A New Scalable Hybrid Model Approach of High-Dimensional Time Series Forecasting Applications

Numan Yaqoob[1], Fatimah M. Alghamdi[2], Anhar Saad Aloufi[3], Riffat Jabeen[1] and Azam Zaka[4,*]

[1] Department of Statistics, COMSATS University Islamabad, Lahore Campus, Lahore, 54000, Pakistan

[2] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

[3] Department of Mathematics and Statistics, College of Science, University of Jeddah, Jeddah, 21959, Saudi Arabia

[4] Department of Statistics, Government Graduate College of Science, Wahdat Road, Lahore, 54810, Pakistan

INFORMATION

Keywords:
Time series forecasting
hybrid models
prophet model
symmetry
lasso regression
high-dimensional data
model development
feature selection

DOI: 10.23967/j.rimni.2025.10.72646

SCIPEDIA
Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería
RIMNI

A New Scalable Hybrid Model Approach of High-Dimensional Time Series Forecasting Applications

Numan Yaqoob[1], **Fatimah M. Alghamdi**[2], **Anhar Saad Aloufi**[3], **Riffat Jabeen**[1] **and Azam Zaka**[4,*]

[1]Department of Statistics, COMSATS University Islamabad, Lahore Campus, Lahore, 54000, Pakistan

[2]Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

[3]Department of Mathematics and Statistics, College of Science, University of Jeddah, Jeddah, 21959, Saudi Arabia

[4]Department of Statistics, Government Graduate College of Science, Wahdat Road, Lahore, 54810, Pakistan

ABSTRACT

This study introduces a Lasso–Prophet hybrid framework developed to deal with the limitations and gap of Facebook's Prophet model. The approach begins with Prophet's decomposition of a time series into its fundamental components trend, seasonality, and holiday effects, and then applies Lasso regression to the residuals to capture additional structural patterns that fail to capture by the base model. This layered methodology boosts predictive accuracy by enabling the model to learn both systematic and irregular variations within temporal data by incorporating Lasso's feature selection capability, the framework efficiently handles high-dimensional datasets, retaining only the most informative predictors. The outcome is a hybrid model which achieves an optimal balance among interpretability, scalability, sparsity, and forecasting precision. Validation on simulated high-feature datasets and real-world electricity consumption data demonstrates that the Lasso–Prophet hybrid consistently outperforms the Prophet and other baseline models.

OPEN ACCESS

Received: 31/08/2025

Accepted: 17/11/2025

Published: 23/01/2026

DOI
10.23967/j.rimni.2025.10.72646

Keywords:
Time series forecasting
hybrid models
prophet model
symmetry
lasso regression
high-dimensional data
model development
feature selection

1 Introduction

Time series forecasting is important in many areas because it helps businesses predict future trends by using their resources wisely, and make better decisions based on data. Real-world data is complex, characterized by high dimensionality, and changing patterns over time. There is a strong need for new forecasting techniques or a framework that can precisely predict future values, handle large, unbalanced, high dimensional and complicated datasets efficiently, and endure easily interpretable. This study introduces the Lasso-Prophet Hybrid Model, by joining the advantages of Facebook's Prophet method (which breaks down time series into understandable components) with

*Correspondence: Azam Zaka (drazamzaka@gcslahore.edu.pk).

Lasso regression, a technique that is good at selecting relevant predictors from large sets of variables. This method is purposes to improve prediction accuracy and clarity, this makes it suitable for real-world applications.

Historically, traditional statistical models like ARIMA, Holt Winters and Exponential Smoothing have dominated the field of forecasting because they are comparatively upfront and easily interpretable (Box et al., 2015) [1]. These methods have clear parameters and no black box methods like machine learning models. Durbin & Koopman (2012) [2] decorated how these models successfully identify trends and seasonal patterns in data, but they also acknowledged that advanced variants like Seasonal ARIMA (SARIMA) typically need extensive manual adjustments. This makes them very challenging to apply efficiently in complex and high-dimensional environments where there is a high number of predictors.

Time series forecasting has advanced from classical linear approaches to hybrid and deep learning–based models that can balance interpretability, scalability, and predictive accuracy. Taylor and Letham (2018) [3] presented work on scalable and interpretable forecasting through the development of the Prophet model. This model decomposes time series data into trend, seasonality, and holiday components inside an additive regression framework. This was designed and presented for business applications, Prophet highlights flexibility, robustness to missing data and outliers, and ease of analyst intervention. In that way, this makes it forecasting at scale without the deep statistical expertise. While Prophet delivers strong interpretability and modularity, its residual components frequently contain unexplained variation driven by external predictors or high-dimensional relationships that the model's additive structure cannot entirely cover. Prophet's limitations in modeling complex temporal dependencies were addressed first by Triebe et al. (2021) [4]. They introduced NeuralProphet, which is a deep learning extension that integrates autoregressive components and covariate networks into the classical decomposition framework. It was first built on PyTorch, NeuralProphet, it retains Prophet's transparency while leveraging neural networks to capture short-term dependencies and local dynamics. This showed improvements in short to medium term forecast accuracy in all multiple datasets. This evolution showed the increasing credibility that Prophet's statistical decomposition can be effectively improved by further modeling layers, where we can capture residual patterns. In the same way, many other studies have shown us that combining Prophet with corresponding models can significantly improve forecasting performance.

Arslan (2022) [5] presented a hybrid model that merges a long short-term memory (LSTM) network with Prophet for energy consumption forecasting, with Seasonal–Trend Decomposition through Loess (STL) to isolate trend, seasonal, and residual components in the model. Although Prophet modeled the overall trend and periodicity, LSTM captured residual non-linearities and temporal dependencies, which resulted in better forecasting accuracy across all seven national energy datasets. Liu et al. (2022) [6] further proceed this concept through the P-gLSTNet model. This wraths Prophet with an upgraded LSTM network to forecast complex environmental and pollutant emission data from more than 3500 vehicles. Results established this framework of combining Prophet's competence for catching long-term seasonality with LSTM's strength in short-term dynamics produced lower MAE, RMSE, and MAPE. Huang et al. (2022) [7] extended Prophet's hybridization to financial forecasting, by mixing and integrating it with Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), and multi-model optimization for nonferrous metal price prediction. Prophet modeled the primary trend and seasonality, and while residual components were decomposed and forecasted using multiple complementary algorithms. These models are ARIMA, BPNN, LSTM, and Elman networks. An optimization layer carefully chooses the best-performing forecasts from each sub-model of the study. This multi-stage hybrid model accomplished better predictive performance and robustness on London Metal Exchange data. Stressing Prophet's effectiveness as a base model in hybrid frameworks when merged with methods that are able to model complex residuals. Furthermore, in this field, Kenyi and Yamamoto (2024) [8] proposed a SARIMA–Prophet model for predicting streamflow in South Sudan. In this model, Prophet captured non-linear long-term patterns and SARIMA modeled short-term stationarity, this produced a reduction in MAE and RMSE. The model performed well with other comparative models this proved the value of complementing Prophet with additional statistical structures for hydrological data.

Recently, Sherly, Christo, and Elizabeth (2025) [9] worked on ARIMA–Prophet hybrid framework which is designed for forecasting in edge computing contexts. In this study, IoT-generated data often exhibits both stationary and non-stationary characteristics. Prophet handled global non-linearities and seasonality, and while ARIMA captured local autocorrelation and short-term dependencies, this gave output in improved performance in all multiple datasets. Together, these studies reveal a consistent research trajectory. Prophet's modular, interpretable frameworks' provides a robust basis, with hybridization with other models whether neural, statistical, decomposition-based provide us good results. In spite of these advances, most Prophet model frameworks efforts focus on deep learning or multi-model combinations. There is a gap in methods that handle sparsity, feature selection, additive framework and high-dimensional residual modeling using interpretable methods. In this background, we can say that the integration of Lasso regression offers a convincing extension to the Prophet model. As compared to Neural or ensemble-based extensions, Lasso performs regularization and feature selection simultaneously, which makes it mainly effective in high-dimensional data environments where residual variation arises from numerous external predictors. Prophet proficiently models large-scale trends and seasonality. But its residuals may still contain structured variance driven by auxiliary features like sensor signals, exogenous variables process indicators that are not directly modeled. When we apply Lasso to Prophet's residuals, sparse and interpretable relationships among high-dimensional predictors can be uncovered. This also helps to handle overfitting in the model. Prophet + Lasso forms a semiparametric hybrid model that unites Prophet's additive decomposition with Lasso's penalized regression strength. This makes it scalable, and interpretable.

The Prophet model has been used in many fields for forecasting. In recent years, many applications have been observed, as it is flexible. Sivaramakrishnan et al. (2021) [10] applied ARIMA and Prophet for predicting air passenger numbers and it was seen that Prophet is easier to use as it automatically handles trends and seasonality. Similarly on these lines Chaturvedi et al. (2022) [11] compared SARIMA, LSTM-RNN, and Prophet for forecasting India's energy demand, showing that Prophet gave more accurate and understandable results. In the same year, Zhao et al. (2022) [12] analyzed ARIMA, MLR, and Prophet for forecasting global Omicron COVID-19 cases, finding that ARIMA performed best, but Prophet still gave dependable results. Basak et al. (2022) [13] applied Prophet for drought forecasting in semi-arid India and confirmed it as a consistent and trustworthy model for environmental data. Shakeel et al (2023) [14] improved the FB-Prophet model by adding positional encoding layers for heat load forecasting achieving better precision than other models. Ma et al. (2023) [15] worked on a Multiscale Superpixelwise Prophet Model (MSPM) for analyzing hyperspectral images. In this, they proved more resistant to noise and accurate than advanced alternatives. Sardar et al. (2023) [16] combined Prophet with ARIMA and other machine learning models to forecast COVID-19 cases in SAARC countries, this helped to show Prophet's flexibility. Most recently, Elneel et al. (2024) [17] used Prophet, ARIMA, and VAR to predict global and regional sea levels. They showed that Prophet is especially effective in capturing long-term climate trends. We can say overall these studies show how Prophet very helpful in forecasting and more work is expected to improve and extend this model methodology for different data problems.

Machine learning methods popularity have grown for forecasting tasks in recent years, mainly because of their ability to handle nonlinear patterns and uncover subtle relationships in data. Methods such as Gradient Boosted Trees, Random Forests, and especially Long Short-Term Memory (LSTM) neural networks have shown good results. Hochreiter and Schmidhuber (1997) [18] first introduced LSTM networks and following studies have presented their powerful predictive capabilities. Although their robust performance, machine learning models often operate like a black box which means users can't easily see or understand how the model arrived at its predictions. Molnar (2020) [19] noted that its lack of transparency can limit its practical use, especially in sectors where understanding the underlying reasons behind predictions is very important. To handle interpretability issues, Facebook introduced the Prophet model. This is designed explicitly for forecasting business and economic time series (Taylor & Letham, 2018) [3]. Lasso shrinks coefficients toward zero using L_1 Plenty. This effectively removes irrelevant features and simplifies the model. Zou and Hastie (2005) [20] introduced the Elastic Net, which is also a regularization method that combines the penalties of both Lasso and Ridge regression to improve variable selection. This method has successfully dealt with multicollinearity and selects groups of correlated predictors.

The prophet's incapability to efficiently include and regularize a large set of external predictors is a major limitation in high-dimensional external predictors, along with seasonality and trend. When numerous correlated predictors like macroeconomic, environmental, and behavioral variables influence the target, Prophet handles them as independent regressors without automatic feature selection or shrinkage. This leads to overfitting, multicollinearity, poor generalization and poor performance in deployment. Prophet therefore struggles to identify the actually useful predictors and cannot exploit the sparsity structure that high-dimensional data naturally requires. This contribution highlights the development of a hybrid modeling framework that integrates Prophet's additive decomposition with Lasso regression. This approach creates methodological balance in a way that Prophet captures structured temporal patterns while Lasso identifies the most influential external drivers or predictors, and a systematic hyperparameter tuning procedure makes sure that the Lasso penalty (λ) is optimally selected by enhancing both prediction accuracy and interpretability.

2 Methodology

Let the observed time series be represented as $y(t)$, where $t = 1, 2, \ldots, n$. The objective is to model $y(t)$ as a combination of structured components (trend, seasonality, holidays) and external predictors:

$$y(t) = g(t) + s(t) + h(t) + X(t)\beta + \epsilon t, \tag{1}$$

where:

- $g(t)$: Trend component (long-term changes).
- $s(t)$: Seasonal component (cyclic patterns).
- $h(t)$: Holiday component (irregular spikes or dips due to special events).
- $X(t)$: High-dimensional external predictors matrix.
- β: Coefficient vector for external predictors.
- $\epsilon_t \sim N(0, \sigma^2)$: Error term assumed to follow a normal distribution.

The primary goal is to estimate $g(t)$, $s(t)$, $h(t)$, β, and σ^2 while selecting significant predictors from $X(t)$.

2.1 *Decomposition via Prophet*

The Prophet model decomposes $y(t)$ into three additive components:

2.1.1 *Trend Component ($g(t)$)*

Trend components have two further types, one is a piecewise linear trend, and the other is logistic growth. Piece wise linear trend is as follows:

$$g(t) = k(t) + m \text{ if no change point} \tag{2}$$

$$g(t) = k + \sum_{j=1}^{j} a_j I_j(t) \text{ if changpoint occur at } t, \tag{3}$$

where k is the growth rate, m is the initial offset, J is the number of changepoints, a_j adjusts the growth rate at each changepoint, and $I_j(t)$ is an indicator function for the changepoints.

Similarly logistic growth is described as follows

$$g(t) = \frac{C}{1 + exp(-k(t - t_0)}, \tag{4}$$

where C is the carrying capacity, k is the growth rate, and t_0 is the midpoint.

2.1.2 *Seasonal Component ($s(t)$)*

Seasonal effects are modeled using a Fourier series:

$$s(t) = \sum_{n=1}^{N}(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right), \tag{5}$$

where P is the periodicity, and N is the number of Fourier terms.

2.1.3 *Holiday Component ($h(t)$)*

Holidays are modeled as a sum of indicator functions:

$$h(t) = \sum_{k=1}^{K} \lambda_k I_k(t), \tag{6}$$

where λ_k represents the effect of holiday k, and $I_k(t)$ is an indicator function for the holiday.

2.2 *Residual Modeling with Lasso Regression*

After extracting the structured components, the residuals are computed as:

$$r(t) = y(t) - (g(t) + s(t) + h(t)). \tag{7}$$

These residuals are modeled using Lasso regression to account for the influence of external predictors $X(t)$. The Lasso regression objective function is:

$$L(\beta) = \frac{1}{2n} \sum_{t=1}^{n} (r(t) - X(t)\beta)^2 + \lambda \sum_{j=1}^{p} \beta_j \tag{8}$$

where:

- λ: Regularization parameter controlling the sparsity of β.

- p: Number of predictors in $X(t)$.
- The L_1 penalty $\sum_{j=1}^{p} \beta_j$ ensures that irrelevant predictors are excluded by shrinking their coefficient to zero.

The optimal value of λ is estimated through k-fold cross-validation, where multiple λ values are tested and the one minimizing the prediction error on validation folds is selected. This method balances model sparsity and predictive accuracy. Larger λ values produce simpler and more interpretable models and smaller λ values retain more predictors.

2.3 *Error Variance* (σ^2)

The estimated from the residuals are given

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{t=1}^{n} \left(y(t) - \left(g(t) + s(t) + h(t) + X(t)\beta\right)\right)^2 \tag{9}$$

2.4 *Model Evaluation*

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y(t) - \hat{y}(t)\right)^2} \tag{10}$$

2.5 *Feature Significance*

Significant predictors are identified based on non-zero coefficients in β.

3 Model Algorithm Framework

The application of the hybrid model follows a two-step method yet interconnected approach which combine statistical decomposition with machine learning based residual learning. The target variable $y(t)$ is broken down using Prophet into three core components which are trend $g(t)$, seasonality $s(t)$, holiday effects $h(t)$. This step helps separate the structured part of the time series. Once these components are extracted, the next logical move is to calculate the residuals, which are given in Eq. (11).

$$r(t) = y(t) - (g(t) + s(t) + h(t)). \tag{11}$$

These residuals are thought to contain extra patterns or noise not captured by Prophet so to model this leftover variation Lasso regression is applied using a set of external predictors $X(t)$, effectively learning the sparse relationships from a potentially high-dimensional feature set. The final forecast is made by joining the main components with the Lasso correction, as shown in the Eq. (12) below.

$$\hat{y} = g(t) + s(t) + h(t) + X(t)\hat{\beta} \tag{12}$$

This framework in Fig. 1 not only improves prediction power but also makes sure that both clear patterns and hidden factors are included, which gives us a more flexible and detailed way to forecast.

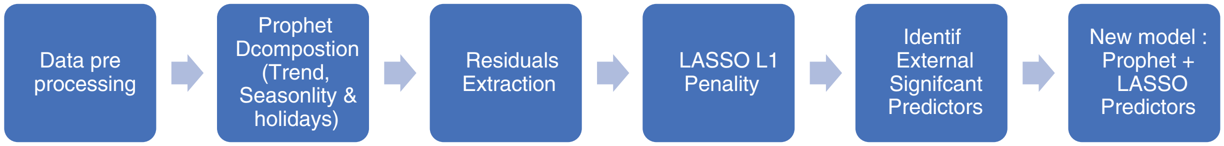

Figure 1: Flow chart of methodology

4 Synthetic Dataset and Time Series Decomposition

The dataset was generated by creating 100 features using a normal distribution with values roughly ranging between 10 and 50, a mean of about 29, and a standard deviation between 11 and 15. The target variable (y) was then constructed using selected features with added random noise to introduce realistic variability. This design allowed the regularized models to effectively recognize the most influential predictors, as all variables were standardized using z-score normalization to ensure consistency in scale and improve the stability of model estimation and for evaluation, the dataset was divided using the first 80% of observations for training and the remaining 20% for testing. This preserved the temporal order and avoided data leakage. This time-based division provided a realistic assessment of forecasting performance on unseen future data.

The first phase of the methodology is about breaking down the time series into its fundamental components and these components are trend, seasonality, and residuals. Fig. 2 shows the decomposition results for the target variable. The decomposition shows a clear long-term trend component that explains the overall growth and decline in the time series and seasonal patterns with recurring fluctuations and residuals that capture the remaining variability in the data after accounting for trend and seasonality.

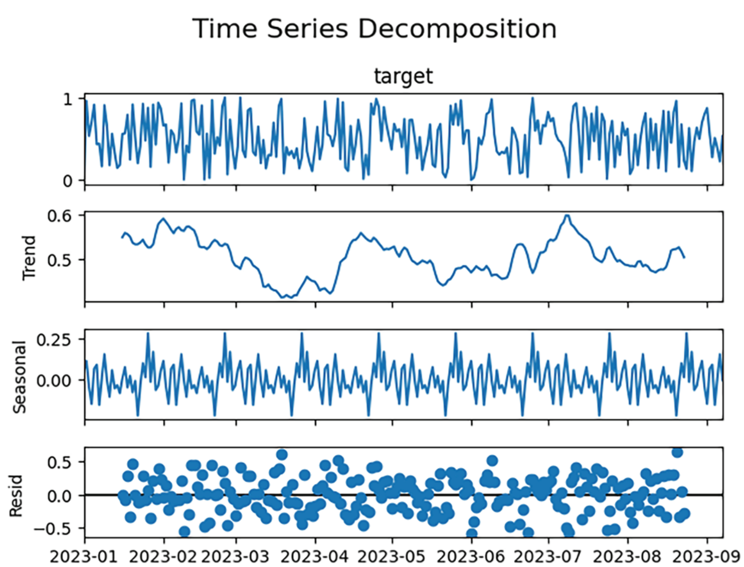

Figure 2: Time series decomposition of target variable

Table 1 presents the performance evaluation of Prophet, its extended hybrid variants, and the Holt-Winters baseline with three error metrics like MAE, RMSE, and MAPE. The Prophet+Lasso model has gained the lowest error values in all metrics. This shows us its superior ability to refine residuals and capture underlying patterns in the synthetic data. Ridge and ElasticNet hybrids both of these also improved over the baseline Prophet model. The Holt-Winters model showed comparatively higher error values. These results confirm that the Lasso-enhanced Prophet hybrid offers the most accurate and stable forecasts among all tested models. The Lasso regularization parameter (λ) was optimized with the GridSearchCV procedure. A logarithmic grid of 120 values for α (equivalent to λ) was defined in the range 10^{-7} to 10^{-2} this makes sure fine resolution with all low penalty regions. Each configuration was assessed using 10-fold cross validation with the negative mean squared error ($-$MSE) it is used as the scoring criteria. The λ value minimizing the mean validation error was selected as optimal, and this cross-validated tuning balances predictive accuracy and sparsity. A higher λ promotes feature elimination, and a lower λ retains more predictors.

Table 1: Forecast performance metrics (synthetic dataset)

Model	MAE	RMSE	MAPE (%)
Prophet	2.356	2.999	10.68
Prophet + Lasso	1.622	2.052	7.53
Prophet + Ridge	1.771	2.242	7.97
Prophet + ElasticNet	1.676	2.114	7.74
Holt-Winters	3.225	4.021	14.17

Fig. 3 shows the comparison between the actual synthetic time-series values and the predictions generated by Prophet, its hybrid extensions (Prophet + Lasso, Prophet + Ridge, Prophet + Elastic-Net), and the Holt-Winters model, which serves as another baseline forecasting approach. The black line represents the true synthetic observations, while the colored dashed lines depict the predictions from each model. All of the forecasting models capture the general trend and seasonality of the data, but Prophet + Lasso gives the nearest fit to actual values. This indicates that the new proposed model effectively refines Prophet's residuals by achieving improved accuracy over the standard Prophet and other baseline models, such as Holt-Winters.

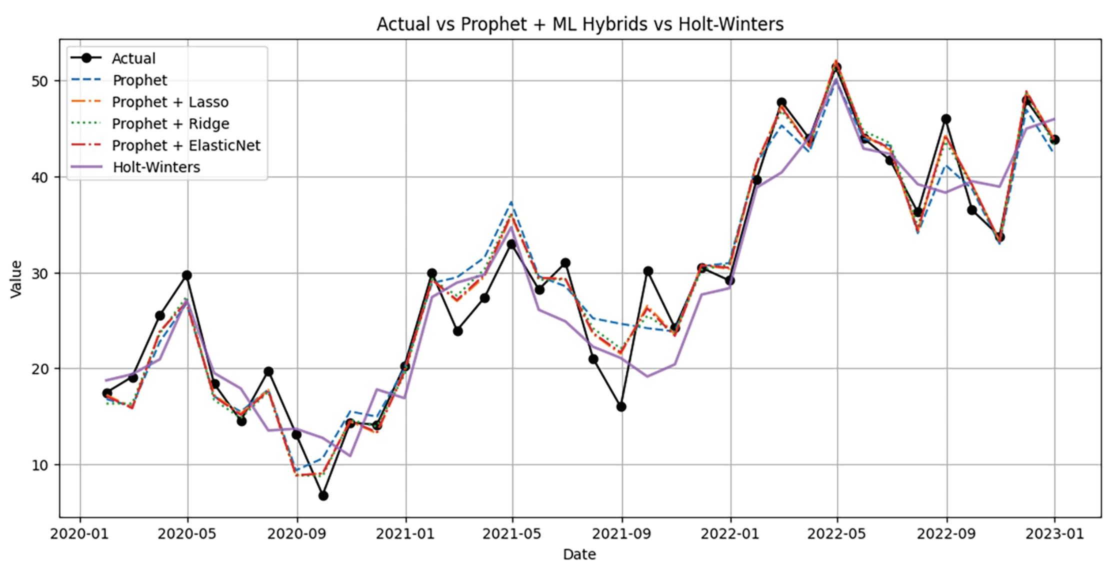

Figure 3: Comparison of actual vs. prophet-based hybrid forecasting (synthetic data)

Fig. 4 displays the top 10 most influential features/predictors selected by the Lasso model from the synthetic dataset, and in this figure, features like feature_47, feature_19, and feature_38 contributed the most to reducing residual forecast errors. This plot highlights the sparsity and interpretability advantage of Lasso in a high-dimensional environment.

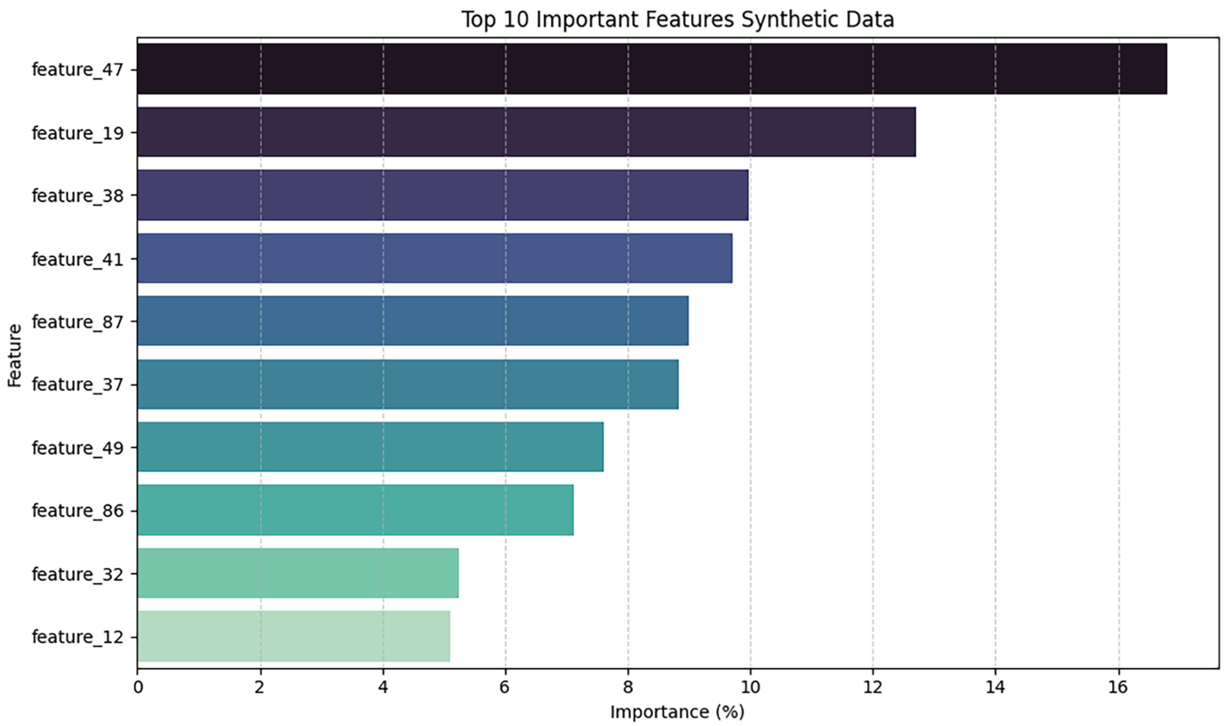

Figure 4: Top 10 important features (synthetic data)

Figs. 5 and 6 show the distribution and temporal movement of forecast errors which are produced by the Prophet model. The histogram shows that most residuals are centered around zero and indicate that the model captures the underlying trend effectively, with only a few large deviations. However, occasional outliers suggest under- or over-prediction in certain periods.

The time series plot of forecast errors also explains the positive and negative deviations, which show us that errors are not systematically biased but fluctuate around the zero line. The Prophet model overall shows balanced performance with moderate variability and no persistent trend in forecast errors.

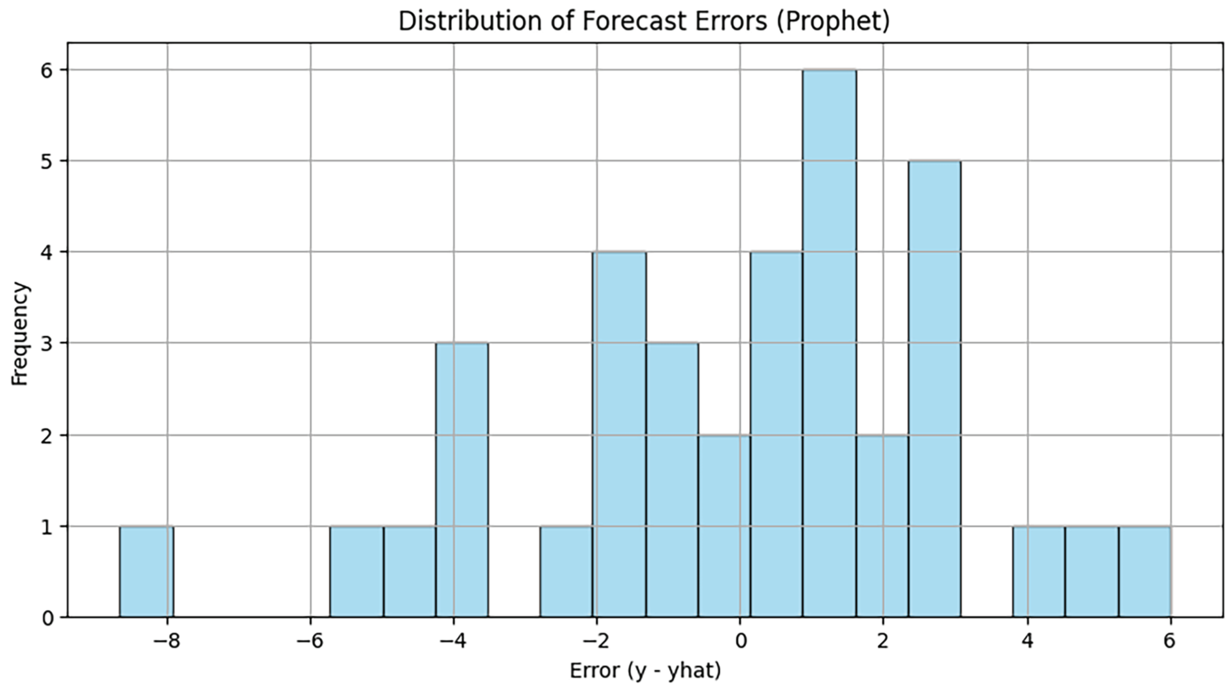

Figure 5: Distribution of forecast error of Prophet model

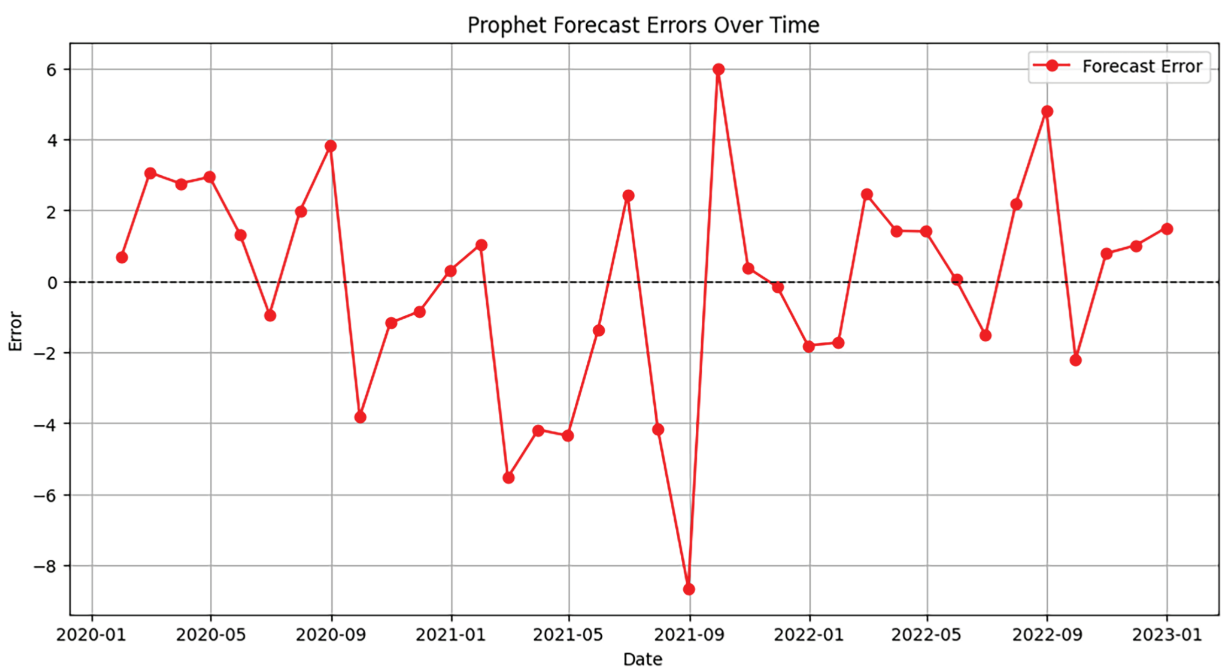

Figure 6: Prophet forecast errors over time

5 Prophet Forecast Performance on Real Data

The real-world dataset and synthetic datasets both used for this experiment, real word data set is sourced from the UCI Electricity Load Diagrams 2011–2014. It includes 15-min aggregated electricity consumption for 370 clients over a 4-year period in kW, which are 140,256 rows, and we converted to kWh for monthly-level forecasting. The target variable was chosen as MT_001 (representing a single client), and all other meter columns (MT_002 to MT_370) were treated as potential predictive features. The dataset was aggregated to a monthly frequency, resulting in a time series of 48 monthly observations from 2011–01 to 2014–12. The first model employed was Facebook Prophet, a decomposition-based forecasting model. Fig. 7 shows the comparison between the actual electricity consumption values and the forecasts generated by Prophet, its hybrid variants (Prophet + Lasso, Prophet + Ridge, Prophet + ElasticNet), and the Holt-Winters baseline model. The black line shows the actual electricity demand and the colored dashed lines show the predictions from different models, all models were able to follow the main trend and seasonal changes in electricity use, showing high and low periods we can see the Prophet + Lasso model matched the real data most closely, especially around the peaks and drops showing it can better capture changes in demand.

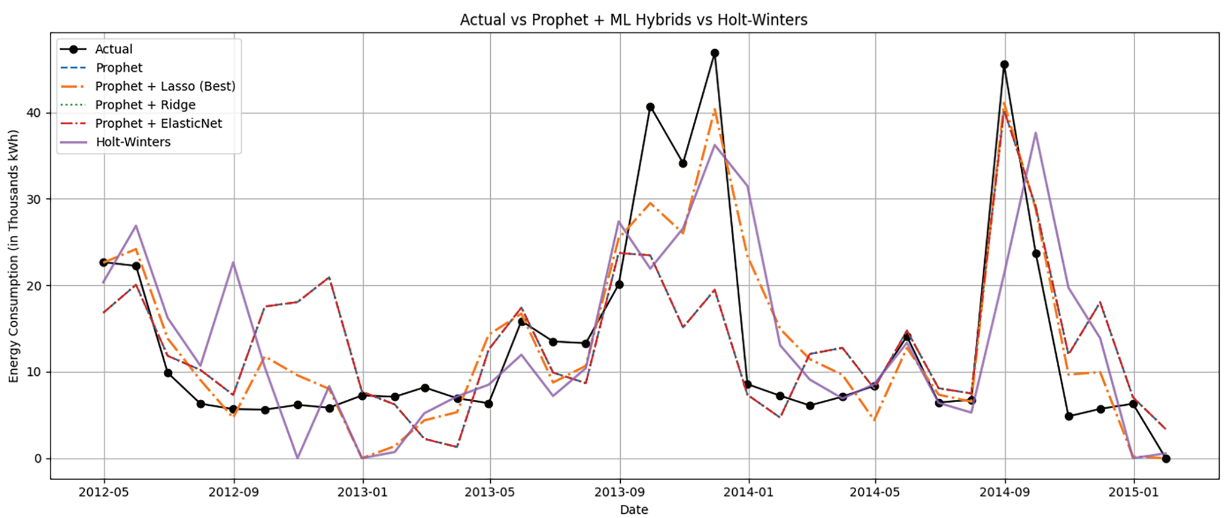

Figure 7: Actual vs. prophet + ML hybrids vs. Holt-Winters **(real data)**

Table 2 shows how well each model predicted electricity use and the Prophet–Lasso hybrid model gave the best results with the lowest MAE, RMSE, and MAPE values, showing it could capture hidden demand patterns and seasonal changes more accurately. The regular Prophet model and its Ridge and ElasticNet versions made fairly good predictions but were slightly less accurate. The Holt-Winters model, which is a traditional method had the highest errors proving that the Prophet–Lasso hybrid provided the most accurate forecasts for electricity consumption.

Fig. 7 clearly shows that the improvement in the hybrid forecast (green dashed line) almost matches the actual electricity demand during the test period, and it is shows strong prediction ability. The Lasso part selected a small but very useful set of features from the 369 m readings and other variables, including nearby meter signals and past values of the target series.

Table 2: Forecast performance metrics (real dataset)

Model	MAE	RMSE	MAPE (%)
Prophet	5.986	8.554	4.47
Prophet + Lasso	4.374	5.414	3.27
Prophet + Ridge	5.986	8.554	4.47
Prophet + ElasticNet	5.986	8.554	4.47
Holt-Winters	6.677	9.188	4.98

Fig. 8 shows the top ten most important features based on their scores. The experiment tested the hybrid model in the same way on both simulated and real datasets using the same accuracy measures and graphs. In each case, the results from the basic Prophet model were compared with the Lasso–Prophet hybrid to highlight the improvement in accuracy and understanding of the model.

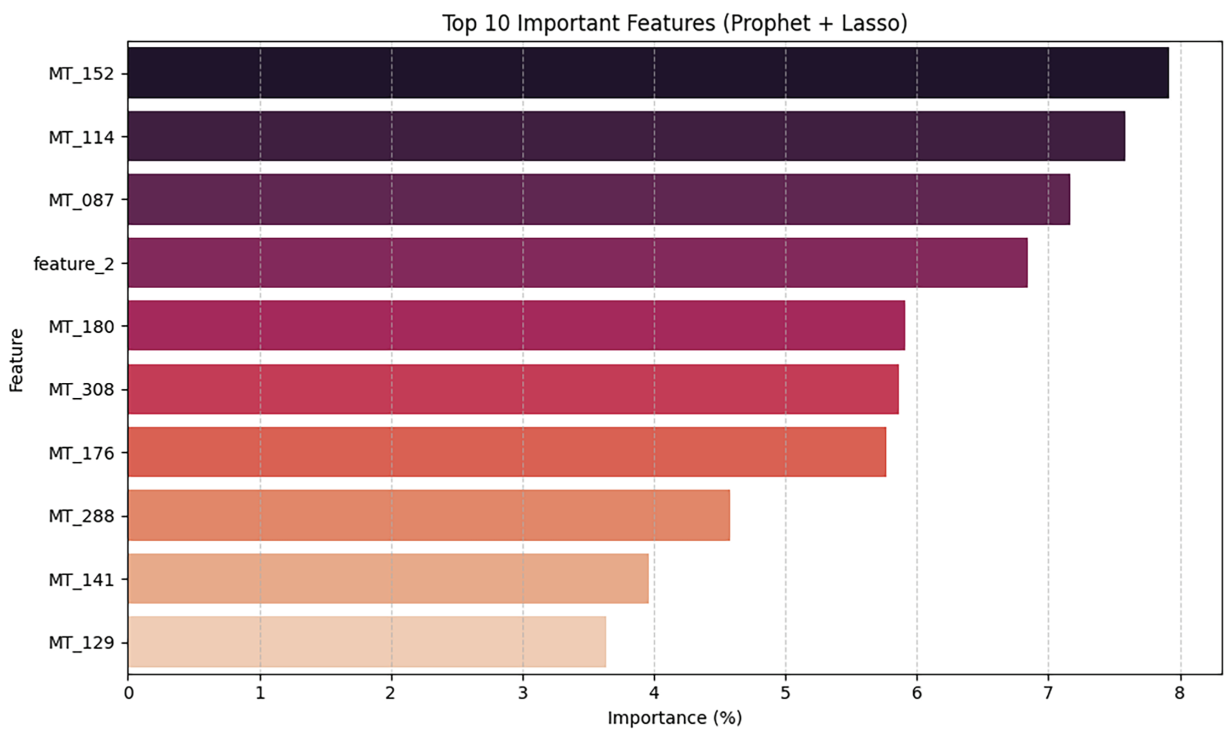

Figure 8: Top 10 important features (real dataset)

Fig. 9 shows the histogram of forecasting errors (Actual–Predicted) for the Prophet model used on electricity data. This figure shows that most of the errors are close to zero, which means that the model predicts the overall pattern quite well. This figure also shows that a few large errors on both sides show that the model sometimes overestimates or underestimates demand, likely due to sudden changes in electricity use, but overall, the error pattern suggests that the Prophet model gives fairly balanced and unbiased forecasts.

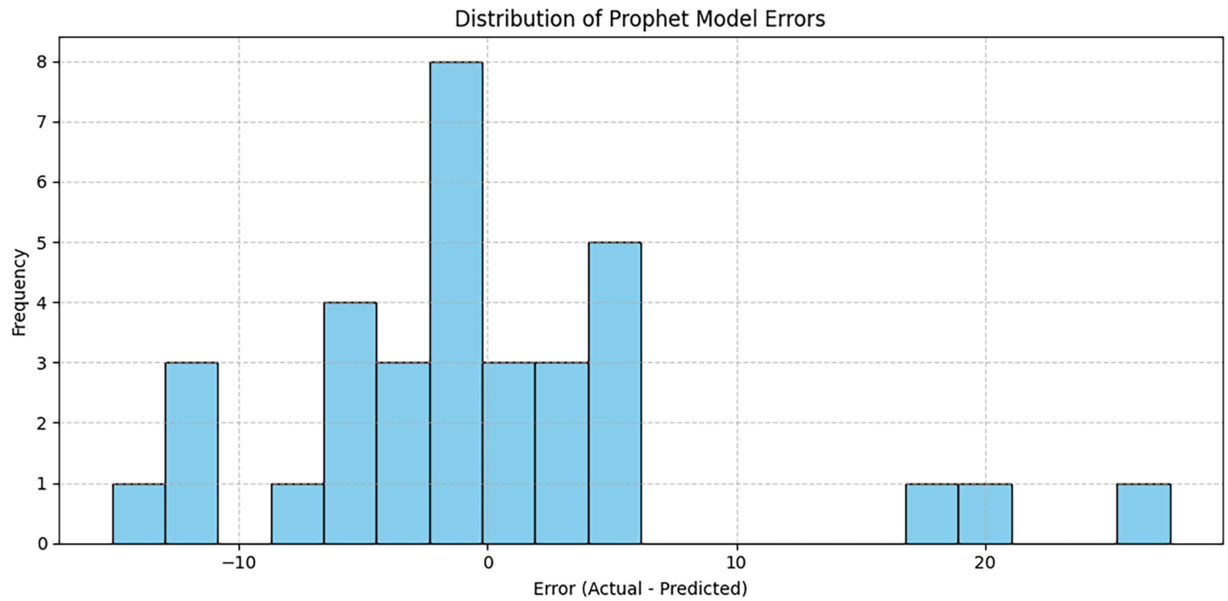

Figure 9: Distribution of Prophet model errors (real data)

Fig. 10 shows how the Prophet model's errors change over time, the red line moves around the zero line it explain that periods when the model slightly overestimates or underestimates demand. This means the errors are not one-sided and the model stays stable across time. A few sharp spikes appear at certain points, matching times of sudden changes in electricity demand. The Prophet model follows the general pattern well, by keeping a steady level of accuracy throughout the observed period.

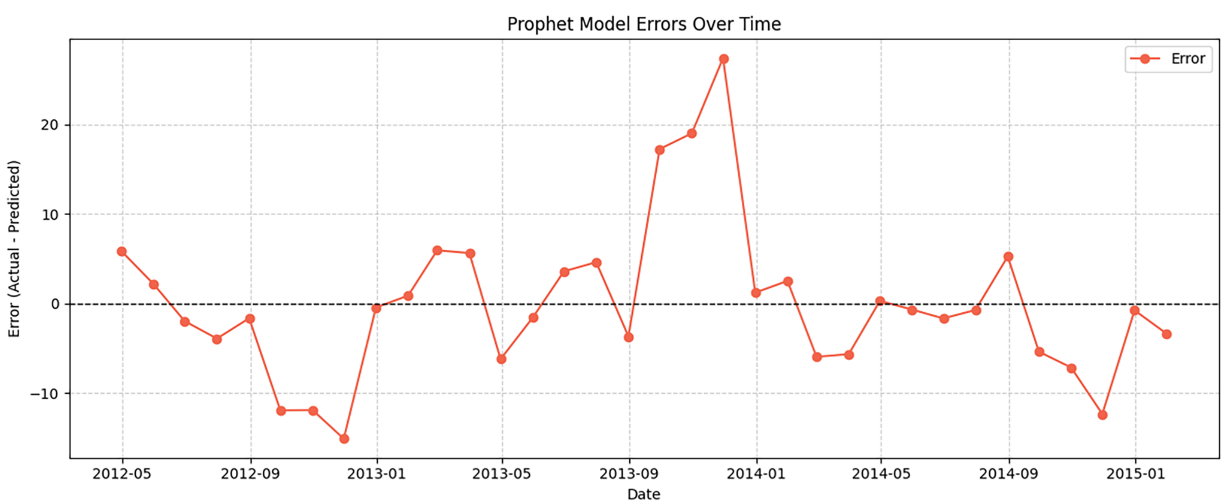

Figure 10: Prophet model errors over time (real data)

6 Conclusion

This research introduced and tested the Lasso–Prophet Hybrid Model, a new and effective approach for forecasting high-dimensional time series data by combining Prophet's ability to break down a time series into trend and seasonal parts with Lasso regression's strength in selecting important variables and avoiding overfitting. The hybrid models showed clear improvements in prediction accuracy for both simulated and real electricity data by capturing local changes and reducing errors. It also stayed easy to interpret and worked well with datasets containing hundreds of features, making it useful for real-world problems. The findings show that combining statistical models with machine learning can lead to more accurate and reliable forecasts, especially in fields like energy, finance, and healthcare. Future improvements could include using nonlinear models, such as kernel or neural-network regressors, to capture more complex patterns, and adding adaptive tuning methods or Bayesian techniques to make the model even more stable and explainable.

Acknowledgement: This work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R735), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R735), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Numan Yaqoob: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Validation, Data Curation, Writing—Original Draft Preparation, Writing—Review and Editing, Visualization. Fatimah M. Alghamdi: Conceptualization, Methodology, Benchmark Model Selection, Writing—Review and Editing. Anhar Saad Aloufi: Conceptualization, Methodology, Formal Analysis, Writing—Review and Editing. Dr. Riffat Jabeen: Conceptualization, Validation, Resources, Supervision, Project Administration, Writing—Review and Editing. Dr. Azam Zaka: Validation, Code Optimization, Supervision, Writing, Review and Editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The dataset analyzed in this study is publicly available from the UCI Machine Learning Repository under the name Individual Household Electric Power Consumption Dataset and all data were accessed and processed according to the repository's usage guidelines.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. 5th ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015.
2. Durbin J, Koopman SJ. Time series analysis by state space methods. 2nd ed. Oxford, UK: Oxford University Press; 2012.
3. Taylor SJ, Letham B. Forecasting at scale. Am Stat. 2018;72(1):37–45. doi:10.1080/00031305.2017.1380080.
4. Triebe O, Hewamalage H, Pilyugina P, Laptev N, Bergmeir C, Rajagopal R. NeuralProphet: explainable forecasting at scale. arXiv:2111.15397. 2021. doi:10.48550/arxiv.2111.15397.
5. Arslan S. A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data. PeerJ Comput Sci. 2022;8(3):e1001. doi:10.7717/peerj-cs.1001.
6. Liu W, Yu X, Zhao Q, Cheng G, Hou X, He S. Time series forecasting fusion network model based on prophet and improved LSTM. Comput Mater Contin. 2023;74(2):3199–219. doi:10.32604/cmc.2023.032595.

7. Huang YT, Bai YL, Yu QH, Ding L, Ma YJ. Application of a hybrid model based on the Prophet model, ICEEMDAN and multi-model optimization error correction in metal price prediction. Resour Policy. 2022;79(8):102969. doi:10.1016/j.resourpol.2022.102969.
8. Kenyi MGS, Yamamoto K. A hybrid SARIMA-Prophet model for predicting historical streamflow time-series of the Sobat River in South Sudan. Discov Appl Sci. 2024;6(9):457. doi:10.1007/s42452-024-06083-x.
9. Sherly A, Christo MS, Elizabeth JV. A hybrid approach to time series forecasting: integrating ARIMA and prophet for improved accuracy. Results Eng. 2025;27(20):105703. doi:10.1016/j.rineng.2025.105703.
10. Sivaramakrishnan S, Fernandez TF, Babukarthik RG, Premalatha S. Forecasting time series data using ARIMA and facebook prophet models. In: Big data management in Sensing. New York, NY, USA: River Publishers; 2022. p. 47–59. doi:10.1201/9781003337355-4.
11. Chaturvedi S, Rajasekar E, Natarajan S, McCullen N. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. Energy Policy. 2022;168(August 2017):113097. doi:10.1016/j.enpol.2022.113097.
12. Zhao D, Zhang R, Zhang H, He S. Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models. Sci Rep. 2022;12(1):18138. doi:10.1038/s41598-022-23154-4.
13. Basak A, Rahman ATMS, Das J, Hosono T, Kisi O. Drought forecasting using the Prophet model in a semi-arid climate region of western India. Hydrol Sci J. 2022;67(9):1397–417. doi:10.1080/02626667.2022.2082876.
14. Shakeel A, Chong D, Wang J. Load forecasting of district heating system based on improved FB-Prophet model. Energy. 2023;278(2):127637. doi:10.1016/j.energy.2023.127637.
15. Ma P, Ren J, Sun G, Zhao H, Jia X, Yan Y, et al. Multiscale superpixelwise prophet model for noise-robust feature extraction in hyperspectral images. IEEE Trans Geosci Remote Sens. 2023;61:5508912. doi:10.1109/TGRS.2023.3260634.
16. Sardar I, Akbar MA, Leiva V, Alsanad A, Mishra P. Machine learning and automatic ARIMA/prophet models-based forecasting of COVID-19: methodology, evaluation, and case study in SAARC countries. Stoch Environ Res Risk Assess. 2023;37(1):345–59. doi:10.1007/s00477-022-02307-x.
17. Elneel L, Zitouni MS, Mukhtar H, Al-Ahmad H. Examining sea levels forecasting using autoregressive and prophet models. Sci Rep. 2024;14(1):14337. doi:10.1038/s41598-024-65184-0.
18. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
19. Molnar C. Interpretable machine learning: a guide for making black box models explainable. Lulu Com. 2020;23(1):92–3. doi:10.1177/09726225241252009.
20. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005;67(2):301–20. doi:10.1111/j.1467-9868.2005.00503.x.