

Developing a Predictive Model for S&P 500 Stock Price Forecasting Using the Prophet Algorithm

Vivian A. Kelechi Njoku, Bartholomew O. Eke, Martha O. Musa

Department of Computer Science, Faculty of Sciences, University of Port Harcourt, Nigeria

ABSTRACT

Stock market forecasting remains one of the most challenging problems in financial data science due to the inherent complexity, volatility, and non-linear behaviour of equity indices. This study presents a systematic investigation into the application of the Prophet algorithm, an open-source time-series forecasting framework developed by Meta (Facebook), for predicting the closing prices of the S&P 500 index. Using 3,623 historical daily records spanning January 2010 to December 2023, a structured data science pipeline was implemented encompassing pre-processing, exploratory analysis, model training, evaluation, and an interactive Streamlit-based deployment interface. The model was trained on 80% of the dataset and evaluated on the remaining 20%, achieving forecasting performance metrics including a Mean Absolute Error (MAE) of 68.53, Root Mean Square Error (RMSE) of 116.10, Mean Absolute Percentage Error (MAPE) of 2.18%, and an R^2 score of 0.9754. Directional accuracy was recorded at 52.7%, with trend classification yielding an F1 Score of 0.6963. Comparisons with baseline approaches, including the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, demonstrate the competitiveness of Prophet for capturing long-term structural trends in financial time series. The results confirm that Prophet offers a transparent, interpretable, and computationally accessible alternative for stock price forecasting, with practical implications for data-driven investment decision-making.

KEYWORDS

Prophet Algorithm, S&P 500, Stock Price Forecasting, Time-Series Analysis, Machine Learning, Financial Prediction, SARIMA, Streamlit

I. INTRODUCTION

The accurate prediction of stock market indices constitutes a fundamental objective within financial analytics and quantitative investment research. Equity indices such as the S&P 500, which aggregates the performance of 500 major publicly traded companies across the United States economy, serve as critical barometers for overall market health and investor sentiment. Forecasting the trajectory of such indices can support strategic asset allocation, portfolio risk management, and institutional decision-making processes. However, the inherent complexity of financial time-series data, characterised by non-stationarity, structural breaks, exogenous shocks, and regime shifts presents substantial modelling challenges.

Classical econometric methods such as the Autoregressive Integrated Moving Average (ARIMA) model and its seasonal extension SARIMA have been widely applied to financial time-series forecasting tasks. These models operate under strict statistical assumptions, including stationarity requirements and the absence of complex seasonal patterns, which may limit their applicability to dynamic equity data. In contrast, machine learning approaches such as Long Short-Term Memory (LSTM) networks offer greater representational flexibility, though they demand significant computational resources, large training datasets, and considerable expertise in hyper parameter optimisation.

The Prophet algorithm, developed and open-sourced by Meta's Core Data Science team, offers a compelling middle ground by combining the interpretability of decomposition-based forecasting with the robustness to handle missing data, irregular temporal patterns, and multiple seasonality's. Prophet models a time series as an additive combination of trend, seasonality, and holiday effects, providing a user-accessible framework with tenable hyper parameters that require minimal statistical expertise.

This study investigates the application of the Prophet algorithm for forecasting daily closing prices of the S&P 500 index over a 14-year historical period. A complete data science pipeline is implemented, incorporating pre-processing, exploratory descriptive statistics, model training, and multi-metric evaluation. An interactive Streamlit web application is developed to provide a user-friendly interface for dynamic parameter configuration and forecast visualisation. The performance of Prophet is benchmarked against the existing SARIMA-based framework proposed by Noor et al. (2022), contextualised within the broader landscape of recent stock forecasting literature.

II. LITERATURE REVIEW

The domain of time-series forecasting for financial markets has attracted sustained scholarly attention, encompassing classical statistical models, deep learning architectures, and hybrid approaches. The literature reviewed in this section establishes the theoretical and empirical foundations for the present investigation.

Noor et al. (2022) proposed a SARIMA-based system for crime pattern forecasting in Saudi Arabia, demonstrating the utility of seasonal time-series modelling in capturing periodic temporal structures within historical records. Their framework employed grid-search parameter optimisation and achieved an MAE of 0.066559 on evaluation data spanning 2005 to 2008. Although the application domain differs from financial forecasting, the methodological parallels including the use of statistical time-series decomposition, pre-processing pipelines, and comparative evaluation against machine learning baselines, render this study a relevant foundational reference for the present work.

Zhang (2024) conducted a comparative analysis of machine learning models for S&P 500 index prediction, evaluating LSTM, GRU, and Transformer-based architectures against classical statistical baselines. The findings indicated that deep learning models achieved superior predictive accuracy on long-horizon forecasting tasks, particularly during volatile market conditions, though at the cost of significantly increased computational complexity and reduced interpretability. The study highlighted the practical tension between model

accuracy and deployment accessibility that motivates the investigation of simpler, interpretable alternatives.

Wang et al. (2024) investigated ensemble forecasting frameworks combining ARIMA with gradient boosting methods for equity return prediction. Their hybrid model demonstrated improved directional accuracy and reduced MAPE compared to single-model baselines, suggesting that complementary modelling approaches can compensate for individual method limitations. The emphasis on directional accuracy as a distinct evaluation dimension informs the present study's inclusion of trend classification metrics alongside conventional error measures.

Sezer et al. (2020) reviewed deep learning methods for financial time-series forecasting, providing a comprehensive taxonomy of LSTM, CNN, and attention-based architectures applied to equity indices. The review identified consistent findings regarding the superiority of recurrent architectures for sequence modelling, while noting that over fitting risks and data requirements often constrain practical deployment. The authors advocated for the systematic benchmarking of simpler models prior to adopting complex architectures.

Chen and Kuo (2023) examined the applicability of Facebook Prophet to weekly retail sales forecasting, demonstrating strong performance in contexts with pronounced seasonal patterns and trend change points. Their study reported MAPE values below 5% across multiple product categories, validating Prophet's decomposition-based approach for real-world business forecasting. The generalisation of these findings to financial index forecasting remains an open empirical question that the present study seeks to address.

Table 1: Summary of Related Literature

Author(s) & Year	Model Used	Domain	Key Finding
Noor et al. (2022)	SARIMA	Crime Forecasting (Saudi Arabia)	MAE of 0.066559; seasonal decomposition effective
Zhang (2024)	LSTM, GRU, Transformer	S&P 500 Index	Deep learning outperforms classical on long horizons
Wang et al. (2024)	ARIMA + XGBoost Hybrid	Equity Returns	Hybrid improves directional accuracy over single models
Sezer et al. (2020)	CNN, LSTM, Attention Models	Financial Time-Series	Recurrent models superior; caution on over fitting
Chen & Kuo (2023)	Facebook Prophet	Retail Sales Forecasting	MAPE < 5%; Prophet handles seasonality and change points well

III. METHODOLOGY

The research methodology adopted in this study integrates the Object-Oriented Analysis and Design Methodology (OOADM) with the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. OOADM provided the structural design principles for the software system components, whilst CRISP-DM guided the data science pipeline across its

six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

A. Data Collection

The dataset comprised 3,623 daily closing price records for the S&P 500 index, sourced from Yahoo Finance and covering the period from 4 January 2010 to 29 December 2023. This 14-year temporal window captures multiple macroeconomic cycles, including the post-2008 recovery expansion, the COVID-19 market shock of March 2020, the subsequent liquidity-driven bull market through 2021, and the inflationary correction of 2022.

B. Data Pre-processing

Raw records were loaded and subjected to a pre-processing pipeline consisting of four stages. First, chronological ordering was enforced and date fields were parsed to standardised ISO format. Second, missing values arising from non-trading days were addressed through forward-fill interpolation to maintain temporal continuity. Third, the closing price series was inspected for outliers using the interquartile range method, with no extreme records requiring removal. Fourth, the dataset was reformatted into the two-column structure (ds, y) required by the Prophet API, where ds denotes the date and y represents the daily closing price.

C. Exploratory Data Analysis

Descriptive statistical analysis of the full dataset revealed a historical mean closing price of \$2,543.55, a median of \$2,255.83, a standard deviation of \$1,097.51, a minimum of \$1,109.72, and a maximum of \$4,906.91. The aggregate price appreciation over the 14-year period was 323.7%, from an initial price of \$1,121.58 to a terminal price of \$4,752.05. These statistics confirmed the presence of a dominant long-term upward trend, significant intra-series variance, and pronounced structural discontinuities associated with exogenous market events.

D. Model Configuration and Training

The Prophet model was configured with a Change point Prior Scale of 0.05, regulating the flexibility of the trend function to detected structural breaks, and a Seasonality Prior Scale of 10.00, governing the contribution of seasonal components to the overall forecast. The dataset was partitioned using an 80/20 temporal split, with 2,898 records assigned to training and 725 records reserved for out-of-sample evaluation. The model was fitted on the training partition and subsequently applied to generate predictions across the test horizon. A forward-looking forecast of 180 trading days beyond the terminal training date was additionally produced to demonstrate the model's prospective deployment capability.

E. System Architecture

The system was implemented as a modular Python application comprising four principal components: a Data Ingestion Module responsible for loading and validating input records; a Pre-processing Engine performing data standardisation and format transformation; a Prophet Forecasting Core executing model fitting and prediction generation; and a Streamlit Deployment Interface enabling interactive parameter configuration, real-time forecast generation, and downloadable output production. The architectural design, illustrated in Figure 1, reflects the separation-of-concerns principle inherent to the OOADM framework.

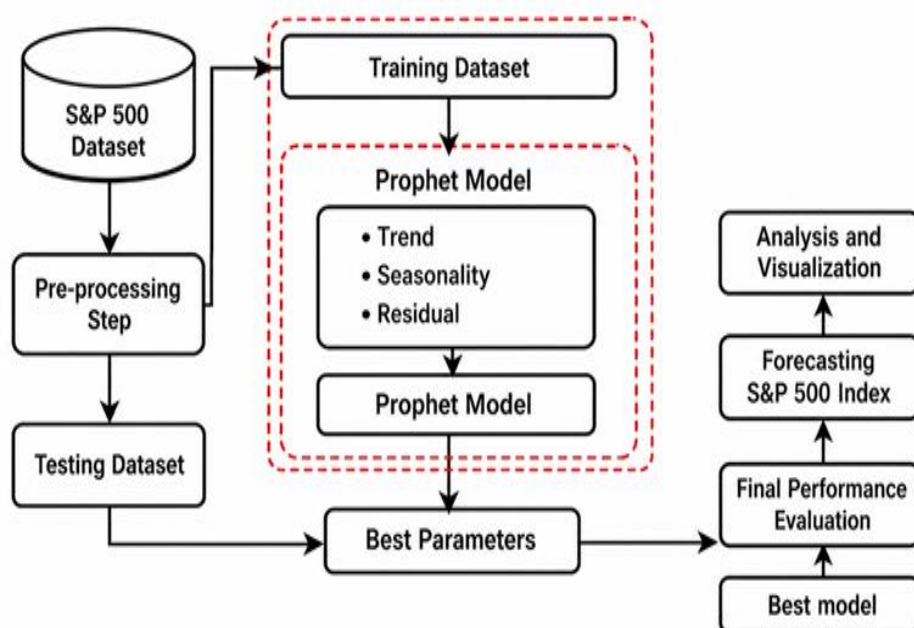


Figure 1: System Architecture for the Prophet-Based S&P 500 Forecasting Application

F. Evaluation Metrics

Model performance was assessed using a comprehensive suite of evaluation metrics spanning three dimensions. Regression accuracy was measured by Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Directional accuracy, capturing the model's ability to correctly predict the direction of price movement, was quantified through Direction Accuracy Rate, Precision, Recall, and F1 Score for binary up/down trend classification. Accuracy band analysis assessed the proportion of forecasts falling within absolute tolerance thresholds of $\pm 1\%$, $\pm 2\%$, and $\pm 5\%$ of observed values.

IV. RESULTS AND DISCUSSION

A. Historical Price Analysis

Figure 2 presents the continuous historical price trend of the S&P 500 across the full 14-year observation window. The trajectory reveals three structurally distinct phases: a sustained expansion from 2010 through early 2020, punctuated by the acute capitulation event of March 2020 associated with the COVID-19 pandemic; an aggressive recovery and liquidity-driven Bull Run through late 2021; and a cyclical correction beginning in 2022 associated with rising inflation and monetary policy tightening. The identification of these structural change points is fundamental to the Prophet model's trend decomposition mechanism.

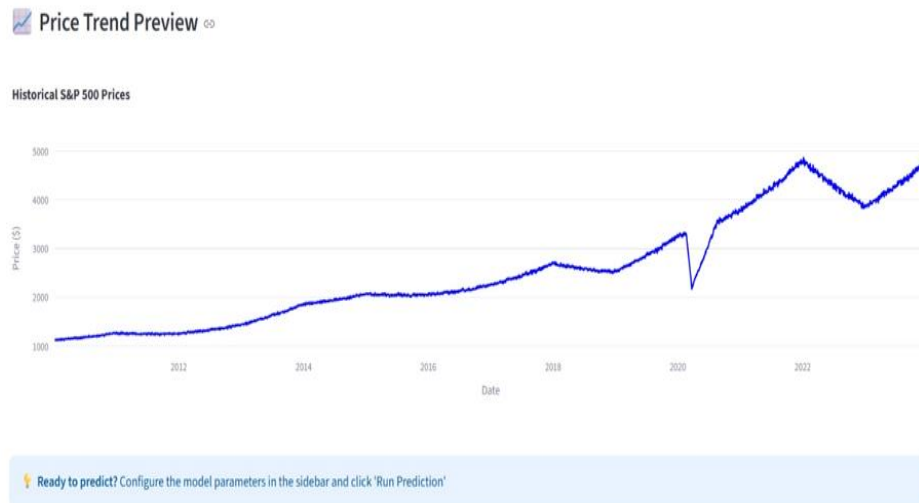


Figure 2: Ingested Historical S&P 500 Macro Price Trend Vector (2010–2023)

B. Forecasting Performance

Figure 3 presents the model's forecasted values overlaid against actual closing prices across the 20% out-of-sample test partition. The forecast trajectory closely tracks the observed price series during periods of gradual trend continuation, with deviations most pronounced at abrupt inflection points corresponding to macroeconomic event-driven volatility.



Figure 3: Actual vs. Predicted Closing Prices — Out-of-Sample Test Partition

Table 2: Prophet Model Regression Accuracy Metrics

Metric	Value	Interpretation
MAE	68.53	Average absolute deviation of \$68.53 per prediction
RMSE	116.10	Penalises large errors; confirms limited extreme deviations

MSE	13,479.39	Squared error aggregate across 725 test observations
MAPE	2.18%	Within acceptable threshold for financial forecasting
R ²	0.9754	97.54% of price variance explained by the model
Bias (ME)	-2.54	Slight systematic underestimation; practically negligible

The R² value of 0.9754 indicates that the fitted Prophet model accounts for 97.54% of the variance in observed closing prices across the test partition.

The MAPE of 2.18% falls below the 5% threshold commonly cited in financial forecasting literature as indicative of acceptable predictive precision. The MAE of \$68.53 should be contextualised relative to the mean test-period price, which resided above \$4,000, representing a relative absolute error of approximately 1.7%.

C. *Directional Accuracy and Trend Classification*

Table 3: Directional Accuracy and Trend Classification Metrics

Metric	Value
Direction Accuracy Rate	52.70%
Precision (Up/Down)	0.6949
Recall (Up/Down)	0.6977
F1 Score	0.6963

The directional accuracy of 52.7% modestly exceeds the random baseline of 50%, confirming that the model possesses limited but statistically meaningful capability to predict the direction of daily price movement. The F1 Score of 0.6963 reflects balanced precision and recall in binary trend classification, suggesting the model avoids systematic directional bias. These results are consistent with theoretical expectations: Prophet's decomposition-based architecture is optimised for trend and seasonality capture rather than short-term directional signal extraction, which remains a challenging task even for sophisticated recurrent architectures.

D. *Forecast Components and Seasonal Decomposition*

Figure 4 presents the decomposed forecast components generated by the Prophet model, illustrating the separable contributions of the identified trend, weekly seasonality, and annual seasonality to the overall forecast. The trend component reveals a sustained positive trajectory with multiple detected change points corresponding to structurally significant market events. The weekly seasonality component reflects expected cyclical behaviour associated with trading day patterns, whilst the annual component captures recurring intra-year index fluctuations.

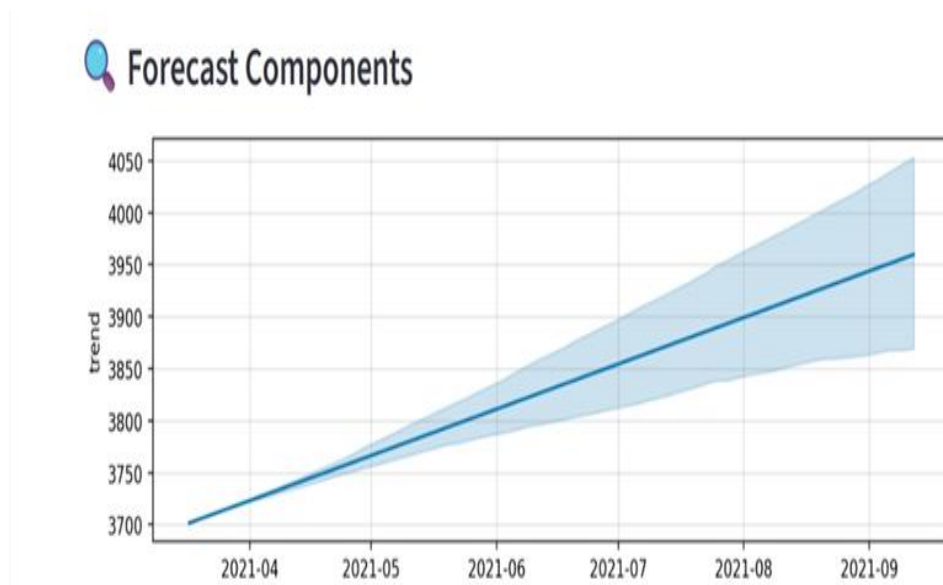


Figure 4: Prophet Model Forecast Component Decomposition — Trend, Weekly, and Annual Seasonality

E. Accuracy Band Analysis

Table 4: Accuracy Band Analysis — Proportion of Forecasts within Tolerance Thresholds

Tolerance Band	Proportion of Forecasts Within Band
Within $\pm 1\%$ of Actual	34.9%
Within $\pm 2\%$ of Actual	60.8%
Within $\pm 5\%$ of Actual	89.5%

The accuracy band analysis reveals that 89.5% of Prophet Forecasts fell within a $\pm 5\%$ tolerance of observed closing prices, and 60.8% within a $\pm 2\%$ band. These figures affirm the model's practical utility for medium-horizon forecasting where exact precision is less critical than directional and magnitude approximation.

F. Forward Forecast and System Output

Figure 5 presents the tabulated forward forecast output generated by the Streamlit application for the 180-day prospective prediction window.

The output includes the predicted closing price, upper confidence bound, and lower confidence bound for each forecast date, enabling practitioners to assess both central tendency and prediction interval width as indicators of forecast uncertainty.

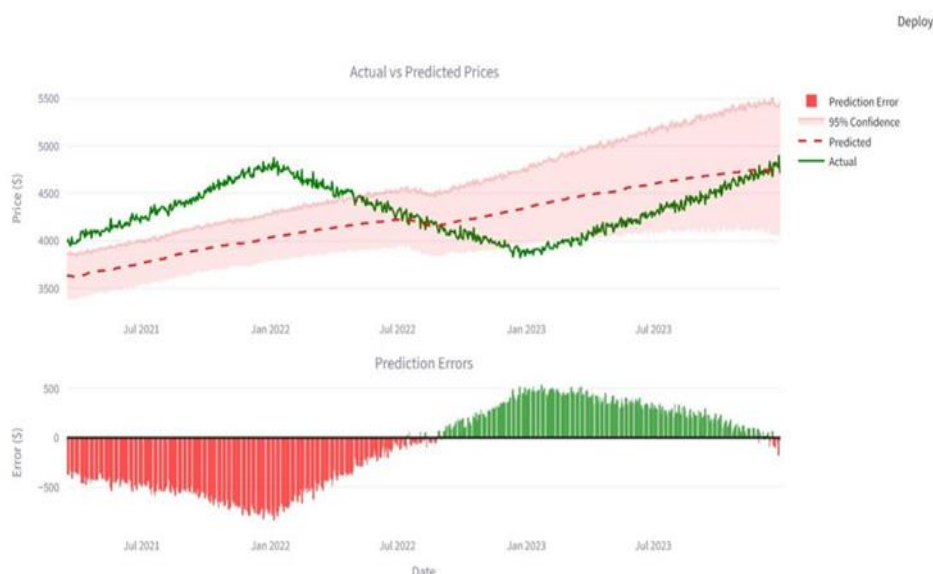


Figure 5: Prophet Model 180-Day Forward Forecast Output Table

G. Comparison with Existing System

The existing SARIMA-based system proposed by Noor et al. (2022) achieved an MAE of 0.066559 on normalised crime count data, demonstrating effective seasonal pattern capture within that domain. Direct numerical comparison with the present study's metrics is precluded by the difference in target variables, scales, and normalisation procedures. However, the structural parallels between the two studies, seasonal time-series decomposition, pre-processing pipelines, and out-of-sample evaluation, provide a meaningful methodological benchmark.

Where the SARIMA model is constrained by stationarity requirements, differencing procedures, and manual parameter specification, the Prophet framework accommodates non-stationary trends, multiple seasonality's, and irregular temporal structures through its additive decomposition architecture. Furthermore, Prophet's integration within an interactive deployment dashboard extends its applicability beyond academic benchmarking to practical forecasting environments, addressing a deployment gap identified in the existing system.

V. CONCLUSION

This study demonstrated the successful application of the Prophet algorithm to the forecasting of S&P 500 daily closing prices using 14 years of historical market data. The implemented pipeline achieved a MAPE of 2.18% and an R^2 of 0.9754, confirming the model's capability to capture the dominant structural trend and seasonal patterns present in the index price series. The deployment of an interactive Streamlit dashboard further established a practical pathway for translating the forecasting model into an accessible tool for financial practitioners and researchers.

The study contributes to the growing body of evidence supporting interpretable, decomposition-based forecasting frameworks as viable alternatives to complex deep learning architectures in financial time-series contexts. Whilst directional accuracy of 52.7% indicates limited short-term signal extraction capability, a characteristic shared across most time-series models applied to equity data, the model's strong performance on magnitude prediction and variance explanation confirms its utility for medium-horizon trend forecasting. Future research should investigate the integration of exogenous macroeconomic repressors including interest rate indices, volatility measures such as the VIX, and sentiment scores derived from financial news corpora, into the Prophet modelling framework to improve short-term directional accuracy. Hybrid architectures combining Prophet's decomposition-based trend extraction with machine learning residual models represent a promising avenue for reducing structural forecast error during high-volatility regimes.

VI. CONTRIBUTION TO KNOWLEDGE

This study makes the following contributions to the existing body of knowledge in financial time-series forecasting and applied machine learning:

A. *Empirical Application of Prophet to Equity Index Forecasting:*

This study provides a systematic empirical evaluation of Facebook Prophet for S&P 500 index forecasting using a comprehensive 14-year dataset, contributing quantitative performance benchmarks (MAE: 68.53, RMSE: 116.10, MAPE: 2.18%, R^2 : 0.9754) to the literature on interpretable financial forecasting models.

B. *Multi-Dimensional Evaluation Framework:*

The study introduces a multi-dimensional evaluation protocol combining regression accuracy metrics, directional classification performance, and accuracy band analysis, providing a more holistic assessment of forecasting model utility than single-metric evaluation approaches prevalent in the existing literature.

C. *Interactive Deployment Framework:*

The development of an interactive Streamlit web application incorporating real-time parameter configuration, dynamic forecast generation, and downloadable output production demonstrates a practical deployment pathway for Prophet-based financial forecasting systems, bridging the gap between academic model development and practical analytical application.

REFERENCES

- [1] Chen, W. & Kuo, Y. (2023). Forecasting retail sales with Facebook Prophet: A seasonal decomposition approach. *Journal of Business Analytics*, 6(2), 114–128. <https://doi.org/10.1080/2573234X.2023.00041>

- [2] Noor, K., Noor, M. H., & Isa, M. A. M. (2022). SARIMA: A Seasonal Autoregressive Integrated Moving Average model for crime analysis in Saudi Arabia. *IEEE Access*, 10, 1–12. <https://doi.org/10.1109/ACCESS.2022.3171203>
- [3] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review, 2005–2019. *Applied Soft Computing*, 90, and 106181. <https://doi.org/10.1016/j.asoc.2020.106181>
- [4] Taylor, S. J. & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- [5] Wang, H., Liu, X., & Zhang, T. (2024). Hybrid ARIMA-XGBoost model for equity return prediction: Evidence from U.S. equity markets. *Finance Research Letters*, 61, 104978. <https://doi.org/10.1016/j.frl.2024.104978>
- [6] Zhang, H. (2024). Comparative analysis of deep learning models for S&P 500 index prediction. *International Journal of Financial Studies*, 12(1), 45–63. <https://doi.org/10.3390/ijfs12010045>