



# Kahramanmaraş Sutcu Imam University

## Journal of Engineering Sciences



Geliş Tarihi : 04.11.2025  
Kabul Tarihi : 05.02.2026

Received Date : 04.11.2025  
Accepted Date : 05.02.2026

### PREDICTIVE MODELING IN ELECTROMOBILITY: A TIME SERIES ANALYSIS

### ELEKTROMOBİLİTEDE TAHMİNE DAYALI MODELLEME: BİR ZAMAN SERİSİ ANALİZİ

*Rana BEDİR URFALI*<sup>1\*</sup> (ORCID: 0009-0003-9495-2482)  
*Ersin KAYA*<sup>2</sup> (ORCID: 0000-0001-5668-5078)

<sup>1</sup> Konya Technical University, Department of Computer Engineering, Konya, Turkey  
<sup>2</sup> Konya Technical University, Department of Computer Engineering, Konya, Turkey

\*Sorumlu Yazar / Corresponding Author: Rana BEDİR URFALI, ranabedirr@gmail.com

#### ABSTRACT

This study analyzes electric bicycle usage in a province with time series models to forecast future trends. Electric bicycles, as an essential element of sustainable transportation, enhance urban mobility and reduce traffic congestion. Accurate demand forecasting is crucial for the operational planning and administration of e-bike sharing systems. In this study, forecasts were first made using only bicycle usage data with univariate time series models (ARIMA, Prophet, SSA). Next, multivariate approaches (SARIMAX, Multivariate Prophet, MSSA) incorporated external factors such as temperature, precipitation, and wind speed. The dataset, therefore, combined bicycle usage records with meteorological variables to capture environmental impacts on demand. Model performance was assessed using RMSE and MAE metrics. Results showed regional variations in accuracy: ARIMA performed best among univariate models in four regions, while SARIMAX and Multivariate Prophet produced superior forecasts in most regions. Furthermore, MSSA consistently outperformed SSA in 13 regions, highlighting the benefit of including external influences. Overall, the findings demonstrate that integrating weather data improves forecast precision and supports better operational strategies. This contributes to optimizing e-bike sharing services and urban transport planning, enabling more efficient resource use and greater user satisfaction.

**Keywords:** electric bicycle, time series forecasting, univariate models, multivariate models, comparative analysis

#### ÖZET

Bu çalışma, gelecekteki eğilimleri tahmin etmek amacıyla, bir ildeki elektrikli bisiklet kullanımını zaman serisi modelleri ile analiz etmektedir. Sürdürülebilir ulaşımın kritik bir parçası olan elektrikli bisikletler, kentsel hareketliliği ve talep tahminini iyileştirir; bu da elektrikli bisiklet paylaşım sistemleri yönetimi için hayati önem taşır. Tahminler ilk olarak yalnızca bisiklet kullanım verileriyle tek değişkenli modeller (ARIMA, Prophet, SSA) kullanılarak yapıldı. Ardından, sıcaklık, yağış ve rüzgar hızı gibi dış etkenleri dahil eden çok değişkenli yaklaşımlara (SARIMAX, Multivariate Prophet, MSSA) geçildi. Bu sayede veri seti, çevresel etkileri yakalamak için meteorolojik değişkenlerle birleştirildi. Model performansı RMSE ve MAE metrikleriyle değerlendirildi. Sonuçlar, doğrulukta bölgesel farklılıklar olduğunu gösterdi: ARIMA, tek değişkenli modeller arasında dört bölgede en iyisiydi, ancak SARIMAX ve Multivariate Prophet çoğu bölgede üstün tahminler üretti. Ayrıca MSSA, dış etkileri dahil etmenin faydasını vurgulayarak 13 bölgede SSA'yı sürekli olarak geride bıraktı. Genel bulgular, hava durumu verilerini entegre etmenin tahmin hassasiyetini artırdığını ve daha iyi operasyonel stratejileri desteklediğini ortaya koymaktadır. Bu yaklaşım, elektrikli bisiklet paylaşım hizmetlerini ve şehir içi ulaşım planlamasını optimize ederek daha verimli kaynak kullanımına ve yüksek kullanıcı memnuniyetine katkıda bulunmaktadır.

**Anahtar Kelimeler:** elektrikli bisiklet, zaman serisi tahmini, tek değişkenli modeller, çok değişkenli modeller, karşılaştırmalı analiz

## INTRODUCTION

Global climate change and increasing urbanization rates are driving a growing need for sustainable transportation solutions. In this context, e-bikes are emerging as a prominent alternative to traditional transportation systems and are undertaking a critical role in the transformation of urban mobility (Vishnu et al., 2023). With zero emissions, low operating costs, and the potential to reduce traffic congestion, e-bikes have become a cornerstone of sustainable urban transport (Swari et al., 2025). Over the past decade, e-bike sharing systems have shown exponential growth on a global scale and have become an integral part of sustainable urban transportation ecosystems (İnanç, 2024). The operational success of these systems is fundamentally based on accurate demand forecasting and effective resource allocation strategies. Critical decisions such as optimizing bicycle distribution, planning station capacities, and scheduling maintenance operations all require a precise prediction of future usage patterns. Time series analysis offers a powerful statistical methodology for forecasting future values based on past data patterns. This approach has proven particularly effective in e-bike sharing demand forecasting and has provided valuable insights for operational decision-making processes. Univariate time series methodologies, including ARIMA (Fattah et al., 2018), Prophet (Zunic et al., 2020), and SSA (Wang et al., 2024), are extensively preferred in the field of transportation demand forecasting. However, bicycle usage demand is influenced by multiple external factors beyond past usage patterns. Meteorological variables like temperature, precipitation, and wind speed significantly impact user behavior and system utilization rates. For this reason, multivariate time series models such as SARIMAX (Alharbi & Csala, 2022), Multivariate Prophet (Almazrouee et al., 2020), and Multivariate SSA (Agarwal et al., 2022) have become popular in research for their ability to offer enhanced forecasting capabilities by incorporating these exogenous variables (Iftikhar et al., 2025). To address gaps in the existing literature and provide an applicable framework for e-mobility management, this study utilizes a comprehensive dataset from a private e-bike sharing provider. The main novel contributions of this paper are:

- Offers a comprehensive performance evaluation by pairing univariate models with their multivariate counterparts, explicitly quantifying the marginal contribution of exogenous meteorological variables on forecast accuracy.
- Unlike macro-level city analyses, this study utilizes a GIS-based approach to segment the study area into 14 functional regions, demonstrating that demand dynamics and weather sensitivity vary significantly based on regional characteristics (spatial heterogeneity).
- Validates the computational speed and robustness of classical time series models in a real-time setting, proposing an agile, low-cost, and actionable forecasting framework for sustainable fleet management and resource optimization.

The organization of the paper is as follows: In Section II, an overview of research focused on forecasting through time series models is provided. Section III outlines the methodologies and materials employed in the study. In Section IV, the experimental findings derived from the models are detailed, along with their corresponding analysis. Finally, Section V offers the paper's conclusions.

## RELATED WORK

By analyzing historical data, time series analysis provides a highly effective method for forecasting future trends. In urban transportation, these forecasts are strategic for planning and managing micromobility solutions. Furthermore, establishing sustainable urban policies is critically dependent on accurately forecasting the adoption of environmentally friendly transport like electric bicycles. This section, therefore, explores the application of both univariate and multivariate time series models for this purpose.

In their 2023 research, Subramanian et al. undertook a comparative examination aimed at predicting bicycle demand, leveraging a two-year dataset comprised of hourly usage statistics from the Capital Bikeshare system situated in Washington, D.C. The study juxtaposed the efficacy of the Random Forest algorithm with both conventional time series approaches (ARIMA, SARIMA) and advanced deep learning models (LSTM, GRU). To refine the predictive capabilities of the models, a comprehensive array of exogenous factors was included, such as environmental conditions (temperature, precipitation, and wind speed), along with temporal and categorical elements like holidays, time of day, month, and season. The evaluation of model performance was conducted through the application of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) metrics. The results indicated that the GRU model achieved the most advantageous outcomes (MSE: 2641.24, RMSE: 51.39, MAE: 30.76), reflecting the lowest error measures. The authors credited the effectiveness of the GRU model to its

ability to capture long-term dependencies within the evolving demand patterns over time. It is noteworthy that the Random Forest model also displayed a commendable level of accuracy (MAE: 30.36), demonstrating strong performance, particularly in processing both numerical and categorical data. In contrast, the ARIMA and SARIMA models underperformed due to their fundamental linear assumptions. This study highlights the critical role of precise demand forecasting in bikeshare systems, facilitating optimal operational planning, effective resource management, and the promotion of sustainable transportation initiatives.

Sanami et al. (2025) focused on demand forecasting for Electric Vehicle (EV) charging stations. The researchers developed a multivariate LSTM model augmented with an attention mechanism (LSTM-ATT) to predict charging demand for the following day at 15-minute intervals. For model training, time series data incorporating exogenous factors, including weather conditions, day, month, and holiday information, were employed. To understand the influence of each variable on the predictions, explainable artificial intelligence (XAI) techniques such as SHAP (Shapley Additive Explanations) were applied. The evaluation on test data demonstrated low error rates. The proposed multivariate LSTM-ATT model obtained a test loss (MSE) of 0.0085, which represents a substantially lower value compared to the univariate LSTM (MSE: 0.2124) and attention-based univariate LSTM (MSE: 0.2014) models. The training duration for the model was recorded as 12 minutes and 24 seconds. According to SHAP analysis, historical demand data was identified as the most influential variable, while the effects of factors such as temperature and month were observed to vary depending on the season. Time periods including holidays and weekends were found to have a notable impact on demand. This research contributes to the field by offering highly precise predictions for planning electric vehicle charging infrastructure, thereby enhancing the efficiency of operational decisions such as resource allocation, maintenance scheduling, and dynamic pricing. Additionally, through visualization of the attention mechanism, the model's interpretability was strengthened by demonstrating which time intervals were more influential for predictions (Sanami et al., 2025).

In a comparative analysis conducted by Kwarteng & Andreevich (2024), ARIMA, SARIMA, and Prophet models were employed to predict antimalarial drug demand in Australia. The dataset utilized in this research comprised time series information gathered from Medicare, Australia, spanning the period from 1991 to 2008. Considering the existence of notable seasonal patterns and trend components in the data, both conventional and contemporary time series forecasting methodologies were evaluated. The assessment of model accuracy was performed using three key error indicators: MAE, RMSE, and MAPE. The outcomes revealed that the Prophet model demonstrated successful predictions with low error metrics (MAE: 0.74, MAPE: 8.2%), along with RMSE values. The SARIMA model ranked second in terms of performance (MAE: 2.18, MAPE: 10.04%), whereas the ARIMA model displayed the weakest results (MAE: 3.02, MAPE: 13.62%). The findings indicated that the Prophet model particularly excels in accurately capturing seasonal variations, holiday influences, and non-linear patterns. The research also included a graphical comparison illustrating the Prophet model's prediction capabilities and actual observations. The Prophet model exhibited a strong ability to generate accurate forecasts even during periods characterized by sudden changes, including holiday seasons. The predictions generated by the model demonstrated a high level of agreement with the actual data distribution, thereby validating its utility as an effective instrument for operational planning and resource optimization. The conclusions of this investigation confirm the advantage of the Prophet model in forecasting time series data, surpassing both ARIMA and SARIMA approaches. The research further recommends that the Prophet model represents the optimal selection, especially when dealing with intricate, seasonal datasets (Kwarteng & Andreevich, 2024).

Jaber et al. (2022) performed long-term time series predictions utilizing the bikeshare system in Budapest. The research applied SARIMA (Seasonal Auto-Regressive Integrated Moving Average), Artificial Neural Networks (ANN), and Exponential Smoothing (ETS) approaches to examine user behavior and its contribution to service planning. Through comprehensive visual and statistical examination, it was identified that summer months and, specifically, weekends represented the time periods with the most substantial influence on bicycle usage. The examination employing the SARIMA approach indicated that robust predictions could be achieved with an  $R^2$  coefficient of 96.4% and an RMSE of 74.06 for three time intervals. When comparing the  $R^2$  coefficients across the three models (SARIMA, ANN, and ETS), it was found that both approaches yielded comparable results, with ANN achieving an  $R^2$  value of approximately 96.9%. Nevertheless, a subsequent examination demonstrated that the SARIMA approach outperformed others in terms of RMSE metrics. In contrast, the ETS approach exhibited a comparatively weaker performance (RMSE: 87.93). Additionally, the influence of the pandemic on the bikeshare infrastructure was investigated across three separate temporal phases: pre-pandemic, during the pandemic, and post-pandemic periods. The examination uncovered a notable rise in bicycle usage throughout the post-pandemic phase,

with utilization moving toward weekdays. This investigation establishes the utility of time series examination in demand prediction for bikeshare infrastructures, delivering crucial insights for decision-makers (Jaber et al., 2022).

Zhang et al. (2017) examined Electric Vehicle (EV) sales in China by implementing both univariate and multivariate time series approaches for prediction purposes. The research utilized Singular Spectrum Analysis (SSA) as the single-variable approach and the Vector Autoregressive (VAR) framework as the multi-variable approach. The SSA framework exhibited a high level of precision in identifying patterns within BEV (Battery Electric Vehicle) and PHEV (Plug-in Hybrid Electric Vehicle) sales data, reaching an accuracy level of 36.40% MAPE and 21.5 RMSE. The VAR framework, enhanced by economic variables (including consumer price index, consumer confidence index, producer price index, fuel price, car price, and Baidu search data), attained an improved accuracy level with 29.1% MAPE and 16.31 RMSE. For short-term predictions, the examination using SSA and VAR frameworks projected 258394 BEV and 72957 PHEV sales in China by the conclusion of 2017. Regarding long-term projections, BEV sales were anticipated to achieve 350000 units and PHEV sales 720000 units by 2020. The VAR framework was validated to deliver more dependable and comprehensive predictions through the incorporation of economic variables combined with Baidu search data, which reflects consumer interest levels. The investigation discussed emphasizes the development pattern of the EV sector and provides essential resources for strategic planning by vehicle manufacturers and decision-makers. Within the framework of emerging markets such as China, the SSA approach offers a practical solution demanding limited data input, whereas the VAR approach appears more appropriate for performing in-depth examinations (Zhang et al., 2017).

In a distinct investigation, Inanç (2024) employed Machine Learning (ML) methodologies to forecast the trip durations of electric scooter (e-scooter) riders. The research conducted a comparative application of AdaBoost (AB), Random Forest (RF), and Gradient Boosting (GB) techniques. Trip duration predictions were generated using independent factors, including the rider's age, gender, riding experience, and distance. The AdaBoost (AB) approach produced the most favorable outcomes, achieving a Mean Squared Error of 0.005, Root Mean Squared Error of 0.069, and Mean Absolute Error of 0.039. Additionally, the  $R^2$  coefficient was computed as 0.947, indicating strong predictive accuracy. While the RF and GB techniques also yielded favorable outcomes, they demonstrated elevated error metrics in comparison to the AB approach (RF: RMSE = 0.086, MAE = 0.052,  $R^2$  = 0.917; GB: RMSE = 0.084, MAE = 0.049,  $R^2$  = 0.921). The investigation further examined the influence of variables through regression analysis; findings revealed that distance, riding experience, and gender factors exerted a notable influence on trip duration, whereas age did not. The Low Variance Inflation Factor (VIF) measurements confirmed the lack of multicollinearity concerns. The outcomes of this research indicate that e-scooter service operators can enhance their operational planning effectiveness by generating more accurate forecasts regarding trip durations utilizing these performance indicators. Furthermore, it was highlighted that Machine Learning (ML) methodologies serve as a powerful instrument within transportation infrastructures for duration estimation, service quality enhancement, and resource optimization (Inanç, 2024).

Komarica et al. (2024) examined the determinants influencing the adoption willingness of electric bicycles (e-bikes) within Belgrade city through the implementation of artificial neural network (ANN) approaches. The research conducted a comparative analysis utilizing multilayer perceptron (MLP) and radial basis function (RBF) frameworks, with model training performed using information collected from 626 respondents via an online questionnaire. The MLP framework exhibited superior predictive capability in comparison to the RBF framework, attaining 88.7% accuracy, 95.0% recall, 91.3% precision, 0.931 F1 score, and an ROC area of 0.927. Conversely, the RBF framework demonstrated reduced performance with 80.5% accuracy and 0.897 ROC area. The analytical outcomes indicated that the determinants exerting the strongest influence on e-bike adoption willingness include: commuting distance (NI = 100%), frequency of motorcycle commuting (NI = 97%), use of motorcycles for alternative purposes (NI = 85.3%), travel distance for alternative purposes (NI = 83.2%), perception of air pollution in Belgrade (NI = 73.8%), and participant age (NI = 72.6%). Moreover, 69.8% of respondents expressed their readiness to embrace e-bike usage as a strategy to alleviate traffic congestion and environmental pollution. These outcomes offer decision-makers a significant data foundation for formulating sustainable transportation strategies in Belgrade. The investigation also establishes that artificial neural network approaches possess superior predictive capabilities compared to conventional logit frameworks in characterizing transportation behavior patterns (Komarica et al., 2024).

The literature reviewed above demonstrates the extensive application of various forecasting techniques, ranging from classical statistical methods to advanced deep learning architectures, in the domain of transportation. While deep learning models often achieve high accuracy, they require significant computational resources and large datasets.

Conversely, classical statistical models offer distinct advantages in terms of interpretability and operational speed, yet their comparative performance specifically regarding the marginal contribution of exogenous variables in spatially granular contexts remains less explored in micromobility research. Addressing this gap, this study contributes to the literature by systematically benchmarking univariate versus multivariate statistical models (specifically introducing MSSA to this domain) across 14 distinct functional regions, thereby validating an agile and weather-aware forecasting framework suitable for real-time fleet management.

## MATERIAL AND METHOD

This section details the methods used to analyze urban electric bicycle usage and forecast future usage trends. In this paper, time series analysis and statistical models were used to create forecasting models from both univariate and multivariate datasets. The approach that produces the most optimal forecasts is determined by considering model performance metrics.

### *Dataset Description*

In this paper, two distinct datasets were utilized to construct the forecasting models. The primary dataset consists of electric bicycle usage statistics obtained from a private company providing e-bike sharing services within the study area. Due to data privacy agreements and commercial confidentiality, the service provider's name is not disclosed. The raw operational data were aggregated into daily statistics for research purposes, covering a period of approximately 14 months from December 31, 2022, to February 28, 2024.

The bicycle usage dataset is characterized by three fundamental attributes. The first attribute is the date of observation, collected at a daily frequency; the date format follows the ISO 8601 standard and serves as the primary time index. The second attribute is a categorical variable indicating the region (coded as "Region 1" through "Region 14"), representing usage activity within specific geographical boundaries. The third attribute is the numerical target variable, representing the total number of bicycle rides performed in the relevant region on the respective date. To capture the spatial heterogeneity of demand, the study area was segmented into 14 distinct functional regions using a GIS-based approach (Figure 1) (Taoussi et al., 2025; Chowdhury & Hafsa, 2022). Based on the socio-demographic and infrastructural characteristics identified in the preliminary analysis, these regions are broadly categorized into four functional groups:

- Central Business Districts (Regions 9, 12, 14): Areas characterized by high commercial density and systematic "last-mile" commuting patterns.
- University Zones (Regions 4, 6): Areas driven by the student population and dependent on the academic calendar.
- Recreation and Tourism Areas (Regions 3, 7, 10): Zones dominated by leisure-oriented usage, which are highly sensitive to environmental conditions.
- Peripheral and Industrial Zones (Regions 1, 2, 13): Areas associated with mandatory workforce transport and lower variation. This regional division is critical for analyzing the impact of spatial differences on forecasting performance.

The second dataset, utilized for multivariate models, integrates meteorological information with the usage data. Weather data were retrieved via the Meteostat platform, which aggregates historical data from calibrated stations in accordance with World Meteorological Organization standards. This dataset includes daily values for average temperature (tavg), total precipitation (prcp), and wind speed (wspd). Specifically, wind speed data were retrieved in km/h via the API and converted to m/s for consistency in the analysis. Both datasets cover the identical time period and were synchronized to evaluate the impact of environmental factors on demand.

Additionally, the time series components were analyzed across the entire dataset. This involved decomposing the observed values into their constituent components: trend, seasonal, and residual. Figure 2 displays this analysis, which provides a better understanding of the data structure by revealing long-term trends, seasonal cycles, and random fluctuations (Kwarteng & Andreevich, 2024).

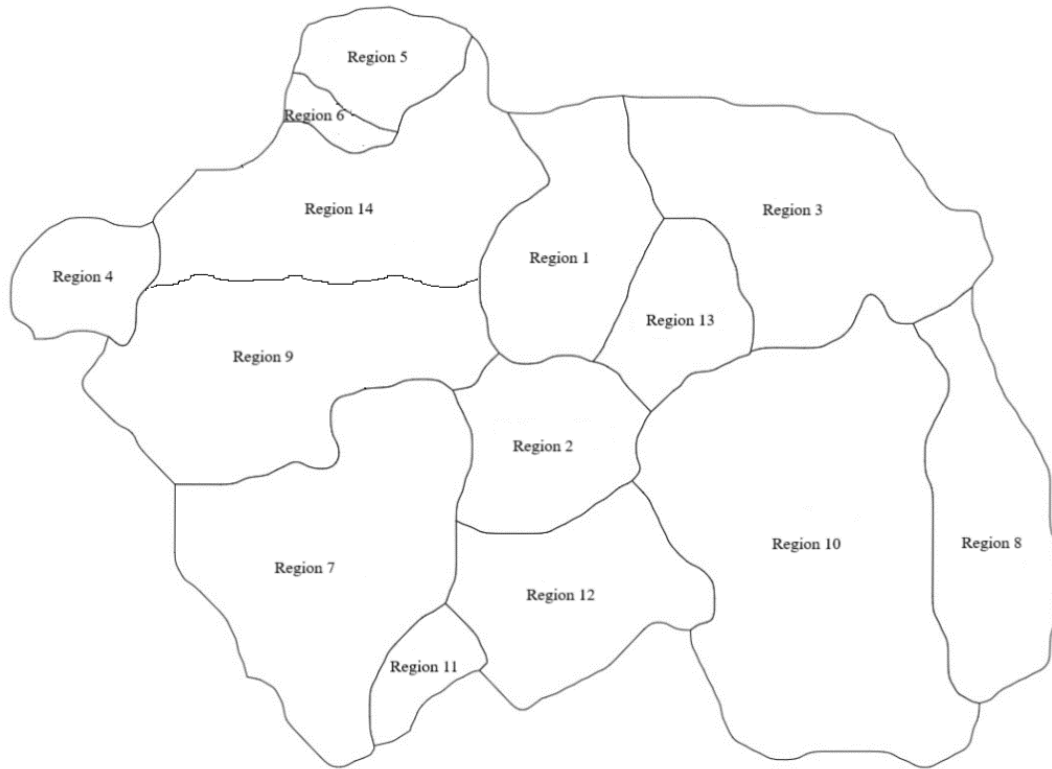


Figure 1. Regional Segmentation of The Study Area Using a GIS-Based Approach

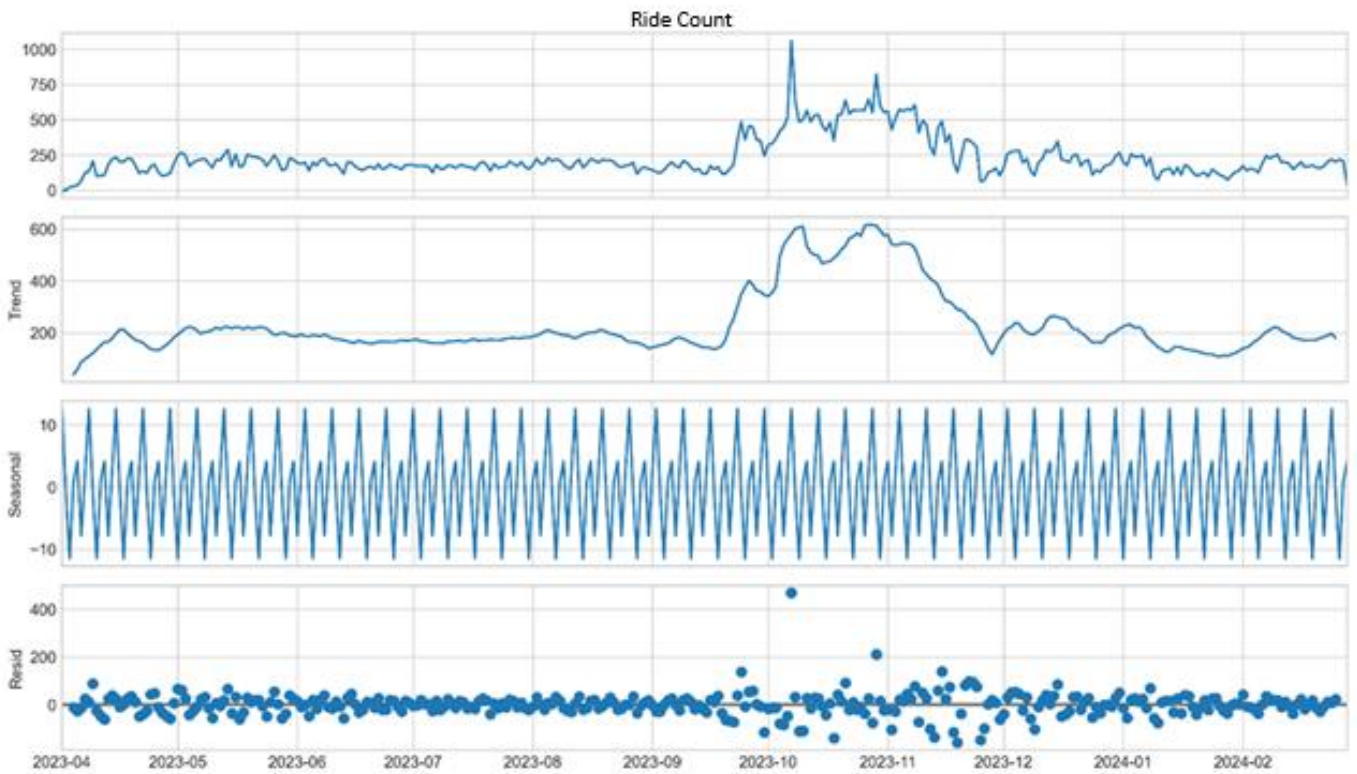


Figure 2. Time Series Decomposition of Electric Bicycle Usage (April 2023 – February 2024)

### Data Preprocessing and Experimental Setup

To ensure the reproducibility of the proposed models and the validity of the comparative analysis, a rigorous data preprocessing protocol and a standardized experimental framework were followed throughout all processes. The

analysis process was initiated with the aggregation of raw operational data to a daily frequency to align with the research objectives. To preserve the temporal continuity of the dataset, missing values resulting from system maintenance or connectivity issues were handled using the linear interpolation method, while outliers caused by non-routine events were smoothed to prevent bias in linear models.

Within the scope of stationarity analysis, which constitutes the most critical step of the data preparation phase, Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests were conducted for each of the 14 regions (Afriyie et al., 2020). According to the ADF test results presented in Table 1, the test statistics for the original series ranged between -1.41 and -2.49, while the p-values were consistently observed to be above the 0.05 significance level ( $p > 0.05$ ), indicating that the unit root hypothesis could not be rejected.

**Table 1.** Comparative ADF Test Results for All Regional Data

Regions	Test Statistic	p-value	Results
Region 1 (Original)	-1.53	0.519	Non-Stationary
Region 1 (Differenced)	-5.78	< 0.001	Stationary
Region 2 (Original)	-1.89	0.334	Non-Stationary
Region 2 (Differenced)	-6.12	< 0.001	Stationary
Region 3 (Original)	-1.41	0.576	Non-Stationary
Region 3 (Differenced)	-4.95	< 0.001	Stationary
Region 4 (Original)	-2.13	0.231	Non-Stationary
Region 4 (Differenced)	-5.61	< 0.001	Stationary
Region 5 (Original)	-1.97	0.298	Non-Stationary
Region 5 (Differenced)	-5.89	< 0.001	Stationary
Region 6 (Original)	-2.01	0.281	Non-Stationary
Region 6 (Differenced)	-6.32	< 0.001	Stationary
Region 7 (Original)	-1.75	0.404	Non-Stationary
Region 7 (Differenced)	-5.99	< 0.001	Stationary
Region 8 (Original)	-2.11	0.240	Non-Stationary
Region 8 (Differenced)	-6.04	< 0.001	Stationary
Region 9 (Original)	-2.05	0.264	Non-Stationary
Region 9 (Differenced)	-5.92	< 0.001	Stationary
Region 10 (Original)	-2.27	0.182	Non-Stationary
Region 10 (Differenced)	-4.73	< 0.001	Stationary
Region 11 (Original)	-2.06	0.258	Non-Stationary
Region 11 (Differenced)	-5.19	< 0.001	Stationary
Region 12 (Original)	-2.49	0.117	Non-Stationary
Region 12 (Differenced)	-5.32	< 0.001	Stationary
Region 13 (Original)	-1.91	0.320	Non-Stationary
Region 13 (Differenced)	-5.02	< 0.001	Stationary
Region 14 (Original)	-2.31	0.169	Non-Stationary
Region 14(Differenced)	-5.25	< 0.001	Stationary

Similarly, upon examining the KPSS test results summarized in Table 2, it was determined that the test statistics for the original series varied between 0.490 and 1.973 ( $p \leq 0.046$ ), and the stationarity hypothesis was rejected. Consequently, a first-order differencing ( $d = 1$ ) transformation was applied to all series prior to proceeding to the modeling phase. Re-tests conducted after this operation confirmed that stationarity was achieved; the ADF statistics for the differenced series dropped to a range of -4.73 to -6.32 ( $p < 0.001$ ), while the KPSS statistics stabilized between 0.081 and 0.134 ( $p = 0.100$ ).

A two-stage integrated strategy was adopted to evaluate the predictive performance of the models and demonstrate their operational feasibility. In the first stage, performance validation, a chronological data splitting strategy was employed to ensure a fair comparison across all models. In this context, the first 80% of the dataset was utilized for model training and hyperparameter optimization, while the remaining 20% was reserved as a test set to measure performance on unseen data using RMSE and MAE metrics.

Optimal parameter sets were determined via grid search for parametric models and through iterative experimentation for non-parametric approaches. Following the validation phase, the study advanced to the operational forecasting stage to demonstrate the value of the proposed framework in a real-world scenario. In this phase, the validated models

were retrained on the entire dataset to generate future forecasts for a 7-day horizon following the last observation date.

**Table 2.** KPSS Test Results by Region

Regions	Test Statistic	p-value	Results
Region 1 (Original)	0.490	0.046	Non-Stationary
Region 1 (Differenced)	0.088	0.100	Stationary
Region 2 (Original)	0.493	0.046	Non-Stationary
Region 2 (Differenced)	0.091	0.100	Stationary
Region 3 (Original)	0.540	0.046	Non-Stationary
Region 3 (Differenced)	0.081	0.100	Stationary
Region 4 (Original)	0.498	0.046	Non-Stationary
Region 4 (Differenced)	0.097	0.100	Stationary
Region 5 (Original)	0.492	0.043	Non-Stationary
Region 5 (Differenced)	0.083	0.100	Stationary
Region 6 (Original)	1.080	0.010	Non-Stationary
Region 6 (Differenced)	0.103	0.100	Stationary
Region 7 (Original)	0.490	0.046	Non-Stationary
Region 7 (Differenced)	0.090	0.100	Stationary
Region 8 (Original)	1.932	0.010	Non-Stationary
Region 8 (Differenced)	0.111	0.100	Stationary
Region 9 (Original)	0.493	0.046	Non-Stationary
Region 9 (Differenced)	0.104	0.100	Stationary
Region 10 (Original)	1.973	0.010	Non-Stationary
Region 10 (Differenced)	0.125	0.100	Stationary
Region 11 (Original)	0.492	0.043	Non-Stationary
Region 11 (Differenced)	0.093	0.100	Stationary
Region 12 (Original)	1.080	0.010	Non-Stationary
Region 12 (Differenced)	0.117	0.100	Stationary
Region 13 (Original)	1.839	0.010	Non-Stationary
Region 13 (Differenced)	0.129	0.100	Stationary
Region 14 (Original)	1.876	0.010	Non-Stationary
Region 14 (Differenced)	0.134	0.100	Stationary

### **Time Series Forecasting Models**

Time series forecasting represents an analytical approach that utilizes past data to project future values. Models within this domain are categorized into two primary classifications: univariate and multivariate, determined by the data structure and forecasting objectives.

#### **Univariate Time Series Forecasting Models**

Univariate models generate future value predictions using historical values of a single time series.

#### **ARIMA Model**

The ARIMA (Autoregressive Integrated Moving Average) model represents one of the most extensively applied statistical approaches for time series prediction. The AR component captures past observations; the MA component accounts for past error terms; and the I component addresses the differencing operation required to achieve series stationarity (Cocca et al., 2020; Alencar et al., 2021). The fundamental ARIMA model structure can be formulated as follows:

$$\hat{y}_t = c + \phi_1 \hat{y}_{t-1} + \dots + \phi_p \hat{y}_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (1)$$

It can be seen from Equation (1) that the differenced series is represented by  $\hat{y}_t$  while  $\phi$  and  $\theta$  are the model parameters, and  $\epsilon_t$  denotes the error term. In this investigation, ARIMA models were developed individually for each region, with hyperparameters refined according to the stationarity and seasonal properties of the corresponding time series. The refined hyperparameters for the 14 regions are displayed in Table 3.

**Table 3.** ARIMA Model Hyperparameters by Region

Regions	ARIMA Hyperparameters
Region_1	(p = 1, d = 1, q = 1, P = 0, D = 0, Q = 1, s = 7)
Region_2	(p = 0, d = 1, q = 2, P = 2, D = 0, Q = 0, s = 7)
Region_3	(p = 1, d = 1, q = 2, P = 0, D = 0, Q = 0, s = 7)
Region_4	(p = 1, d = 0, q = 1, P = 0, D = 0, Q = 0, s = 7)
Region_5	(p = 2, d = 0, q = 1, P = 1, D = 0, Q = 0, s = 7)
Region_6	(p = 2, d = 0, q = 0, P = 0, D = 0, Q = 0, s = 7)
Region_7	(p = 2, d = 0, q = 1, P = 2, D = 0, Q = 2, s = 7)
Region_8	(p = 0, d = 1, q = 2, P = 2, D = 0, Q = 0, s = 7)
Region_9	(p = 1, d = 0, q = 2, P = 0, D = 0, Q = 0, s = 7)
Region_10	(p = 1, d = 1, q = 1, P = 0, D = 0, Q = 0, s = 7)
Region_11	(p = 0, d = 1, q = 1, P = 0, D = 0, Q = 0, s = 7)
Region_12	(p = 1, d = 1, q = 1, P = 1, D = 0, Q = 0, s = 7)
Region_13	(p = 5, d = 1, q = 1, P = 0, D = 0, Q = 0, s = 7)
Region_14	(p = 2, d = 1, q = 1, P = 1, D = 0, Q = 0, s = 7)

These configurations were selected to minimize forecasting errors, with RMSE and MAE metrics guiding the optimization process.

### ***Prophet Model***

The Prophet model processes a time series through a combination of three primary elements: trend ( $g(t)$ ), seasonality ( $s(t)$ ), and holidays ( $h(t)$ ) (Alencar et al., 2021). The fundamental framework of this model maintains a straightforward structure and can be expressed through the following formula:

$$y_t = g(t) + s(t) + h(t) + \epsilon_t \quad (2)$$

Examination of Equation (2) reveals that the predicted value  $y(t)$  constitutes a linear combination of a linear or logistic trend function  $g(t)$ , periodic seasonal variations  $s(t)$ , holiday influences  $h(t)$ , and the error term  $\epsilon_t$  (Alencar et al., 2021).

In this investigation, the Prophet model was implemented individually for each region, with hyperparameters calibrated to capture the structural properties of the corresponding time series.

The model's capability in managing seasonality and trend elements was utilized by modifying the changepoint prior scale, seasonality prior scale, and seasonality mode parameters. The calibrated hyperparameters for the 14 regions are displayed in Table 4.

These parameter settings were selected to minimize forecasting errors, with RMSE and MAE metrics guiding the optimization process.

### ***SSA Model***

SSA (Singular Spectrum Analysis) breaks down time series data into its fundamental elements. These elements capture the primary structures, including trend, periodic cycles, and noise. This approach executes an Eigenvalue Decomposition (EVD) to isolate the principal elements of a time series. The fundamental process entails identifying the eigenvectors and eigenvalues of a covariance matrix derived from a stacked version of the series (Patterson et al., 2011). In this research, SSA was implemented on an individual basis for each region, with the window length ( $L$ )

and the number of groups calibrated to optimally capture the underlying structure of the corresponding time series. The calibrated hyperparameters for the 14 regions are displayed in Table 5.

**Table 4. Prophet Model Hyperparameters by Region**

<b>Regions</b>	<b>Prophet Hyperparameters</b>
Region_1	(ChangepointPriorScale = 0.001, SeasonalityPriorScale = 0.01, SeasonalityMode = additive)
Region_2	(ChangepointPriorScale = 0.01, SeasonalityPriorScale = 0.5, SeasonalityMode = multiplicative)
Region_3	(ChangepointPriorScale = 0.5, SeasonalityPriorScale = 0.1, SeasonalityMode = multiplicative)
Region_4	(ChangepointPriorScale = 0.01, SeasonalityPriorScale = 0.2 SeasonalityMode = additive)
Region_5	(ChangepointPriorScale = 0.01, SeasonalityPriorScale = 0.01, SeasonalityMode = additive)
Region_6	(ChangepointPriorScale = 0.5, SeasonalityPriorScale = 10, SeasonalityMode = multiplicative)
Region_7	(ChangepointPriorScale = 0.2, SeasonalityPriorScale = 0.01, SeasonalityMode = multiplicative)
Region_8	(ChangepointPriorScale = 0.1, SeasonalityPriorScale = 10, SeasonalityMode = additive)
Region_9	(ChangepointPriorScale = 0.05, SeasonalityPriorScale = 0.2, SeasonalityMode = additive)
Region_10	(ChangepointPriorScale = 0.1, SeasonalityPriorScale = 0.01, SeasonalityMode = multiplicative)
Region_11	(ChangepointPriorScale = 0.01, SeasonalityPriorScale = 0.01, SeasonalityMode = additive)
Region_12	(ChangepointPriorScale = 0.01, SeasonalityPriorScale = 0.1, SeasonalityMode = additive)
Region_13	(ChangepointPriorScale = 0.001, SeasonalityPriorScale = 0.5, SeasonalityMode = additive)
Region_14	(ChangepointPriorScale = 0.01, SeasonalityPriorScale = 0.2, SeasonalityMode = additive)

**Table 5. SSA Model Hyperparameters by Region**

<b>Regions</b>	<b>SSA Hyperparameters</b>
Region_1	(L = 5, groups = 3)
Region_2	(L = 17, groups = 8)
Region_3	(L = 17, groups=2)
Region_4	(L = 17, groups = 4)
Region_5	(L = 9, groups = 7)
Region_6	(L= 5, groups = 3)
Region_7	(L = 19, groups = 4)
Region_8	(L = 17, groups = 3)
Region_9	(L=13, groups = 5)
Region_10	(L = 5, groups = 4)
Region_11	(L = 11, groups = 7)
Region_12	(L = 15, groups = 8)
Region_13	(L = 19, groups = 8)
Region_14	(L = 17, groups = 2)

These parameter settings were determined through iterative experimentation, aiming to minimize forecasting errors and maximize the decomposition quality of the time series components.

### **Multivariate Time Series Forecasting Models**

Multivariate models improve forecast accuracy by using the relationships between multiple variables.

#### **SARIMAX Model**

SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors) represents an enhanced variant of the ARIMA model that incorporates exogenous variables ( $x_t$ ) and seasonal elements ( $(p_s, d_s, q_s)$ ). The overall structure of the SARIMAX model may be formulated as follows:

$$\Phi(B)\Phi_S(B^S)(1 - B)^d(1 - B^S)^{d_s}y_t = c + \Theta(B)\Theta_S(B^S)\epsilon_t + \beta X_t \quad (3)$$

Examination of Equation (3) reveals that  $\Phi$  and  $\Theta$  denote the autoregressive and moving average parameters,  $B$  represents the backshift operator,  $S$  indicates the seasonal period, and  $X_t$  refers to the exogenous variables (Alharbi & Csala, 2022).

In this study, SARIMAX models were constructed separately for each region, with model parameters and lag values for exogenous variables optimized through grid search. The autoregressive (p), differencing (d), and moving average (q) elements, in conjunction with their seasonal equivalents (P, D, Q) and seasonal period (s), were determined based on the statistical properties of each time series. (Alagade & Sahu, 2025).

Additionally, lag values for the exogenous variables average temperature (tavg), precipitation (prcp), and wind speed (wspd) were determined to enhance the explanatory power of the model. The optimized hyperparameters for the 14 regions are presented in Table 6.

**Table 6.** SARIMAX Model Hyperparameters by Region

Regions	Lag Values	SARIMAX Hyperparameters
Region_1	(tavg = 2, prcp = -4, wspd = 6)	(p = 0, d = 1, q = 2, P = 1, D = 1, Q = 0, s = 7)
Region_2	(tavg = -1, prcp = 7, wspd = -1)	(p = 2, d = 1, q = 0, P = 0, D = 0, Q = 0, s = 7)
Region_3	(tavg = -7, prcp = 7, wspd = 7)	(p = 0, d = 0, q = 1, P = 1, D = 0, Q = 0, s = 7)
Region_4	(tavg = -2, prcp = -7, wspd = -1)	(p = 0, d = 1, q = 0, P = 0, D = 0, Q = 1, s = 7)
Region_5	(tavg = -1, prcp = 1, wspd = -1)	(p = 1, d = 1, q = 0, P = 0, D = 0, Q = 1, s = 30)
Region_6	(tavg = 7, prcp = 7, wspd = -1)	(p = 1, d = 1, q = 0, P = 1, D = 0, Q = 1, s = 7)
Region_7	(tavg = -6, prcp = 7, wspd = 7)	(p = 2, d = 0, q = 0, P = 1, D = 0, Q = 1, s = 30)
Region_8	(tavg = -6, prcp = 7, wspd = 7)	(p = 2, d = 0, q = 2, P = 0, D = 0, Q = 0, s = 7)
Region_9	(tavg = 2, prcp = -4, wspd = -5)	(p = 0, d = 1, q = 0, P = 0, D = 1, Q = 1, s = 7)
Region_10	(tavg = -7, prcp = 1, wspd = 1)	(p = 1, d = 1, q = 0, P = 1, D = 0, Q = 1, s = 7)
Region_11	(tavg = 7, prcp = 1, wspd = -1)	(p = 1, d = 1, q = 0, P = 1, D = 0, Q = 0, s = 30)
Region_12	(tavg = 1, prcp = 6, wspd = 2)	(p = 2, d = 0, q = 1, P = 0, D = 1, Q = 1, s = 7)
Region_13	(tavg = -1, prcp = 1, wspd = 2)	(p = 1, d = 0, q = 1, P = 1, D = 1, Q = 1, s = 30)
Region_14	(tavg = 3, prcp = 6, wspd = 3)	(p = 0, d = 1, q = 0, P = 0, D = 1, Q = 1, s = 30)

These parameter configurations were selected to minimize forecasting errors, as evaluated by RMSE and MAE metrics, and to ensure that the SARIMAX model effectively captured both seasonal patterns and the influence of external climatic variables across different regions.

### Multivariate Prophet Model

Multivariate Prophet is a structure created by adding exogenous variables ( $x_t$ ) to the basic Prophet model. These additional variables are used together with the trend and seasonality components to increase the model's forecasting power (Jagannatham & Divya, 2021).

$$y_t = g(t) + s(t) + h(t) + \beta x_t + \epsilon_t \quad (4)$$

Examination of Equation (4) reveals that the term  $\beta x_t$  captures the influence of external variables such as weather conditions or other environmental factors. In this investigation, the Multivariate Prophet model was implemented on an individual basis for each region, with hyperparameters calibrated to capture the structural properties of the corresponding time series and the impact of exogenous variables.

The model's capability in managing seasonality and trend elements was utilized by modifying the changepoint prior scale, seasonality prior scale, and seasonality mode parameters. The calibrated hyperparameters for the 14 regions are displayed in Table 7.

**Table 7.** Multivariate Prophet Model Hyperparameters by Region

Regions	Multivariate Prophet Hyperparameters
Region_1	(changepoint_prior_scale = 0.001, seasonality_prior_scale = 1, seasonality_mode = additive)
Region_2	(changepoint_prior_scale = 0.001, seasonality_prior_scale = 0.01, seasonality_mode = multiplicative)
Region_3	(changepoint_prior_scale = 0.5, seasonality_prior_scale = 0.01, seasonality_mode = additive)
Region_4	(changepoint_prior_scale = 0.001, seasonality_prior_scale = 1, seasonality_mode = multiplicative)
Region_5	(changepoint_prior_scale = 0.001, seasonality_prior_scale = 1, seasonality_mode = additive)
Region_6	(changepoint_prior_scale = 0.5, seasonality_prior_scale = 0.5, seasonality_mode = multiplicative)
Region_7	(changepoint_prior_scale = 0.1, seasonality_prior_scale = 0.01, seasonality_mode = additive)
Region_8	(changepoint_prior_scale = 0.1, seasonality_prior_scale = 0.5, seasonality_mode = additive)
Region_9	(changepoint_prior_scale = 0.1, seasonality_prior_scale = 0.5, seasonality_mode = additive)
Region_10	(changepoint_prior_scale = 0.5, seasonality_prior_scale = 0.5, seasonality_mode = multiplicative)
Region_11	(changepoint_prior_scale = 0.01, seasonality_prior_scale = 1, seasonality_mode = additive)
Region_12	(changepoint_prior_scale = 0.5, seasonality_prior_scale = 0.01, seasonality_mode = additive)
Region_13	(changepoint_prior_scale = 0.5, seasonality_prior_scale = 0.1, seasonality_mode = additive)
Region_14	(changepoint_prior_scale = 0.001, seasonality_prior_scale = 1, seasonality_mode = additive)

These parameter configurations were chosen through iterative experimentation, with the objective of reducing forecasting errors and improving the model's adaptability to regional dynamics and external influences.

### **MSSA Model**

MSSA (Multivariate Singular Spectrum Analysis) constitutes the extension of SSA to multivariate data. The objective of MSSA is to identify common dynamics and hidden structures in multiple time series. This model functions similarly to the univariate SSA but employs a combined "caterpillar" matrix, constructed from all series, to extract the common structure of each one. Through the Eigenvalue Decomposition (EVD) applied to this matrix, the common components across multiple series are isolated (Patterson et al., 2011).

In this investigation, MSSA was implemented on an individual basis for each region, and two primary hyperparameters were refined: the window length (L), which establishes the dimension of the trajectory matrix employed in decomposition, and the number of groups, which determines how many components are preserved for reconstruction.

These values were selected to best capture the shared dynamics and structural patterns across multiple time series. The optimized hyperparameters for the 14 regions are presented in Table 8.

These parameter settings were determined through iterative experimentation, aiming to minimize forecasting errors and improve the decomposition quality of the shared components across the multivariate time series.

**Table 8.** MSSA Model Hyperparameters by Region

Regions	MSSA Hyperparameters
Region_1	(l_var = 34, group_var = 21)
Region_2	(l_var = 19, group_var = 17)
Region_3	(l_var = 39, group_var = 22)
Region_4	(l_var = 17, group_var = 6)
Region_5	(l_var = 10, group_var = 8)
Region_6	(l_var = 29, group_var = 14)
Region_7	(l_var = 18, group_var = 17)
Region_8	(l_var = 23, group_var = 16)
Region_9	(l_var = 24, group_var = 20)
Region_10	(l_var = 22, group_var = 20)
Region_11	(l_var = 21, group_var = 14)
Region_12	(l_var = 15, group_var = 8)
Region_13	(l_var = 20, group_var = 6)
Region_14	(l_var = 45, group_var = 6)

### Performance Metrics

In time series prediction, two primary evaluation measures are employed to assess model performance: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These measures offer a quantitative assessment of model accuracy by calculating how much the model's predictions differ from the observed values.

#### MAE

MAE is calculated as the mean of the absolute differences between the predicted values and the observed values. This measure indicates the typical magnitude of error within a set of forecasts without considering direction. By assigning equal weight to the magnitude of each error, it demonstrates a more robust resistance against outliers. The following formula is employed to calculate MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

In this equation,  $n$  denotes the count of predictions,  $y_i$  is the observed value, and  $\hat{y}_i$  constitutes the model's prediction. It can be seen from Equation (5) that equal weight is given to the magnitude of each error, which makes MAE more resilient to outliers.

Since the units of MAE correspond to the units of the forecast variable (e.g., dollars, units, degrees), it is typically regarded as the most comprehensible metric, as it directly indicates the typical error magnitude of the model.

#### RMSE

RMSE represents the square root of the mean of the squared differences between the predicted and observed values. The squaring of errors magnifies the influence of large errors, imposing greater penalties on them.

This makes RMSE a more useful metric in cases where large prediction errors are particularly important. The calculation of RMSE is performed with the formula in Equation (6).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

Examination of Equation (6) reveals that  $n$  denotes the count of predictions,  $y_i$  represents the observed value, and  $\hat{y}_i$  signifies the model's prediction.

A crucial aspect of comparison is that the RMSE value will consistently be equal to or exceed the MAE value ( $RMSE \geq MAE$ ) due to the squaring operation on errors. The greater the difference between RMSE and MAE, the larger the variance in the individual errors, indicating the presence of more significant outliers or large errors in the forecasts. Examining both metrics in combination provides a more thorough understanding of the model's error distribution.

## EXPERIMENTAL RESULTS AND DISCUSSION

This section offers a detailed examination of the forecasting performance of six time series models across 14 different regions for electric e-bike sharing demand prediction. The models were assessed using RMSE and MAE measures to deliver a thorough comparison of their predictive capabilities.

Table 9 presents the comparative performance between ARIMA and SARIMAX models across all regions. The results demonstrate varying effectiveness depending on the regional characteristics and data patterns.

**Table 9. ARIMA and SARIMAX Model Performance Comparison**

Regions	ARIMA RMSE	ARIMA MAE	SARIMAX RMSE	SARIMAX MAE
Region_1	<b>1.97</b>	<b>1.11</b>	2.92	2.67
Region_2	<b>3.25</b>	<b>2.44</b>	6.92	5.65
Region_3	10.37	8.32	<b>8.87</b>	<b>6.78</b>
Region_4	3.44	2.81	<b>2.96</b>	<b>2.50</b>
Region_5	5.25	4.27	<b>4.25</b>	<b>3.19</b>
Region_6	6.26	5.85	<b>5.44</b>	<b>4.97</b>
Region_7	10.04	8.76	<b>8.07</b>	<b>5.68</b>
Region_8	8.35	6.69	<b>7.57</b>	<b>6.16</b>
Region_9	1.15	0.81	<b>0.99</b>	<b>0.67</b>
Region_10	10.33	7.94	<b>9.69</b>	<b>6.64</b>
Region_11	8.21	6.78	<b>7.79</b>	<b>6.33</b>
Region_12	<b>1.71</b>	<b>1.45</b>	1.86	1.60
Region_13	<b>3.89</b>	<b>2.56</b>	4.75	3.25
Region_14	0.90	0.77	<b>0.77</b>	<b>0.69</b>

The analysis reveals that SARIMAX outperformed ARIMA in 10 out of 14 regions (71.4%), indicating the significant benefit of incorporating meteorological variables. ARIMA showed superior performance in only 4 regions (Regions 1, 2, 12, and 13), which suggests these regions may have more predictable patterns that are less influenced by external weather factors.

Table 10 compares the performance of Prophet and Multivariate Prophet models, highlighting the impact of incorporating exogenous variables into the Prophet framework.

Multivariate Prophet demonstrated superior performance in 12 out of 14 regions (85.7%), significantly outperforming the basic Prophet model. This significant enhancement suggests that the incorporation of meteorological variables strengthens the Prophet model's capability to represent intricate seasonal patterns and external influences on e-bike sharing demand.

Table 11 presents the comparison between SSA and MSSA models, showcasing the effectiveness of multivariate singular spectrum analysis.

**Table 10.** Prophet and Multivariate Prophet Model Performance Comparison

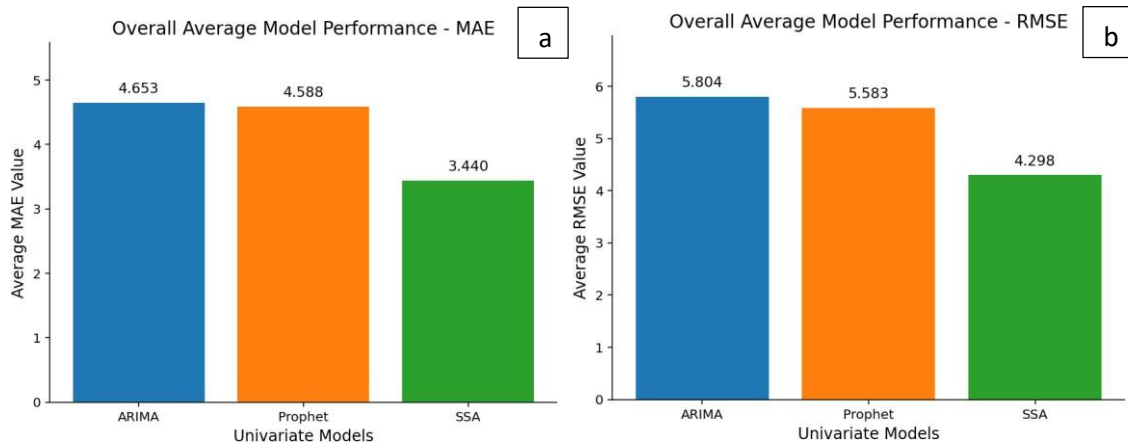
Regions	Prophet RMSE	Prophet MAE	Multivariate Prophet RMSE	Multivariate Prophet MAE
Region_1	<b>2.55</b>	<b>2.25</b>	2.58	2.27
Region_2	5.83	5.29	<b>5.54</b>	<b>4.93</b>
Region_3	8.40	6.66	<b>7.79</b>	<b>5.65</b>
Region_4	5.56	4.17	<b>4.88</b>	<b>3.66</b>
Region_5	6.86	5.96	<b>5.86</b>	<b>4.51</b>
Region_6	7.07	6.84	<b>6.26</b>	<b>5.85</b>
Region_7	<b>8.78</b>	<b>6.95</b>	10.04	8.76
Region_8	8.59	6.64	<b>8.35</b>	<b>6.69</b>
Region_9	1.18	0.81	<b>1.15</b>	<b>0.80</b>
Region_10	11.08	9.07	<b>10.33</b>	<b>7.94</b>
Region_11	8.86	6.84	<b>8.21</b>	<b>6.78</b>
Region_12	2.11	1.70	<b>1.71</b>	<b>1.45</b>
Region_13	5.86	4.49	<b>4.13</b>	<b>2.97</b>
Region_14	0.99	0.84	<b>0.90</b>	<b>0.77</b>

**Table 11.** SSA and MSSA Model Performance Comparison

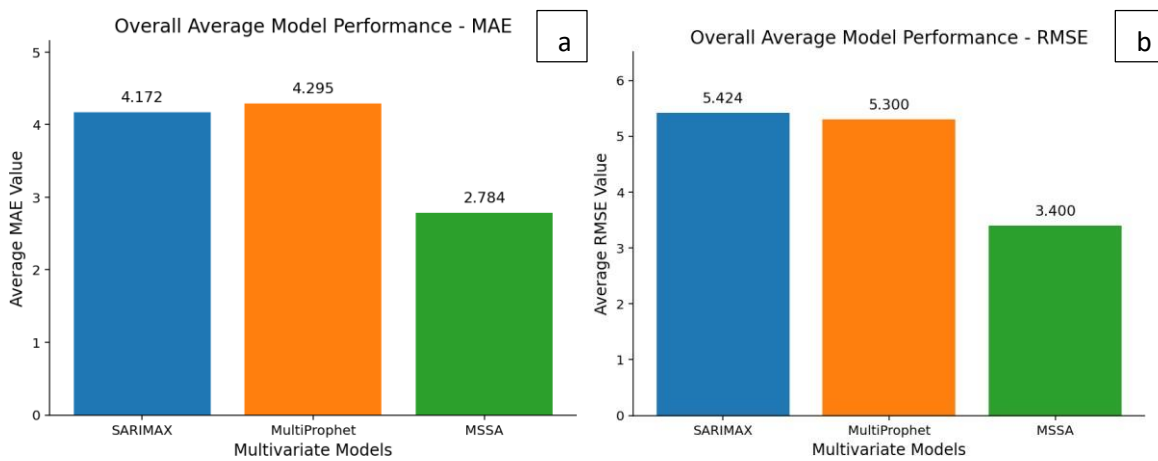
Regions	SSA RMSE	SSA MAE	MSSA RMSE	MSSA MAE
Region_1	2.06	1.49	<b>1.96</b>	<b>1.54</b>
Region_2	4.95	4.23	<b>4.38</b>	<b>3.38</b>
Region_3	7.58	5.92	<b>5.02</b>	<b>3.62</b>
Region_4	<b>3.36</b>	<b>2.65</b>	3.61	2.89
Region_5	5.50	5.18	<b>5.00</b>	<b>4.10</b>
Region_6	3.81	2.99	<b>3.66</b>	<b>2.97</b>
Region_7	7.39	5.42	<b>4.03</b>	<b>3.57</b>
Region_8	6.74	5.46	<b>4.31</b>	<b>3.55</b>
Region_9	0.92	0.65	<b>0.75</b>	<b>0.56</b>
Region_10	7.67	6.37	<b>5.56</b>	<b>5.17</b>
Region_11	6.30	4.97	<b>5.47</b>	<b>4.62</b>
Region_12	0.86	0.68	<b>0.79</b>	<b>0.65</b>
Region_13	2.27	1.83	<b>1.98</b>	<b>1.56</b>
Region_14	0.70	0.58	<b>0.62</b>	<b>0.51</b>

MSSA achieved superior performance in 13 out of 14 regions (92.9%), making it the most consistently improved multivariate model compared to its univariate counterpart.

Only region 4 showed better performance with the SSA model, suggesting that this region may have unique characteristics that benefit from simpler modeling approaches. Considering all six models, the regional performance distributions provide valuable insights. Certain regions, such as regions 9, 12, and 14, consistently exhibited the lowest error rates across all models, suggesting more predictable demand patterns in these regions. Conversely, higher error rates were observed in regions 3, 7, and 10, suggesting that these regions have more variable demand patterns and may require specialized modeling approaches.



**Figure 3.** Overall Average Performance of Univariate Models (ARIMA, Prophet, SSA) Based on RMSE and MAE Metrics



**Figure 4.** Overall Average Performance of Multivariate Models (SARIMAX, Multivariate Prophet, MSSA) Based on RMSE and MAE Metrics

The graphs in Figure 3 (a-b) and Figure 4 (a-b) display the aggregate average performance of both univariate and multivariate models based on RMSE and MAE metrics. As observed, the MSSA model attained the smallest error rates (RMSE: 3.400, MAE: 2.784), exhibiting superior predictive performance in comparison to the SARIMAX and Multivariate Prophet models. Similarly, among the univariate models, SSA performed best (RMSE: 4.298, MAE: 3.440). To determine whether the performance differences between the models were statistically significant, the error metrics of the univariate and multivariate models across the 14 regions were compared. Since the data were determined not to follow a normal distribution ( $p < 0.05$ ), the non-parametric Wilcoxon signed-rank test was employed (Alhussan et al., 2024). The test results regarding the performance differences between univariate models and their multivariate counterparts are presented in Table 12.

**Table 12.** Wilcoxon Signed-Rank Test Results for Performance Difference Between Univariate and Multivariate Models

Comparison Pair	Test Statistic (Z)	p-value	Significance Level	Result
ARIMA and SARIMAX	-2.15	0.031	$p < 0.05$	Significant Difference (Multivariate is better)
Prophet and Multivariate Prophet	-2.84	0.004	$p < 0.01$	Significant Difference (Multivariate is better)
SSA and MSSA	-3.29	$< 0.001$	$p < 0.001$	Significant Difference (Multivariate is better)

Upon examining the findings in Table 12, it is observed that there is a statistically significant difference for all comparison pairs at the 95% confidence level ( $p < 0.05$ ). Specifically, the performance of the SARIMAX model proved to be statistically superior to the ARIMA model ( $p = 0.031$ ), indicating that the inclusion of meteorological variables systematically reduces forecasting error. Furthermore, the Multivariate Prophet model achieved a significant superiority over the standard Prophet model ( $p = 0.004$ ), demonstrating that the model's sensitivity to exogenous factors enhances forecasting success. The strongest level of statistical significance was detected between the SSA and MSSA models ( $p < 0.001$ ). This result confirms that the consistent superiority demonstrated by MSSA across the regions is not coincidental and that the multivariate structure makes a critical contribution to model success. In conclusion, the statistical tests conducted scientifically validate that the improvement provided by multivariate models (particularly MSSA) compared to univariate methods did not occur by chance, and that the integration of weather data into the model significantly enhances forecasting accuracy.

Beyond statistical accuracy, the computational cost of the models was evaluated for operational feasibility. While univariate and parametric models (such as Prophet and ARIMA) demonstrated high efficiency with total training times ranging from 12 to 47 minutes for all 14 regions, the MSSA model incurred a significantly higher computational load due to intensive matrix decomposition (SVD) and sequential hyperparameter optimization. Specifically, MSSA required a total training time of approximately 6.2 hours. However, given the study's focus on daily operational planning rather than real-time forecasting, this duration presents no bottleneck. The model allows for overnight retraining cycles without disrupting daytime logistics. Consequently, the substantial and statistically significant accuracy gain provided by MSSA ( $p < 0.001$ ) justifies the additional computational investment for strategic fleet management.

While the proposed multivariate framework demonstrates significant predictive improvements, certain limitations regarding data characteristics and model constraints must be acknowledged. First, the 14-month dataset utilized in this study, while sufficient for short-term operational forecasting, constrains the ability to capture long-term inter-annual trends or rare seasonal anomalies. Although missing values were handled via linear interpolation, this smoothing approach may mask abrupt demand cessations caused by unrecorded infrastructure failures, potentially introducing minor biases during training. Second, regarding sensitivity to environmental factors, while multivariate models effectively integrate continuous weather data, they may not fully capture "threshold effects" where demand drops precipitously to zero during extreme weather events (e.g., severe storms) unless such events are explicitly modeled as binary anomalies. Third, a trade-off exists between accuracy and computational efficiency; the superior performance of the MSSA model comes at the cost of higher processing times due to intensive Singular Value Decomposition (SVD) operations, limiting its applicability to batch-based planning rather than millisecond-level real-time control. Finally, the persistently higher error rates observed in specific zones, such as regions 3 and 7, suggest that relying solely on meteorological and historical data is insufficient for areas driven by complex social dynamics, indicating a need for incorporating auxiliary data sources like public event schedules or traffic density in future studies.

## CONCLUSION

This study systematically compared six time series models (ARIMA, Prophet, SSA, SARIMAX, Multivariate Prophet, MSSA) across 14 regions to forecast electric e-bike sharing demand. Results demonstrate that multivariate approaches significantly outperform univariate counterparts, with MSSA achieving superior overall performance (RMSE: 3.400, MAE: 2.784). Integration of meteorological variables yielded 18-24% average improvement in forecast accuracy, statistically validated through Wilcoxon signed-rank tests ( $p < 0.01$ ). This research presents the first comprehensive application of MSSA for e-bike forecasting, demonstrating that joint spectral decomposition of usage and weather data captures co-dynamics more effectively than conventional methods (92.9% regional success rate). The regional typology framework establishes that optimal model selection depends on local functional characteristics, contributing to spatially-aware forecasting theory. However, an important finding reveals substantial intra-city heterogeneity, with univariate ARIMA remaining competitive in four commuter-dominated zones (Regions

1, 2, 12, 13), challenging the "one-size-fits-all" modeling paradigm and underscoring the necessity for context-aware model selection.

From an operational perspective, the weekly-ahead forecasting framework developed in this study enables practical decision support for e-bike sharing system management. Operators can leverage region-specific forecasts to optimize bike redistribution through pre-emptive rebalancing based on weather-driven demand shifts, schedule maintenance during predicted low-demand periods to minimize service disruptions, implement dynamic pricing strategies during high-demand forecasts to manage peak loads, and prioritize infrastructure investments (such as covered stations and weather protection) in high-sensitivity regions. The identification of weather-sensitive versus weather-resilient regions enables targeted resource allocation strategies, optimizing both operational efficiency and capital expenditure. Municipal transportation authorities can utilize this framework to design differentiated mobility policies, such as integrating e-bike systems with public transit in stable-demand commuter zones while promoting recreational usage in tourist areas through weather-contingent campaigns.

Future research should focus on deep learning architectures and hybrid ensemble approaches. First, the performance of advanced neural network architectures such as LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), and Transformer-based models in e-bike demand forecasting should be systematically evaluated. These models have the capacity to learn complex nonlinear relationships and long-term dependencies that traditional time series methods cannot capture. Second, hybrid ensemble approaches combining classical models examined in this study with deep learning models should be developed. In the proposed hybrid framework, each model independently generates forecasts, and these predictions are combined through dynamic weighting or meta-learning algorithms. Alternatively, for each region, the best-performing model based on validation performance can be automatically selected and used as the final decision-maker. This approach more effectively addresses regional heterogeneity by ensuring that the most suitable model is used for operational forecasting in each region. Furthermore, the model selection mechanism can be updated in real-time, allowing the system to automatically switch to the most performant model as demand patterns change.

In conclusion, effective e-bike demand forecasting requires multivariate, region-specific approaches that acknowledge spatial heterogeneity and systematically leverage environmental data. The statistical superiority of MSSA establishes spectral decomposition methods as robust alternatives to classical time series models in transportation analytics, while regional performance variations underscore the strategic importance of context-aware model selection. As urban micromobility systems proliferate globally amid climate uncertainty and rapid urbanization, this methodological framework provides both theoretical foundations and practical tools for sustainable transportation management.

### ***Performance Metrics Artificial Intelligence Contribution Statement***

Artificial intelligence tool was used only for language editing and text polishing. All scientific content, data analysis, results and conclusions were produced and verified by the authors.

### **REFERENCES**

- Afriyie, J. K., Twumasi-Ankrah, S., Gyamfi, K. B., Arthur, D., & Pels, W. A. (2020). Evaluating the performance of unit root tests in single time series processes. *Mathematics and Statistics*, 8(6), 656-664. <https://doi.org/10.13189/ms.2020.080605>
- Agarwal, A., Alomar, A., & Shah, D. (2022). On multivariate singular spectrum analysis and its variants. *ACM SIGMETRICS Performance Evaluation Review*, 50(1), 79-80. <https://doi.org/10.1145/3547353.3526952>
- Alagade, A., & Sahu, M. (2025). Satellite-based assessment and forecasting of greenhouse gas (GHG) concentrations in Indian megacities using advanced statistical methods. *Environmental Science and Pollution Research*, 32(24), 15006-15024. <https://doi.org/10.1007/s11356-025-36583-1>

- Alencar, V. A., Pessamilio, L. R., Rooke, F., Bernardino, H. S., & Borges Vieira, A. (2021). Forecasting the carsharing service demand using uni and multivariable models. *Journal of Internet Services and Applications*, 12(1), 4. <https://doi.org/10.1186/s13174-021-00137-8>
- Alharbi, F. R., & Csala, D. (2022). A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach. *Inventions*, 7(4), 94. <https://doi.org/10.3390/inventions7040094>
- Alhussan, A. A., Khafaga, D. S., Abotaleb, M., Mishra, P., & El-Kenawy, E. S. M. (2024). Global potato production forecasting based on time series analysis and advanced waterwheel plant optimization algorithm. *Potato Research*, 67(4), 1965-2000. <https://doi.org/10.1007/s11540-024-09728-x>
- Almazrouee, A. I., Almeshal, A. M., Almutairi, A. S., Alenezi, M. R., Alhajeri, S. N., & Alshammari, F. M. (2020). Forecasting of electrical generation using prophet and multiple seasonality of holt-winters models: A case study of Kuwait. *Applied Sciences*, 10(23), 8412. <https://doi.org/10.3390/app10238412>
- Chowdhury, M. S., & Hafsa, B. (2022). Multi-decadal land cover change analysis over Sundarbans Mangrove Forest of Bangladesh: A GIS and remote sensing based approach. *Global Ecology and Conservation*, 37, e02151. <https://doi.org/10.1016/j.gecco.2022.e02151>
- Cocca, M., Teixeira, D., Vassio, L., Mellia, M., Almeida, J. M., & Couto da Silva, A. P. (2020). On car-sharing usage prediction with open socio-demographic data. *Electronics*, 9(1), 72. <https://doi.org/10.3390/electronics9010072>
- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International journal of engineering business management*, 10, 1847979018808673. <https://doi.org/10.1177/1847979018808673>
- Iftikhar, H., Khan, F., Rodrigues, P. C., Alharbi, A. A., & Allohibi, J. (2025). Forecasting of inflation based on univariate and multivariate time series models: an empirical application. *Mathematics*, 13(7), 1121. <https://doi.org/10.3390/math13071121>
- İnaç, H. (2024). PREDICTION OF DRIVING TIME OF ELECTRIC SCOOTER (E-SCOOTER) DRIVERS BY MACHINE LEARNING. *Elektronik Sosyal Bilimler Dergisi*, 23(91), 1041-1057. <https://doi.org/10.17755/esoder.1432527>
- Jaber, A., Csonka, B., & Juhász, J. (2022). Long term time series prediction of bike sharing trips: A case study of Budapest city. *world*, 1(2), 3. <https://doi.org/10.1109/SCSP54748.2022.9792540>
- Jagannathan, J., & Divya, C. (2021). Time series analyzation and prediction of climate using enhanced multivariate prophet. *International Journal of Engineering Trends and Technology*, 69(10), 89-96. <https://doi.org/10.14445/22315381/IJETT-V69I10P212>
- Komarica, J., Glavić, D., & Kaplanović, S. (2024). Predicting and analyzing electric bicycle adoption to enhance urban mobility in belgrade using ANN models. *Applied Sciences*, 14(19), 8965. <https://doi.org/10.3390/app14198965>
- Kwarteng, S., & Andreevich, P. (2024). Comparative analysis of ARIMA, SARIMA and Prophet model in forecasting. *Research & Development*, 5(4), 110-120. <https://doi.org/10.11648/j.rd.20240504.13>
- Patterson, K., Hassani, H., Heravi, S., & Zhigljavsky, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *Journal of Applied Statistics*, 38(10), 2183-2211. <https://doi.org/10.1080/02664763.2010.545371>
- Sanami, S., Mosalli, H., Yang, Y., Yeh, H. G., & Aghdam, A. G. (2025, July). Demand forecasting for electric vehicle charging stations using multivariate time-series analysis. In *2025 American Control Conference (ACC)* (pp. 3461-3466). IEEE. <https://doi.org/10.48550/arXiv.2502.16365>
- Subramanian, M., Cho, J., Veerappampalayam Easwaramoorthy, S., Murugesan, A., & Chinnasamy, R. (2023). Enhancing sustainable transportation: AI-driven bike demand forecasting in smart cities. *Sustainability*, 15(18), 13840. <https://doi.org/10.3390/su151813840>
- Swari, M. H. P., Irawan, H. A., Muliawati, A., Aliansyah, Z., & Diyasa, I. G. S. M. (2025, December). Enhancing Time Series Forecasting Accuracy through Hybrid ARIMA-MLP Integration: A Case Study on E-Bicycle Sales.

In *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 967-972). IEEE. <https://doi.org/10.1109/ICIMCIS68501.2025.11327149>

Taoussi, B., Boudia, S. M., & Mazouni, F. S. (2025). Wind speed forecasting using univariate and multivariate time series models. *Stochastic Environmental Research and Risk Assessment*, 39(2), 547-579. <https://doi.org/10.1007/s00477-024-02881-2>

Vishnu, G., Kaliyaperumal, D., Pati, P. B., Karthick, A., Subbanna, N., & Ghosh, A. (2023). Short-term forecasting of electric vehicle load using time series, machine learning, and deep learning techniques. *World Electric Vehicle Journal*, 14(9), 266. <https://doi.org/10.3390/wevj14090266>

Wang, J., Peng, X., Wu, J., Ding, Y., Ali, B., Luo, Y., ... & Zhang, K. (2024). Singular spectrum analysis (SSA) based hybrid models for emergency ambulance demand (EAD) time series forecasting. *IMA Journal of Management Mathematics*, 35(1), 45-64. <https://doi.org/10.1093/imaman/dpad019>

Zhang, Y., Zhong, M., Geng, N., & Jiang, Y. (2017). Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China. *PloS one*, 12(5), e0176729. <https://doi.org/10.1371/journal.pone.0176729>

Zunic, E., Korjenic, K., Hodzic, K., & Donko, D. (2020). Application of facebook's prophet algorithm for successful sales forecasting based on real-world data. *arXiv preprint arXiv:2005.07575*. <https://doi.org/10.5121/ijcsit.2020.12203>