancy between the actual label and the prediction obtained in the preceding step, reflecting the contribution of the newly added tree (Rahman, Chowdhury, & Amrin, 2022).

This selection allows a comparative evaluation between ensemble tree-based methods and additive time-series decomposition approaches.

HYPERPARAMETER OPTIMIZATION

To ensure a fair comparison, hyperparameters for Random Forest and XGBoost were optimized using Grid Search with cross-validation, conducted exclusively on the training dataset while respecting time-series validation principles. This procedure systematically explored predefined parameter combinations to mitigate the risks of underfitting or overfitting and to identify the configuration that minimizes forecasting error. Facebook Prophet parameters were adjusted to control trend flexibility and seasonal components.

Metrics

Three widely used forecasting metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), are used to assess our model's predictive power, as follows (Luo, Zhang, & Yao Fu, 2021):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

$$MAPE_{\%} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (5)$$

where $y_i - \hat{y}_i$ represents the error between the predicted and actual values, and n is the number of observations. MSE is a loss function used to quantify the discrepancy between the true and predicted values. RMSE is the square root of the mean squared error and expresses the prediction error in the same units as the target variable. MAPE measures the average absolute percentage error between the actual and predicted values. It expresses forecasting errors as percentages, making the results easier to interpret. Lower MAPE values indicate better forecasting accuracy.

RESULTS AND DISCUSSION

Model accuracy is measured using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), which ensure a thorough and comparable evaluation framework. This section shows the out-of-sample forecasting performance of the three models, Random Forest, XGBoost, and Facebook Prophet, evaluated on the test dataset.

Forecasting performance was evaluated over the entire out-of-sample test period rather than selected observations, ensuring statistically representative and unbiased results.

Comparative Forecasting Performance

Table 1 reports the out-of-sample performance metrics for all models.

Table 1. Comparative Out-of-Sample Forecasting Performance

Model	MSE	RMSE	MAPE (%)
Random Forest	0.2254	0.4748	1.66
XGBoost	0.1380	0.3715	1.31
Facebook Prophet	1.0927	1.0453	4.19

Source: Authors' calculation based on Python output

As shown in Table 1, XGBoost achieves the lowest error values across all the three metrics of evaluation, indicating superior predictive performance. Random Forest demonstrates competitive accuracy, although with slightly higher error levels compared to XGBoost. Facebook Prophet records substantially higher MSE, RMSE, and MAPE values, suggesting limitations in capturing complex nonlinear market dynamics during the post-pandemic recovery phase.

The results confirm that ensemble tree-based learning methods outperform additive time-series modeling in forecasting JETS during structurally volatile periods.

Random Forest Results

The Random Forest model was optimized using Grid Search on the training dataset while preserving chronological order.

Table 2. Random Forest Optimal Hyperparameters

Parameter	Value
n_estimators	100
max_depth	20
min_samples_split	2
min_samples_leaf	1
max_features	sqrt

Source: Authors' calculation based on Python output

The model demonstrates stable predictive behavior, with relatively low MAPE (1.66%). Its ability to model nonlinear interactions contributes to improved short-term forecasting accuracy compared to Facebook Prophet.

XGBoost Results

XGBoost (Extreme Gradient Boosting) is an ensemble learning algorithm based on gradient-boosted decision trees. It builds the model sequentially by adding decision trees, where each new tree is trained to reduce the errors of the preceding ensemble. The algorithm optimizes a differentiable objective function that combines a loss function with regularization terms to control model complexity and reduce overfitting. In this study, XGBoost was implemented as a regression model (XGBRegressor) to predict the subsequent value of the stock-price series. The model parameters were optimized using GridSearchCV with a predefined parameter grid, and the best-performing parameter combination was selected based on the negative mean squared error criterion.

Table 3. XGBoost Optimal Hyperparameters

Parameter	Value
learning_rate	0.2
max_depth	3
n_estimators	300
Subsample	0.8
colsample_bytree	0.8

Source: Authors' calculation based on Python output

XGBoost achieves the strongest predictive performance, with RMSE = 0.3715 and MAPE = 1.31%. The sequential error correction mechanism of boosting enhances its ability to capture complex market adjustments and short-term fluctuations, explaining its superior accuracy.

FacebookProphet Results

In this study, the Prophet model was implemented using its standard configuration to forecast the closing-price series. The model was fitted to the historical observations, with the date variable specified as the time index (ds) and the closing price as the target variable (y). A forecasting horizon of 120 future periods was specified using the make_future_dataframe (periods=120) function. The model automatically estimates the underlying trend and potential seasonal patterns from the historical data, while its changepoint mechanism allows the trend to adapt to changes over time. Prophet was selected because its time-series framework can capture evolving trends and recurring temporal patterns without requiring the strict stationarity assumptions typically imposed by some conventional time-series models. These characteristics make it suitable for examining financial price series that may exhibit changing trends and temporal patterns.

Table 4. Facebook Prophet Configuration Parameters

Component	Specification
Trend	Automatic
Weekly Seasonality	Automatic
Yearly Seasonality	Automatic
Changepoint Prior Scale	0.05

Source: Authors' calculation based on Python output

Figure 1 illustrates actual versus predicted values during the test period.

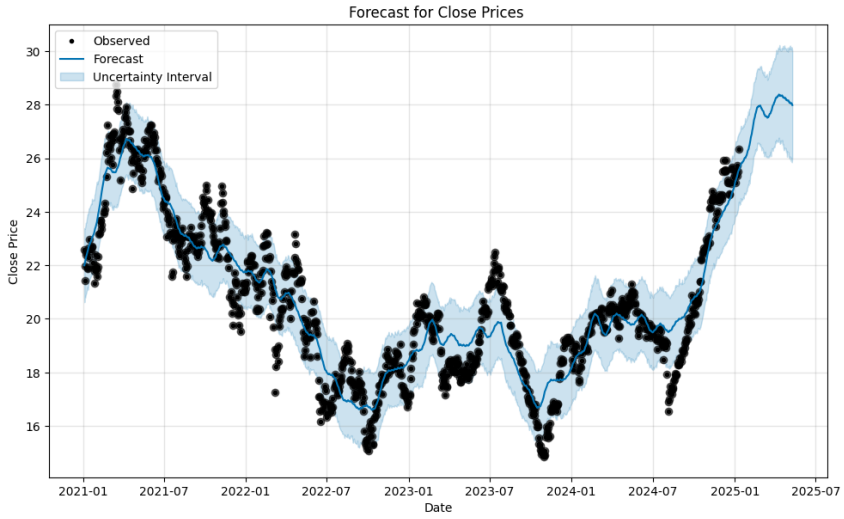


Figure 1. Actual vs Predicted JETS Prices (FacebookProphet)

Source: Authors' calculation based on Python output

Figure 1 illustrates the comparison between actual and forecasted values, together with uncertainty intervals. The widening confidence bands over time reflect increasing forecast uncertainty, while the upward trajectory indicates recovery dynamics in the aviation sector.

Figure 2 presents the decomposed trend and seasonal components.

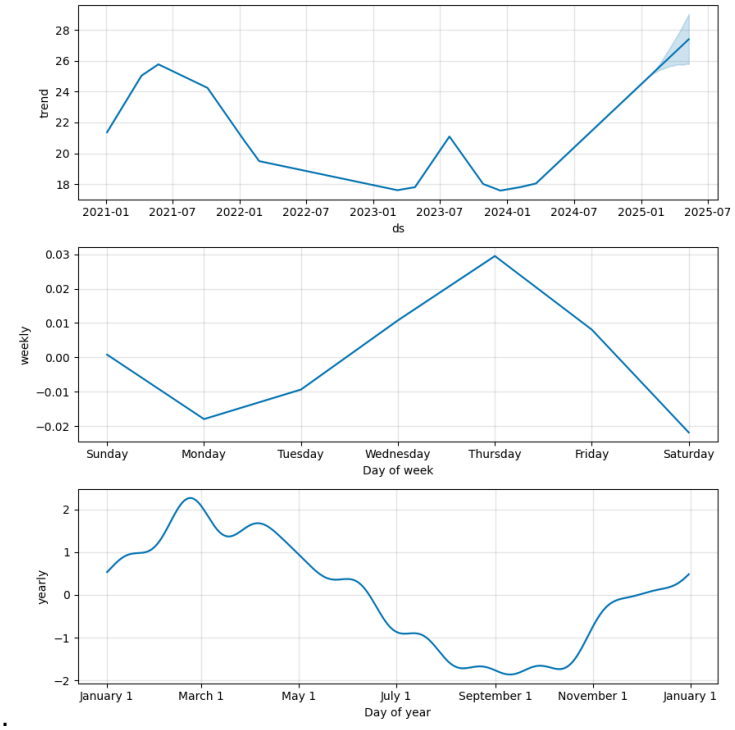


Figure 2. Trend and Seasonal Decomposition

Source: Authors' calculation based on Python output

The trend component shows a decline from 2021 to mid-2023 followed by gradual recovery, reflecting sectoral stabilization. Weekly seasonality reveals recurring trading patterns, while yearly seasonality indicates cyclical fluctuations potentially linked to travel demand dynamics.

Despite capturing structural temporal patterns, Facebook Prophet's forecasting accuracy remains inferior to tree-based ensemble methods.

CONCLUSION

This study examined the forecasting performance of three machine learning models, Random Forest, Facebook Prophet, and XGBoost, in predicting the U.S. Global Jets ETF (JETS). The comparative out-of-sample evaluation demonstrated clear differences in predictive accuracy, with XGBoost achieving the lowest error metrics across MSE, RMSE, and MAPE. The results confirm the effectiveness of gradient-boosting frameworks in modeling nonlinear and structurally complex financial time series, particularly within volatile, sector-specific market environments.

Beyond identifying the most accurate model, this research contributes to the literature by providing a structured comparison between ensemble tree-based learning and additive time-series decomposition in an aviation-focused financial context. The findings suggest that boosting-based architectures are better suited to capturing dynamic market adjustments during post-pandemic recovery phases. For financial analysts and portfolio managers, the results offer practical guidance in selecting forecasting tools capable of handling nonlinear sectoral volatility.

Nevertheless, this study is subject to several limitations. The analysis focuses on a single sector-specific ETF, employs a univariate framework based solely on closing prices, and covers a defined post-pandemic period. Future research may extend this framework by incorporating additional financial indicators, macroeconomic variables, alternative data sources, multi-step forecasting strategies, or hybrid modeling architectures that combine predictive power with interpretability.

REFERENCES

Adetunji, A. B., Akande, O. N., Ajala, F. A., Ololade, Akande, Y. F., & Oluwadara, G. (2021). House Price Prediction using Random Forest Machine Learning Technique. The 8th International Conference on Information Technology and Quantitative Management. Procedia Computer Science.https://doi.org/10.1016/j.procs.2022.01.100

Alim, M., Ye, G.-H., Guan, P., Huang, D.-S., Zhou, B.-S., & Wu, W. (2020). Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time series study. *BMJ Open* .https://doi.org/10.1136/bmjopen-2020-039676

Aliu, F., Hašková, S., & Šuleř, P. (2022). SUSTAINABILITY OF ELECTRICITY PRICES AND THE CONSEQUENCES FOR THE PRAGUE STOCK EXCHANGE. *ENTREPRENEURSHIP AND SUSTAINABILITY ISSUES*10(2): 473-494.http://doi.org/10.9770/jesi.2022.10.2(30)

Al-Maadid, A., Alhazbi, S., & Al-Thelaya, K. (2022). Using machine learning to analyze the impact of coronavirus pandemic news on the stock markets in GCC countries. *Research in International Business and Finance* .https://doi.org/10.1016/j.ribaf.2022.101667

Bharat Kumar Meher; Abhishek Anand; Sunil Kumar; Ramona Birau; Manohar Singh. (2024). Effectiveness of Random Forest Model in Predicting Stock Prices of Solar Energy Companies in India. *International Journal of Energy Economics andPolicy* , 14(2): 426-434.doi:10.32479/ijeep.15581

Caparrini, A., Arroyo, J., & Mansilla, J. E. (2024). S&P 500 stock selection using machine learning classifiers: A look into the changing role of factors. *Research in International Business and Finance* .https://doi.org/10.1016/j.ribaf.2024.102336

DAORI, H., ALHARTHI, M., ALANAZI, A., ALZAHRANI, G., ABOROKBAH, M., & Aljehane, N. (2022). Predicting Stock Prices Using the Random Forest Classifier. PREPRINT (Version 1) available at Research Square, https://doi.org/10.21203/rs.3.rs-2266733/v1

Demir, S., & Paksoy, T. (2024). A Conceptual Framework for Smart and Sustainable Operations in the Aviation Industry. *Smart and Sustainable Operations Management in the Aviation Industry*. CRC Press.

Diane, L., & Brijlal, P. (2024). Forecasting Stock Market Realized Volatility using Random Forest and Artificial Neural Network in South Africa. *International Journal of Economics and Financial Issues* , 14:2,5-14. DOI:10.32479/ijefi.15431

Fang, Z.-g., Yang, S.-q., Lv, C.-x., An, S.-y., & Wu, W. (2022). Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study. BMJ Open .https://doi.org/10.1136/bmjopen-2021-056685

Gao, Z., & Mavris, D. N. (2022). Statistics and Machine Learning in Aviation Environmental Impact Analysis: A Survey of Recent Progress. *Aerospac*e, 9:12, 2-35. https://doi.org/10.3390/aerospace9120750

Goldberg, L. R., & Mouti, S. (2022). Sustainable Investing and the Cross-Section of Returns and Maximum Drawdown. *The Journal of Finance and Data Science*.8, 353-387. https://doi.org/10.1016/j.jfds.2022.11.002

Guerra, P., Castelli, M., & Côrte-Real, N. (2022). Machine learning for liquidity risk modelling: A supervisory perspective. Economic Analysis and Policy, 175-187.https://doi.org/10.1016/j.eap.2022.02.001

Güleryüz, D., & Özden, E. (2020). The Prediction of Brent Crude Oil Trend Using LSTM and Facebook Prophet. *European Journal of Science and Technology.* https://doi.org/10.31590/ejosat.759302

Han, L. (2023). Analysis of Stock Price and Price Movement Prediction based on Machine Learning Models for E-Hualu. BCP Business & Management.

Huang, Y. (2023). Research on the Google Stock Price Prediction Based on SVR, Random Forest, and KNN Models. Highlights in Business, Economics and Management. 24, 1054-1058

Jabed, M. A.& Murad, M. A. A. (2024). STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES.*Heliyon*, 10:24, https://doi.org/10.1016/j.heliyon.2024.e40836

Jha, B. K., & Pande, S. (2021). Time Series Forecasting Model for Supermarket Sales using FB-Prophet. Proceedings of the Fifth International Conference on Computing Methodologies and Communication. Carleton University. DOI: 10.1109/ICCMC51019.2021.9418033

Katterbauer, K., Syed, H., Cleenewerck, L., & Genc, S. Y. (2022). Robo-Sukuk pricing for Chinese equities. *Borsa _Istanbul Review*. 22:5, 854-860. https://doi.org/10.1016/j.bir.2022.06.002

Kulshreshtha, S., & Vijayalakshmi, A. (2020). An ARIMA- LSTM Hybrid Model for Stock Market Prediction Using Live Data. *Journal of Engineering Science and Technology Review.*DOI: 10.25103/jestr.134.11

Li, W., Yin, Y., Quan, X., & Zhang, H. (2019). Gene Expression Value Prediction Based on XGBoost Algorithm. *Front. Genet.* https://doi.org/10.3389/fgene.2019.01077

Li, X., Sengupta, T., Mohammed, K. S., & Jamaani, F. (2023). Forecasting the lithium mineral resources prices in China: Evidence with Facebook Prophet (Fb-P) and Artificial Neural Networks (ANN) methods. *Resources Policy* .82, https://doi.org/10.1016/j.resourpol.2023.103580

Luo, J., Zhang, Z., & Yao Fu, F. R. (2021). Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results in Physics*. 27, https://doi.org/10.1016/j.rinp.2021.104462

MADHURI, C., CHINTA, M., & KUMAR, V. V. (2020, October 3). Stock Market Prediction for Time-series Forecasting. IEEE 7th International Conference on Smart Structures and Systems ICSSS . University of Canberra. DOI: 10.1109/ICSSS49621.2020.9202042

Mao, M. (2023). A Comparative Study of Random Forest Regression for Predicting House Prices Using. *Highlights in Science, Engineering and Technology*. 85, 969-974. https://doi.org/10.54097/bdfe8032

Mohapatra, S., Mukherjee, R., Roy, A., Sengupta, A., & Puniyani, A. (2022). Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators?,*Journal of Risk and Financial Management* .15:8, https://doi.org/10.3390/jrfm15080350

Omar, A. B., Huang, S., Salameh, A. A., Khurram, H., & Fareed, M. (2022). Stock Market Forecasting Using the Random Forest and Deep Neural Network Models Before and During the COVID-19 Period. Frontiers in Environmental Science.https://doi.org/10.3389/fenvs.2022.917047

Papacharalampous, G. A., & Tyralis, H. (2018). Evaluation of random forests and Prophet for daily streamflow forecasting. *European Geosciences Union*, 45, 201-208.https://doi.org/10.5194/adgeo-45-201-2018

Petropoulos, A., & Siakoulis, V. (2021). Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning technique. *Central Bank Review*, 21:4, 141-153.https://doi.org/10.1016/j.cbrev.2021.12.002

Purnomo, K., Wijaya, R. A., Fajar, M., & Suri, P. A. (2024). Comparative Analysis of BiLSTM Deep Learning Model and Random Forest Regressor Performance on Indonesian Nickel Mining Company Stock Prices. *Procedia Computer Science* , 778–786.doi:10.1016/j.procs.2024.10.304

Rahman, M. S., Chowdhury, A. H., & Amrin, M. (2022). Accuracy comparison of ARIMA and XGBoost forecasting models in predicting the incidence of COVID-19 in Bangladesh. PLOS *global public health* , 2:5, 1-13.https://doi.org/10.1371/journal.pgph.0000495

Sharma, K., Bhalla, R., & Ganesan, G. (2022). Time Series Forecasting Using FB-Prophet. *Algorithms Computing and Mathematics Conference*. India.

Talazadeh, S., & Peraković, D. (2024). SARF: Enhancing Stock Market Prediction with Sentiment-Augmented Random Forest. *arXiv preprint* . arXiv:2410.07143.

Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. Heliyon .5:8, https://doi.org/10.1016/j.heliyon.2019.e02310

Tao, S. (2023). Predicting BMW Stock Price Based on Linear Regression, LSTM, and Random Forest Regression. BCP Business & Management.38, 335-340, https://doi.org/10.54691/bcpbm.v38i.3712

Wu, Y. (2022). Stock Price Prediction Based on Simple Decision Tree Random Forest and XGBoost. BCP Business & Management. https://doi.org/10.54691/bcpbm.v38i.4311

Yuan, C. Z., & Ling, S. K. (2020). Long Short-Term Memory Model Based Agriculture Commodity Price Prediction Application. *Proceedings of the 2020 2nd international conference on information technology and computer communications*, https://doi.org/10.1145/3417473.3417481

ZHANG, D., QIAN, L., MAO, B., HUANG, C., HUANG, B., & S, Y. (2018). A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGBoost.*Front. Energy Res*. 9:751066. doi: 10.3389/fenrg.2021.751066

Zhang, Y. (2022). Stock price prediction method based on XGBoost algorithm. International Conference on Bigdata Blockchain and Economy Management. *Atlantis Press*. doi: 10.2991/978-94-6463-030-5_60

Zhao, Y. (2022). Comparison of Stock Price Prediction in Context of ARIMA and Random Forest Models. *BCP Business & Management*.https://doi.org/10.54691/bcpbm.v38i.3996

Zheng, J., Xin, D., Cheng, Q., Tian, M., & Yang, L. (2024). The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance. *arXiv preprint* . https://doi.org/10.48550/arXiv.2402.17194

Article history:	Received: 27.11.2025
	Revised: 3.11.2025
	Accepted: 9.11. 2025