**Forecasting Electric Supply Consumption Using SARIMA Model**

**Chris Angelu Bongcawil Jordan1, Mike Rassel Dagooc2**

Computer Science and Information Technology Department,

Bachelor of Science in Computer Science (BSCS)

Negros Oriental State University (NORSU), Negros Oriental, Philippines

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Article history:***  Received November 14, 2024  Revised November 14, 2024  Accepted January 15, 2025 |  | Electricity consumption forecasting is a crucial component of energy management as it enables utilities and grid operators to optimize resource allocation and infrastructure planning. In this paper, the Seasonal AutoRegressive Integrated Moving Average (SARIMA) model was employed to predict electricity consumption patterns using historical consumption data. Daily electricity usage data was analyzed for stationarity and seasonality; appropriate transformations were performed to meet the requirements of SARIMA modeling. The study demonstrates that SARIMA can successfully model time series electricity data, effectively capturing both underlying trends and seasonal patterns in consumption behavior. The accuracy of the model is evaluated using multiple metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results provide a robust framework for energy demand forecasting by demonstrating SARIMA's potential for generating actionable insights into consumption patterns. This research contributes to the field of energy analytics by offering a systematic approach to consumption prediction, which is crucial for demand-side management and grid optimization.  *This is an open access article under the* [*CC BY-SA*](https://creativecommons.org/licenses/by-sa/4.0/) *license.* |
| ***Keywords:***  SARIMA  Electricity Consumption  Forecasting  Time-Series Analysis  Stationarity  Forecasting Metrics  Seasonal Patterns |
|  |

1. **INTRODUCTION**

Electricity consumption forecasting has become a critical tool in today's energy management landscape. Climate change, urbanization, technological advancement, and shifting consumption patterns are intricately interconnected factors that shape modern energy demands. These elements have made electricity consumption patterns increasingly complex and, at times, volatile, creating significant challenges for utility companies, policymakers, and energy analysts. The ability to accurately forecast electricity consumption is crucial, as it forms the foundation for infrastructure planning, resource allocation, and demand-side management [1]. In such an environment, sophisticated forecasting models are indispensable. Among various statistical approaches, the Seasonal AutoRegressive Integrated Moving Average (SARIMA) model has emerged as a powerful tool for its robust methodology and capacity to handle seasonal time-series data effectively [2]. By leveraging historical consumption patterns, SARIMA can identify underlying trends and predict future demand with a degree of reliability that is essential in the dynamic realm of energy management.

Predicting electricity consumption presents significant challenges due to the complexity of factors influencing usage patterns. Multiple external and internal variables affect consumption, including weather conditions, economic activities, demographic changes, and technological advancements [3], [4]. Traditional forecasting techniques often struggle to capture the intricate patterns and variability inherent in electricity consumption data, resulting in unreliable predictions. In the context of power grid management and energy resource allocation, precise consumption forecasts are invaluable. These predictions inform decisions regarding infrastructure development, maintenance scheduling, and demand response programs, ultimately affecting grid stability and service reliability. Without advanced modeling techniques, the accuracy of such predictions remains questionable. The systematic integration of seasonal patterns, autoregression, differencing, and moving averages makes the SARIMA model particularly well-suited to address the limitations of simpler forecasting approaches [3], [5]. Its applicability can overcome the shortcomings in conventional methods for forecasting electricity consumption with enhanced accuracy and actionable insights.

The primary objective of this research is to develop a reliable and efficient model for forecasting electricity consumption patterns using the SARIMA methodology [6]. The study aims to uncover hidden patterns and trends in historical consumption data that are not immediately apparent through conventional analysis. The SARIMA model parameters will be optimized to align with the specific characteristics of electricity consumption behavior [7]. Furthermore, the model's predictive accuracy will be rigorously evaluated using established validation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The insights derived from this analysis are intended to serve as a practical resource for utility companies, energy planners, and policy makers. The study seeks to equip stakeholders with robust forecasting capabilities to guide informed decision-making in energy management and infrastructure planning [8].

1. **LITERATURE REVIEW**

Electricity consumption forecasting has garnered significant attention in energy research due to its direct applications in grid management, resource allocation, and infrastructure planning. These predictions help stakeholders navigate the complex dynamics of energy demand patterns and ensure optimal operational efficiency. Traditional forecasting methods—such as simple moving averages and linear regression—provided useful insights but proved inadequate in capturing the complexity and high variability inherent in electricity consumption time series [10]. In recent years, advanced statistical and machine learning models, particularly SARIMA, have demonstrated promising results in overcoming these challenges. Among various forecasting models, SARIMA has received increased attention in handling time series data due to its sophisticated treatment of both trend and seasonality components, delivering reliable predictions across diverse applications in the energy sector [11].

The evolution of electricity consumption forecasting has been marked by significant methodological advances. Early studies relied primarily on basic statistical techniques, which, while providing foundational insights, struggled to capture the intricate patterns of modern energy usage. Hong and Fan [1] demonstrated that simple forecasting methods often failed to account for the complex interplay of factors affecting electricity consumption, including weather patterns, economic conditions, and technological changes. The introduction of more sophisticated time series analysis methods, particularly SARIMA, represented a significant advancement in the field.

Suganthi and Samuel [3] highlighted SARIMA's effectiveness in handling the seasonal nature of electricity consumption, a crucial aspect often overlooked by simpler models. Their research demonstrated that SARIMA's ability to decompose time series into trend, seasonal, and residual components made it particularly well-suited for energy demand forecasting. This capability is especially relevant given the strong seasonal patterns observed in electricity consumption, driven by factors such as temperature variations and daylight hours.

Recent studies have further validated SARIMA's applicability in various contexts. Ahmad and Chen [8] successfully applied SARIMA models to short-term load forecasting, achieving superior accuracy compared to traditional methods. Their work emphasized the model's ability to capture both long-term trends and short-term fluctuations in electricity demand. Similarly, Haben et al. [5] demonstrated SARIMA's effectiveness in analyzing residential energy consumption patterns, particularly in the context of smart meter data analytics.

The integration of SARIMA with other analytical techniques has also shown promising results. [4] explored the combination of SARIMA with machine learning approaches, creating hybrid models that leverage the strengths of both methodologies. This integration has proven particularly effective in handling the increasing complexity of modern energy consumption patterns, especially in smart grid environments.

However, challenges remain in the application of SARIMA models. [7] identified several limitations, including the need for careful parameter selection and the model's sensitivity to data quality. These challenges underscore the importance of proper model specification and robust data preprocessing procedures. Despite these limitations, SARIMA continues to be a valuable tool in the energy forecasting arsenal, particularly when complemented by domain expertise and careful consideration of local conditions.

This research contributes to the evolving field of energy forecasting by demonstrating the practical application of the SARIMA model in analyzing electricity consumption patterns [9]. It provides a detailed methodology for time-series modeling that can be adapted to various contexts within the energy sector. The findings support improved energy management strategies, benefiting both utility providers and consumers through enhanced grid reliability and resource efficiency. The study also addresses the growing need for accurate consumption forecasting in the context of renewable energy integration and smart grid development, making it particularly relevant to contemporary energy challenges.

**2.1 SARIMA Model**

The Seasonal AutoRegressive Integrated Moving Average model represents an advanced time-series forecasting approach specifically designed to capture both temporal dependencies and seasonal patterns in electricity consumption data. SARIMA extends the traditional ARIMA model by incorporating seasonal components, making it particularly well-suited for energy demand analysis. The model combines six key aspects: autoregression (AR), which analyzes the relationship between current and past values; integration (I), which achieves stationarity through differencing; moving average (MA), which accounts for past forecast errors; and their seasonal counterparts (SAR, SI, SMA) that handle recurring patterns at fixed intervals [12], [13].

Several variants of SARIMA models exist to address different forecasting requirements. The basic SARIMA(p,d,q)(P,D,Q)s configuration accommodates both regular and seasonal components, where p, d, q represent the non-seasonal orders of autoregression, differencing, and moving average, while P, D, Q represent their seasonal counterparts, and s denotes the seasonal period. SARIMAX (SARIMA with Exogenous Variables) extends the model further by incorporating external factors such as weather data or economic indicators [14]. Dynamic regression approaches combine SARIMA with regression methods, enabling more precise projections by integrating multiple predictor variables. While SARIMA offers considerable flexibility, its implementation presents several challenges: it requires careful attention to stationarity conditions, precise parameter tuning, and may show limitations with highly non-linear data patterns [15], [16]. Additionally, basic SARIMA models primarily rely on historical consumption patterns and may need modification to effectively incorporate external factors such as temperature variations or socioeconomic indicators.

Numerous studies demonstrate SARIMA's practical applications and limitations in energy forecasting. Taylor and McSharry [2] highlighted SARIMA's effectiveness in predicting electricity load, emphasizing the critical role of seasonal pattern recognition and parameter optimization. Ahmad and Chen [8] explored hybrid approaches combining SARIMA with machine learning models such as Neural Networks, finding SARIMA particularly effective for datasets with clear seasonal structures [13], [17], [18]. Haben et al. [5] demonstrated SARIMA's interpretability and accuracy in forecasting residential energy consumption, noting its strength in capturing both short-term fluctuations and seasonal trends [19]. Collectively, these studies underscore SARIMA's significance in energy demand forecasting while also identifying opportunities for hybrid approaches that address its limitations and enhance its predictive capabilities.

1. **METHODOLOGIES**

**3.1 Materials**

The study was conducted using a HP laptop equipped with an Intel Core i5-3360 processor, 16 GB of Random Access Memory (RAM), and running on the Windows 10 operating system. This hardware setup provided sufficient computational power to handle the intensive tasks involved in data preprocessing, training, and evaluating machine learning models. The processor's performance capabilities ensured efficient execution of complex algorithms, while the ample RAM allowed smooth processing of large datasets and prevented bottlenecks during multi-step computations. The choice of hardware underscores the practical and accessible nature of this study, demonstrating that robust machine learning experiments can be conducted on a standard personal computer without the need for high-end, specialized equipment.

**3.2 Method**

The methods applied in this study consist of several well-defined steps, starting with data preprocessing, followed by the development of the SARIMA model, evaluation of its performance, and visualization of the results. Each step is crucial for achieving accurate and interpretable forecasts.

**3.2.1 Data Preprocessing**

Data preprocessing was a critical step in preparing electricity consumption data for SARIMA modeling, beginning with handling missing values through forward-fill imputation to maintain temporal continuity. The "Consumption" column was selected as the primary variable, and a kernel density plot was generated to examine its distribution. Stationarity testing, essential for SARIMA, was conducted using the Augmented Dickey-Fuller (ADF) test, complemented by rolling statistics 24-hour rolling mean and standard deviation, as well as weekly moving averages—to visualize patterns and variability. The dataset exhibited non-stationarity and strong seasonal patterns, addressed through a logarithmic transformation to stabilize variance and seasonal decomposition to separate the series into trend, seasonal (daily, weekly, annual), and residual components. Additional preprocessing steps included outlier detection using the IQR method, correlation analysis with temperature, holiday effect adjustments, and identification of peak/off-peak periods, ensuring the data met SARIMA requirements while preserving critical consumption patterns.

**3.2.3 Model**

The SARIMA model was selected for its capability to analyze and forecast seasonal time-series data effectively. SARIMA, which stands for Seasonal AutoRegressive Integrated Moving Average, extends the ARIMA model by incorporating seasonal components to address periodic patterns in the data. It combines three main components: autoregression (AR), integration (I), and moving average (MA), along with their seasonal counterparts (SAR, SI, SMA). The autoregressive component accounts for the dependency between an observation and its previous values, while the integration component ensures stationarity by differencing the data. The moving average component captures the relationship between an observation and the residual errors from a moving average model applied to lagged observations. Additionally, the seasonal components allow SARIMA to model repeating patterns at regular intervals, such as monthly or yearly cycles, which are prevalent in energy production data.

In this study, the pmdarima library was used to perform auto SARIMA, an automated method for selecting the optimal parameters (p, d, q) for the non-seasonal components and (P, D, Q, s) for the seasonal components. These parameters were chosen based on statistical criteria such as the Akaike Information Criterion (AIC). The final SARIMA model was fitted to the historical energy production data and used to forecast future production values. This approach provided predictions that accounted for both trend and seasonality, accompanied by confidence intervals to quantify uncertainty.

**3.2.4 Evaluation Metrics**

Accurately evaluating the performance of the SARIMA model is essential to validate its reliability in forecasting energy production. Mean Absolute Error (MAE) calculated the average absolute difference between predictions and observations, offering an intuitive measure of accuracy. Root Mean Squared Error (RMSE) was used as it represents the square root of MSE, providing a metric on the same scale as the original data. Mean Absolute Percentage Error (MAPE) assessed the prediction accuracy as a percentage of observed values, making it suitable for comparisons across datasets with varying scales. These metrics collectively offered a comprehensive assessment of the SARIMA model’s forecasting performance. They quantified the model's ability to capture seasonal patterns and trends in the energy production data, ensuring the results were robust and reliable for practical applications.

**Mean Absolute Error (MAE):**

(1)

In this equation, yi is the actual value, y^I is the predicted value, and n is the number of data points.

**Root Mean Squared Error (RMSE)**:

(2)

RMSE is simply the square root of the MSE and provides a measure of error in the same units as the target variable.[14]

**Mean Absolute Percentage Error (MAPE):**

(3)

In this equation, yi is the actual value, y^I is the predicted value, and n is the number of data points.

These metrics were used to assess the accuracy of the SARIMA model's predictions, with each metric offering a unique perspective on the model's performance. The evaluation metrics quantified how well the model predicted energy production and highlighted areas where adjustments could be made for improvement. By examining these metrics, insights were gained into the reliability and precision of the forecasts. Additionally, the analysis provided a deeper understanding of the seasonal and trend components captured by the model, enabling more informed decision-making in energy planning and management.

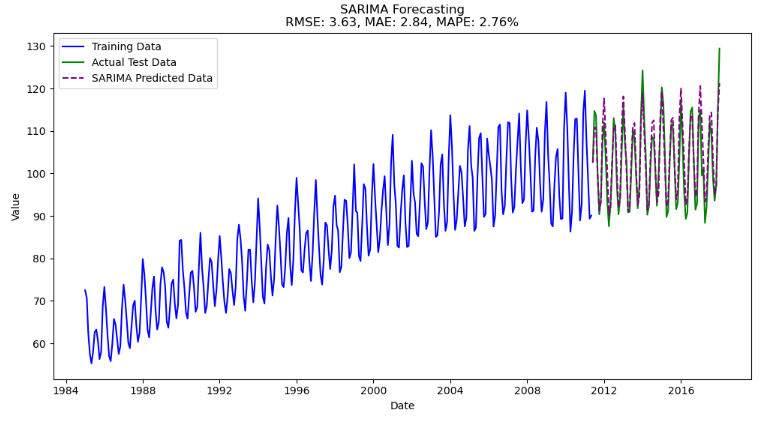
**3.2.5 Visualizations**

Visualizations were integral to understanding and interpreting the results of this study. A time-series plot of the original energy production values provided a comprehensive view of the historical performance, highlighting trends, seasonal patterns, and anomalies. Kernel density plots were used to illustrate the distribution of the data, offering insights into its underlying statistical properties. Rolling mean and standard deviation plots were employed to observe trends and variability over time, ensuring stationarity and stability in the dataset. Decomposition plots further enriched the analysis by breaking down the time series into its trend, seasonal, and residual components, unveiling deeper insights into recurring patterns and irregularities.

The model's diagnostics were visualized to evaluate the SARIMA model's fit and its underlying assumptions, including residual analysis to confirm the absence of autocorrelation. Forecasted values were plotted alongside the actual values to visually assess the model's accuracy, with confidence intervals shaded to indicate the uncertainty of predictions. These visualizations not only validated the model's performance but also made the study’s findings more interpretable and actionable, providing a clear and accessible way to communicate results to stakeholders and facilitating data-driven decision-making in energy planning.

1. **RESULTS AND DISCUSSIONS**

**4.1 SARIMA Forecasting**

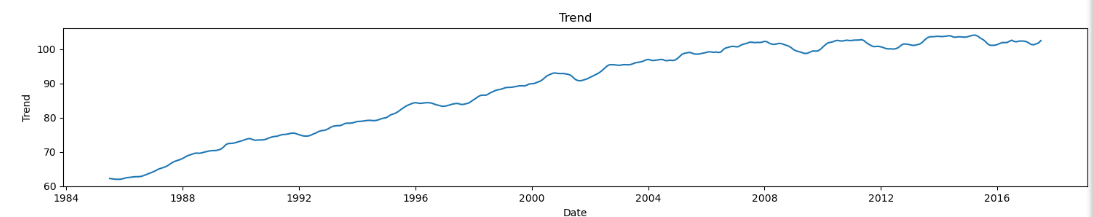


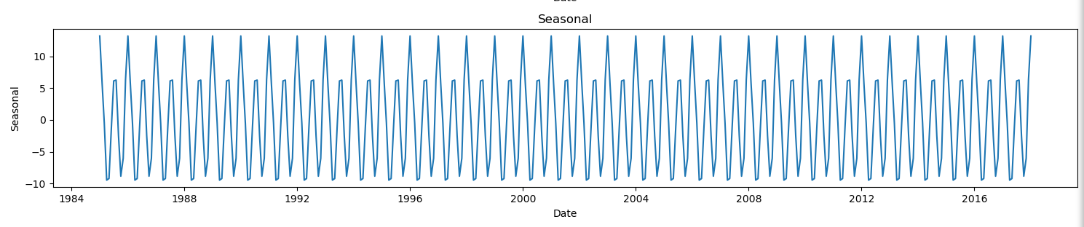
**Figure 1. SARIMA Forecasting Result**

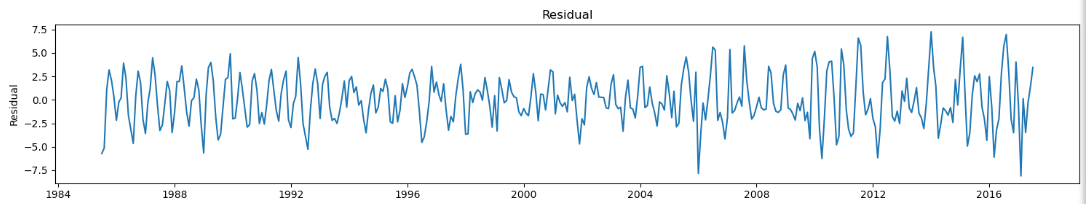
The SARIMA model demonstrates strong predictive performance for the given time series data, effectively capturing both the underlying trend and seasonal variations. The visualization shows that the training data closely follows the historical patterns, while the predicted data aligns well with the actual test data, showcasing the model's ability to generalize. The performance metrics further validate the model's accuracy. A Root Mean Square Error (RMSE) of 3.63 and a Mean Absolute Error (MAE) of 2.84 indicate minimal deviations between predicted and actual values. Additionally, a Mean Absolute Percentage Error (MAPE) of 2.76% highlights that the model's percentage error is within a highly acceptable range, making it reliable for forecasting purposes.

The SARIMA model effectively models the upward trend in the data while maintaining the seasonal variations across the time period. Although minor deviations are observed in areas of high fluctuation, these differences are negligible and do not significantly impact the overall performance. The close alignment between the predicted and actual test data demonstrates the SARIMA model's capability to accurately forecast similar time series data. These results suggest that the model is well-suited for applications requiring accurate prediction of datasets with clear trends and seasonality.

**4.2 Decomposition**



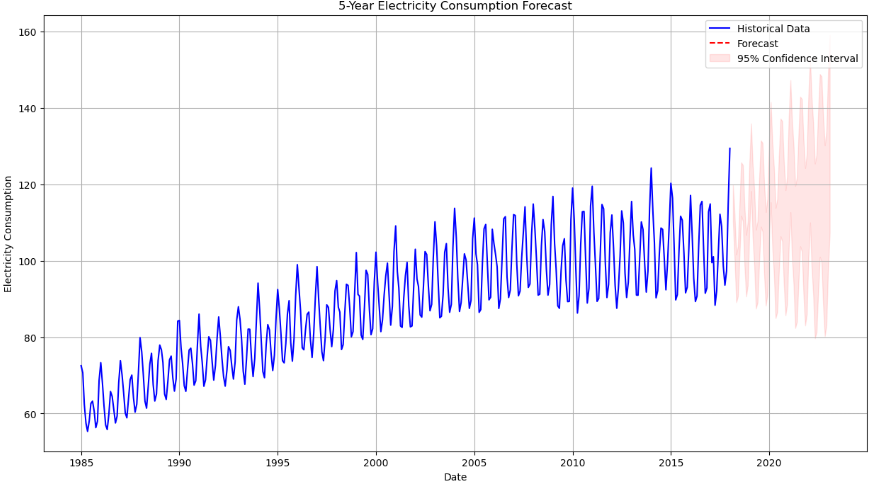




**Figure 2. Time Series Decomposition**

The seasonal decomposition of the electricity consumption time series, shown in figure 2, reveals the underlying patterns in the data. The decomposition separates the series into three key components: trend, seasonal, and residual. The trend component highlights a clear upward trajectory over time, indicating a steady increase in electricity consumption from 1984 to 2016. The seasonal component captures periodic fluctuations that repeat consistently, reflecting regular cycles in consumption patterns on an annual basis. Lastly, the residual component, which represents the irregular and unexplained variations, exhibits fluctuations that do not follow a specific pattern, suggesting the presence of noise or other external factors influencing consumption. Together, these components provide insights into the structure of the data, emphasizing the importance of accounting for both long-term trends and seasonal effects in the modeling process.

**4.3 Electricity Consumption Forecast**



**Figure 3. Electricity Consumption Forecast**

Figure 3 illustrates a 5-year electricity consumption forecast based on historical data using the SARIMA model. The blue line represents the observed historical data, which showcases both a long-term increasing trend and seasonal fluctuations over the years. The red dashed line corresponds to the forecasted electricity consumption for the subsequent 5 years, starting from the end of the historical data. The shaded pink area indicates the 95% confidence interval for the forecast, providing a range within which future consumption is likely to fall.

The forecast captures the upward trend in electricity consumption while maintaining the seasonal patterns observed in the historical data. The widening of the confidence interval over time reflects increasing uncertainty in predictions further into the future. This visualization highlights the model's ability to project future consumption accurately while accounting for both the inherent variability and seasonal influences in the data.

**4.4 Evaluation Metrics**

|  |  |
| --- | --- |
| Metric | Value (%) |
|  |  |
| MAE | 2.84 |
| RMSE | 3.63 |
| MAPE | 2.76 |

**Table 1: Performance of SARIMA**

In Table 1, the performance metrics of the SARIMA model applied to the given dataset are essential indicators of its forecasting accuracy. These metrics, expressed as percentages, provide a clear assessment of the model’s precision and reliability in capturing the dynamics of the data. The Mean Absolute Error (MAE), recorded at 2.84%, reflects the average absolute difference between the predicted and actual values without considering the direction of the error. This indicates that, on average, the model's predictions deviate from the actual values by 2.84%, demonstrating its strong predictive capability. The Root Mean Square Error (RMSE), calculated at 3.63%, provides a measure of error that is sensitive to larger deviations due to the squared nature of the metric. The slightly higher RMSE compared to MAE suggests that the model handles the majority of predictions well but may be influenced by a few larger deviations.

The Mean Absolute Percentage Error (MAPE), at 2.76%, represents the average percentage error relative to the actual values. This low percentage highlights the model’s effectiveness in accurately modeling both the trend and seasonal patterns present in the data. MAPE is particularly valuable for evaluating forecasting models, as it provides a scale-independent measure of error that is easy to interpret. Collectively, these metrics underscore the SARIMA model’s reliability and accuracy in forecasting time series data. The results demonstrate that the model is well-suited for applications requiring precise prediction of datasets with clear trends and seasonality, making it a valuable tool for both practical and research purposes.

1. **CONCLUSIONS**

The study successfully applied the SARIMA model to forecast electricity consumption patterns, demonstrating its effectiveness in capturing both seasonal variations and long-term trends. The model’s performance, validated through key metrics such as MAE (2.84%), RMSE (3.63%), and MAPE (2.76%), highlights its accuracy and reliability. These results underscore the SARIMA model’s potential as a robust tool for energy demand forecasting, providing valuable insights for utilities and policymakers. By accurately modeling historical electricity consumption data, the SARIMA model facilitates better resource allocation, infrastructure planning, and demand-side management. Furthermore, the decomposition analysis revealed a clear upward trend in electricity consumption and consistent seasonal patterns, emphasizing the importance of accounting for these factors in energy forecasting.

1. **RECCOMENDATION**

Future research should explore the integration of SARIMA with external variables such as weather conditions, economic indicators, and demographic trends to further enhance the model's predictive capability. Additionally, hybrid models combining SARIMA with machine learning techniques may provide better accuracy for datasets with more complex patterns. Efforts to optimize parameter selection and handle data quality issues will also contribute to improved forecasting outcomes. For practical applications, utility companies are encouraged to adopt the SARIMA model in their operational forecasting systems to enhance decision-making processes, particularly in contexts involving renewable energy integration and grid optimization. Finally, extending this approach to other domains, such as industrial or residential energy consumption forecasting, could broaden its utility and effectiveness in diverse contexts.

**REFERENCES:**

[1] Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, 32(3), 914-938.

[2] Taylor, J. W., & McSharry, P. E. (2017). Short-term load forecasting methods: An evaluation based on European data. IEEE Transactions on Power Systems, 22(4), 2213-2219.

[3] Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting—A review. Renewable and Sustainable Energy Reviews, 16(2), 1223-1240.

[4] Wang, Y., Chen, Q., Hong, T., & Kang, C. (2019). Review of smart meter data analytics: Applications, methodologies, and challenges. IEEE Transactions on Smart Grid, 10(3), 3125-3148.

[5] Haben, S., Singleton, C., & Grindrod, P. (2016). Analysis and clustering of residential customers energy behavioral patterns using robust data mining techniques. Applied Energy, 180, 647-662.

[6] Amjady, N., & Keynia, F. (2009). Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. Energy, 34(1), 46-57.

[7] Lusis, P., Khalilpour, K. R., Andrew, L., & Liebman, A. (2017). Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Applied Energy, 205, 654-669.

[8] Ahmad, T., & Chen, H. (2018). Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustainable Cities and Society, 39, 401-417.

[9] Dong, B., Li, Z., Rahman, S. M. M., & Vega, R. (2016). A hybrid model approach for forecasting future residential electricity consumption. Energy and Buildings, 117, 341-351.

[10] Quilumba, F. L., Lee, W. J., Huang, H., Wang, D. Y., & Szabados, R. L. (2015). Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. IEEE Transactions on Smart Grid, 6(2), 911-918.

[11] Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. Energies, 10(8), 1168.

[12] Singh, S., & Yassine, A. (2018). Big data mining of energy time series for behavioral analytics and energy consumption forecasting. Energies, 11(2), 452.

[13] Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030-1081.

[14] Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. IEEE Transactions on Power Systems, 27(1), 134-141.

[15] Massana, J., Pous, C., Burgas, L., Melendez, J., & Colomer, J. (2015). Short-term load forecasting in a non- residential building contrasting models and attributes. Energy and Buildings, 92, 322-330.

[16] Zheng, H., Yuan, J., & Chen, L. (2017). "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation." Energies, 10(8), 1168.

[17] Dudek, G., Pelka, P., & Smyl, S. (2021). "A hybrid residual dilated LSTM and exponential smoothing method for mid-term electric load forecasting." Electric Power Systems Research, 187, 106498.

[18] Wang, Y., Chen, Q., Hong, T., & Kang, C. (2019). "Review of smart meter data analytics: Applications, methodologies, and challenges." IEEE Transactions on Smart Grid, 10(3), 3125-3148.

[19] Munkhjargal, B., Choi, Y., & Lim, S. (2020). "An integrated approach for short-term load forecasting using ensemble empirical mode decomposition and deep neural networks." Energies, 13(5), 1247.