



An Evaluation of ARIMA and Persistence Models in IoT-Driven Smart Homes

Ahmed Abdelmgeed¹, Ahmed Mohamed Zaki^{*2}, Marwa Adel Soliman^{3,4}

¹ Department of Communications & Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, Egypt

² Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA

³ Department of Architecture, Faculty of Engineering, Mansoura University, Mansoura, Egypt

⁴ Department of Architecture, Delta Higher Institute of Engineering and Technology, Mansoura, Egypt

Emails: ahmedabdelmgeed284@gmail.com; azaki@jcsis.org; marwa_elfiky@mans.edu.eg

Abstract

Commencing with the transformative fusion of Smart Home and Internet of Things (IoT) technologies, this study scrutinizes the efficacy of predictive modeling approaches, specifically the autoregressive integrated moving average (ARIMA) and persistence algorithms. The primary focus lies in their potential for forecasting and optimizing energy consumption dynamics within the intricate framework of smart homes. The investigation reveals a nuanced comparison between the proposed ARIMA and conventional Persistence models. Smart Home, emblematic of innovative living, integrates seamlessly with IoT, promising an intelligent and interconnected domestic ecosystem. To enhance energy efficiency, this study explores the ARIMA model's capabilities alongside the persistence algorithm. Notably, the proposed ARIMA model showcases exceptional prowess in forecasting, substantiated by a significantly lower *RMSE* compared to the Persistence model. The ARIMA model, with an Root Mean Square Error value of 0.03378, outshines the Persistence model with a higher Root Mean Square Error value of 0.158 when evaluated on the test dataset. This substantial reduction in *RMSE* emphasizes the superior performance of the ARIMA model, making it a compelling choice for time series forecasting tasks. Beyond quantitative metrics, the precision of the ARIMA model holds transformative potential, promising cost-effective energy consumption, proactive maintenance, and an elevated quality of life within smart homes. This research establishes a robust foundation for integrating advanced predictive modeling, particularly the ARIMA model, to enhance the efficiency, sustainability, and inhabitant satisfaction of smart homes in the era of IoT.

Keywords: Smart Home; ARIMA model; Time Series, IOT; Persistence algorithm; ARIMA model; Modern architecture.

1 Introduction

In the ever-evolving terrain of the 21st century, the continued progress of technological innovations has seamlessly woven into the fabric of our daily existence, fundamentally altering our interactions with the external environment. This metamorphosis has permeated the essence of our residential spaces, triggering a paradigmatic transformation in the conventional understanding of living areas. At the forefront of this

revolution stands the visionary idea of "Smart Homes" – a seamless integration of modern architecture and cutting-edge technology that transcends mere convenience [1]. As shown in figure 1, these homes are a departure from the conventional, evolving into sophisticated ecosystems prioritizing efficiency, security, and personalized experiences. Smart homes are not just about physical structure; they embody a fusion of design and technology that revolutionizes how we perceive, control, and secure our domestic settings. [2,3] The impact of these dwellings on our daily routines is profound, necessitating a closer examination of their intricacies. Unraveling the network of interconnected devices that elevate the intelligence of these modern living environments becomes a focal point in understanding the technological forces steering this phenomenon [4].



Figure 1: Modern Architecture Meets Smart Home Technology.

Amid these advancements, exploration doesn't conclude; it evolves. The challenges posed by intelligent homes demand sophisticated solutions; this is where time-series artificial intelligence (AI) comes into play. Time series AI introduces a dynamic and predictive approach, enabling the efficient management and optimization of data streams within smart homes [5]. It responds to the evolving needs of these homes, from predicting energy usage patterns to enhancing security systems through using historical data. Our investigation extends beyond theoretical concepts to the practical applications of artificial intelligence in the real world. As in figure 2, exploring how AI can significantly enhance the efficacy of smart home systems, we unveil the vital role of time-series AI analysis in dissecting practical scenarios [6]. As we navigate the complexities of innovation in smart homes, the study invites you to explore the potential of artificial intelligence, notably time series, in shaping the future of modern life.



Figure 2: Effect of AI on The Efficacy of Smart Homes.

The paper addresses vital research questions aimed at unraveling the intricacies of energy forecasting in smart homes, laying the foundation for a deeper understanding of the dynamic interplay between modern architecture, time series analysis, and sophisticated algorithms. Through this exploration, we seek to illuminate the path toward homes that respond intelligently to our needs and actively contribute to shaping the future of modern living.

- How effective are the autoregressive integrated moving average (ARIMA) and persistence algorithms in predicting energy usage in smart homes?
- What specific advantages does the ARIMA model offer over the persistence algorithm in the

DOI: <https://doi.org/10.54216/JAIM.060201>

Received: May 13, 2023 Revised: August 02, 2023 Accepted: December 03, 2023

context of smart home energy consumption predictions?

- How does the ARIMA model demonstrate its ability to discern complex patterns and interdependencies in smart home energy consumption data?
- What potential benefits may arise from the increased accuracy of ARIMA model predictions, particularly in lowering energy expenditures, promoting proactive maintenance, and enhancing occupant comfort?
- In what ways do advanced forecasting techniques, such as ARIMA, contribute to optimizing energy management within the evolving landscape of smart homes?

The subsequent sections of this work are structured as follows: The paper includes a discussion of the Related Work in Section 2, a description of the Material and Methods in Section 3, the presentation of the Results in Section 4, and a summary of the Conclusion in Section 5.

2 Related Work

A comprehensive literature study [7] examines application of information technologies in engineering. Artificial Intelligence (AI), Big Data, Blockchain, and IoT are all covered. These cutting-edge technologies' intricate connections and seamless integration are the focus. The study focuses on intelligent energy management strategies. AI models for energy consumption, load profiles, and resource scheduling are used to study energy resource efficiency and prediction. Big data platforms and data mining help AI models discover new functions and improve information. IoT devices integrate AI with a variety of hardware and software, facilitated by the IoT platform, to streamline data transmission and storage for stakeholders and data mining. Blockchain and cryptocurrency are being studied for IoT integration and cloud layer enhancement to optimize data storage. The energy industry's low-carbon transition requires the seamless integration of AI, big data, and advanced digital technologies.

In a distinct study initiative [8], the focus is directed toward the intricate task of energy demand forecasting at the individual household level—a crucial aspect within intelligent and sustainable environments. The study unveils the limitations of the conventional one-step forecasting approach characterized by a brief prediction period. Instead, it advocates for a more nuanced two-phase methodology to enhance accuracy. In the initial phase, the study employs long short-term memory (LSTM) models to predict total generative active power over an extensive 500-hour duration. Transitioning to the latter phase, a hybrid deep learning model takes center stage, amalgamating the convolutional features of a neural network with LSTM. This combined architecture, known as Convolutional LSTM (ConvLSTM), surpasses existing models, demonstrating the lowest root mean square error in estimating weekly residential electricity consumption. Another study dimension [9] scrutinizes how the Internet of Things (IoT) ecosystem establishes trust and integrity within smart homes. Emphasizing the critical need for dependability and reliability in the IoT network, given its substantial data conveyance, the study proposes a machine learning algorithm. This algorithm incorporates weather information and energy data from smart homes, employing the SDAR-based Change Finder algorithm to bring unexpected patterns in appliance usage to light. This underscores the paramount importance of privacy in the context of IoT-generated data. The analysis utilizes various time series models, including ARIMA, SARIMA, LSTM, Prophet, Light GBM, and VAR. Evaluation metrics such as *RMSE*, *MSE*, and *MAE* point towards the superior performance of the ARIMA model in effectively addressing the complexities of energy demand forecasting at the individual household level within the IoT ecosystem.

In a parallel investigation [10], the study shifts its lens to the escalating utilization of IoT devices for monitoring electrical energy consumption in intelligent residential homes. Aligned with global commitments to sustainable development, the study accentuates the pursuit of reducing carbon footprints and embracing clean energy. Comparative analyses of energy consumption data from intelligent residential homes with predictive models, including SVR, GRU, CNN-LSTM, and CNN-GRU, unravel nuanced insights. The empirical findings underscore the degradation of machine learning SVR algorithms as data volume increases. At the same time, the proposed CNN-GRU architecture emerges as a standout performer, yielding a 17.4% improvement in Mean Absolute Error (MAE), particularly in daily granularity data. LSTM, though marginally better in MAE for hourly granularity data, along with CNN-LSTM, proves effective in identifying outliers in energy consumption data. An innovative methodology takes center stage in addressing the multi-objective scheduling problem of energy demand planning in Smart Homes [11]. Integrating three distinct Artificial Intelligence techniques to harmonize energy costs and user comfort, the approach employs an Elitist, Non-dominated Sorting Genetic Algorithm II for demand-side management. Considering factors such as electricity price fluctuations, equipment priority, operational cycles, and battery bank resources, the

approach showcases the efficacy of Support Vector Regression in predicting distributed generation for the subsequent day. Numerical simulations with accurate data from smart homes underscore the substantial cost reduction (51.4%) achieved by the proposed combination of AI techniques when comparing Smart Homes with and without distributed generation and a battery bank.

Venturing into the realm of forecasting energy consumption time series in buildings, a unique study [12] employs machine learning (ML) methods. Real-time data from a smart grid in an experimental building form the basis for evaluating the effectiveness of both statistical and ML approaches. The study introduces a 'hybrid model' that intertwines forecasting and optimization techniques, revealing its superiority over single and ensemble models regarding accuracy, positioning it as a preferred choice for energy management planning. Concluding this expansive exploration, another study [13] directs attention toward the transition to autonomous responsive demand and cyber-physical energy systems, specifically within the context of smart homes and their Home Energy Management Systems (HEMS). Delving into smart HEMS's architecture and functional modules, the study reviews advanced HEMS infrastructures and smart home appliances. A comprehensive survey of renewable energy resources, including solar, wind, biomass, and geothermal energies, utilized in HEMS adds depth to the analysis. The investigation culminates by scrutinizing various home appliance scheduling strategies to reduce residential electricity costs and enhance energy efficiency derived from power generation utilities.

3 Material and Methods

3.1 Dataset Collection

Data acquisition is indispensable in realizing the functionality and advantages intrinsic to smart homes within these advanced living environments. Nevertheless, the delicate balance between innovation and privacy emerges as an overarching concern, accentuating the critical need to ensure data collection practices align with the guiding principles of transparency, security, and ethical responsibility. Establishing unambiguous and transparent data policies, robust security measures, and an empowering framework facilitating users' control over their data lays the groundwork for unlocking the complete potential of smart homes, concurrently upholding individual privacy and fundamental rights. Notably, the open-source and regularly updated Smart Home dataset, available on Kaggle [14], served as a valuable resource for this study. This dataset, acknowledged for its openness and accessibility, provides a comprehensive repository of information on various dimensions of smart home environments, encompassing appliance usage, environmental conditions, and occupancy patterns. The dataset's availability and accessibility contribute significantly to fostering research and development endeavors in smart homes.

3.2 Data Preprocessing

Preprocessing the data is necessary to ensure the accuracy of the regression model. To prepare the raw data for analysis, it is essential to employ several analytical methods that involve cleaning, purifying, and formatting the data. Further processing is required to enhance the efficacy of the regression model. The reason behind this is the presence of discrepancies, missing numbers, outliers, and noise in the raw data. All of these issues collectively add to the model's overall inefficiency. Many strategies were employed to carry out qualitative research. The strategy encompassed data completion, including essential rows and columns, aggregation according to specific attributes, and guaranteeing the distinctiveness of each occurrence of the selected characteristics [15,16].

3.3 Descriptive Analysis

The visualization depicts several facets of the case, encompassing diverse time series, patterns, seasonal variations, and residuals. The components are obtained by seasonal decomposition using an additive model with a set duration. The study reveals that the data demonstrates a recurring pattern throughout different seasons, which is supported by both figure 3 and the outcomes of the Kruskal-Wallis test.

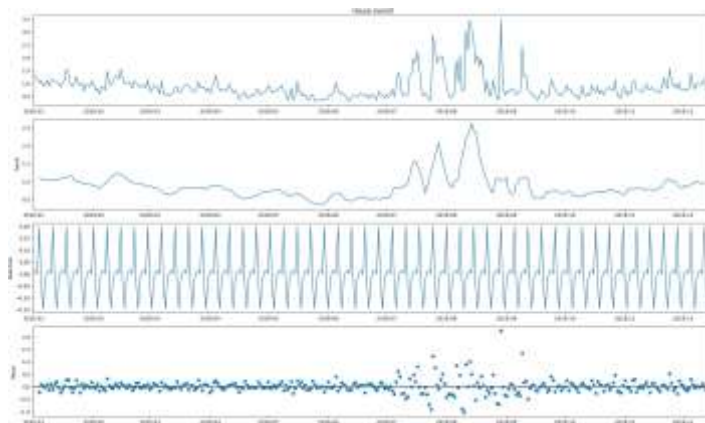


Figure 3: Seasonal Decompose.

The ARIMA technique, an acronym for autoregressive integrated moving average, is extensively employed to examine time series data [17, 18, 19]. Acknowledged for its adaptability, this highly flexible model finds application across diverse scenarios, encompassing predicting future events, identifying distinct patterns, and evaluating economic trends [20, 21]. Particularly well-suited for data manifesting patterns, seasonality, and intricate interconnections, the ARIMA model comprises three fundamental components: the autoregressive (AR) component, the moving average (MA) component, and the integrating (I) component. The autoregressive (AR) component denotes the correlation among historical time series data, while the moving average (MA) component reflects the influence of prior prediction errors [8]. The introduction of the differencing component, symbolized by "I," transforms the time series into a state of stationarity, ensuring constancy in its statistical characteristics over time [22].

The efficacy of the ARIMA model can be ascribed to its adeptness in capturing both temporal dependencies and the overarching trend or seasonality inherent in time series data [15]. Through the determination of appropriate values for the AR(p), I(d), and MA(q) parameters, the ARIMA model yields accurate forecasts, providing valuable insights into the underlying patterns and dynamics of the time series [4]. The fitting of an ARIMA model necessitates the identification of suitable values for the AR (p), I (d), and MA (q) parameters [8]. Techniques such as autocorrelation and partial autocorrelation plots facilitate the visualization of relationships between a time series and its past values for this purpose [17]. Furthermore, selecting the most suitable ARIMA model for a given context may involve using statistical metrics such as the Akaike Information Criterion or the Bayesian Information Criterion [20]. Recognized for its versatility, ARIMA is a valuable tool for academics, economists, and analysts spanning diverse disciplines [6]. Its capacity to capture temporal relationships and overall patterns or seasonality enables precise predictions, anomaly detection, and a deeper comprehension of complex time-series data, thereby facilitating well-informed decision-making [15].

3.4 Evaluation Metrics

The criteria used to evaluate predictions are displayed in Table 1. N represents the magnitude of the dataset under consideration. $(\hat{V}_n)$ and (V_n) respectively denote the estimated and true bandwidth values for the n -th iteration. The average of these estimated and observed values is denoted by $\bar{\hat{V}}_n$ and $\bar{V}_n$ [9]. Mean Absolute mistake (MAE) is a metric used to assess the effectiveness of statistical models. In contrast, Mean Squared Error (MSE) is a measure that quantifies the number of mistakes in statistical models.

The Root Mean Square Error ($RMSE$) is a statistical measure used to compare the prediction errors of different models. This evaluation also considers the relative root mean square error ($RRMSE$). The Pearson correlation coefficient is represented by " r ". The coefficient of determination, sometimes referred to as R squared, is a statistical measure that quantifies the proportion of the variance in the dependent variable that the independent variables can explain.

Table 1: Evaluation criteria for predictions.

Metric	Formula
RMSE	$\sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{V}_n - V_n)^2}$
RRMSE	$\frac{RMSE}{\frac{1}{N} \sum_{n=1}^N \hat{V}_n} \times 100$
MAE	$\frac{1}{N} \sum_{n=1}^N \hat{V}_n - V_n $
R^2	$1 - \frac{\sum_{n=1}^N (V_n - \hat{V}_n)^2}{\sum_{n=1}^N (\sum_{n=1}^N V_n - V_n)^2}$
r	$\frac{\sum_{n=1}^N (\hat{V}_n - \bar{\hat{V}})(V_n - \bar{V})}{\sqrt{(\sum_{n=1}^N (\hat{V}_n - \bar{\hat{V}})^2)(\sum_{n=1}^N (V_n - \bar{V})^2)}}$

4 Results

Table 2 presents a comparative analysis of the performance of the proposed Persistence and ARIMA models. The Root Mean Square Error (*RMSE*) for the proposed model is documented as 0.033778, while the Persistence model registers a higher *RMSE* value of 0.157909 upon evaluation with the test dataset. This noteworthy reduction in *RMSE* for the proposed ARIMA model underscores its superior effectiveness in time series forecasting and prediction tasks compared to the traditional Persistence model. In support of this finding, figure 4 illustrates the Residual Diagnostic Plot. Figure 5 displays the box plot, delineating the distribution of prediction error values for the proposed ARIMA model and the baseline algorithms. These visualizations offer a comprehensive means of comparing their respective performances, particularly in the context of forecasting accuracy.

Table 2: Comparison of the ARIMA Model and Persistence (Baseline) Algorithm.

	RMSE	MAE	r	R^2	RRMSE
Persistence	0.3035	0.194	0.388	0.1511	38.8413
ARIMA	0.263249087	0.174	0.374	0.1403	33.6862

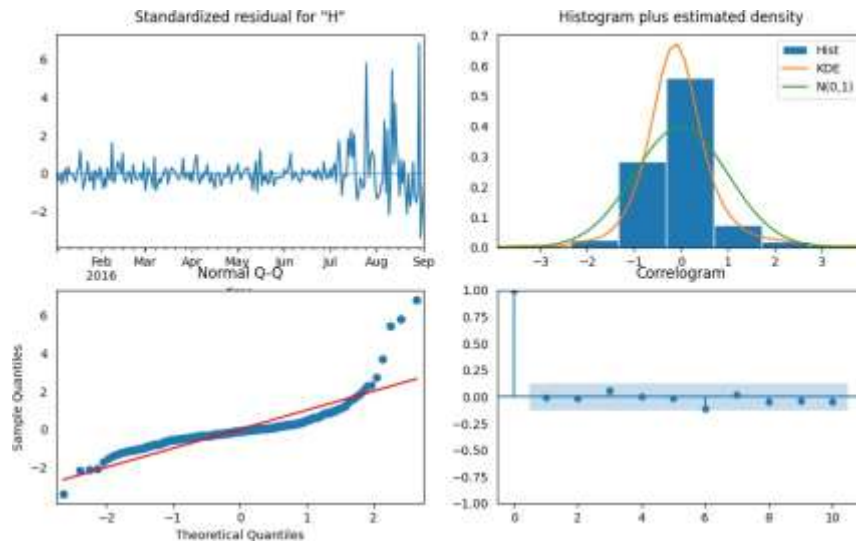


Figure 4: Residual Diagnostic Plot.

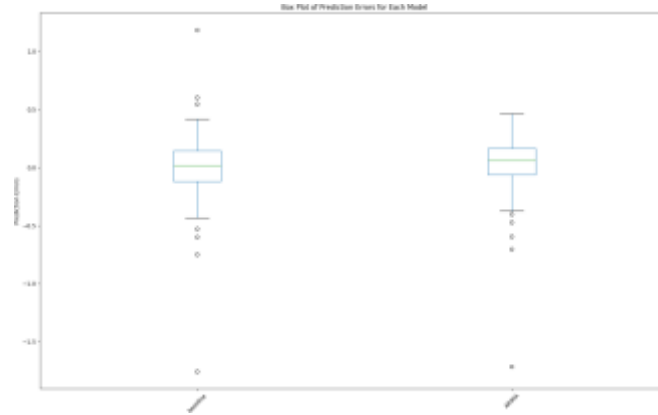


Figure 5: The Box Plots.

5 Conclusion

This comprehensive study delved into the efficacy of employing ARIMA (Autoregressive Integrated Moving Average) and persistence algorithms to forecast energy consumption in smart homes. Our investigation reveals that the proposed ARIMA model outshines the persistence algorithm, showcasing superior performance by attaining a significantly lower Root Mean Square Error (RMSE) value. This noteworthy outcome underscores the ARIMA model's capability to adeptly capture the intricate patterns and interdependencies inherent in the energy consumption data of smart homes. As a result, the forecasts generated by the ARIMA model demonstrate heightened accuracy, paving the way for a myriad of substantial benefits for smart home occupants.

The improved accuracy achieved by the ARIMA model holds profound implications for the residents of smart homes. Among these advantages is the potential for reduced energy costs, as the model's enhanced predictive capabilities contribute to more precise energy consumption forecasts. This heightened precision allows for more proactive maintenance strategies and contributes to the overall enhancement of occupant comfort within smart home environments. In the rapidly evolving landscape of smart homes, characterized by an increasing reliance on data-driven technologies, the role of sophisticated time series forecasting techniques, such as ARIMA, is set to become even more critical. These forecasting methodologies will play a pivotal role in optimizing energy consumption, ensuring efficient resource utilization, and ultimately contributing to an enriched and seamless smart home experience. As smart homes continue to advance and integrate cutting-edge technologies, the significance of accurate and advanced forecasting models becomes paramount in steering the trajectory of energy management within these intelligent living spaces.

References

- [1] Leal-Junior, A., Avellar, L., Blanc, W., Frizera, A., & Marques, C. (2023). Opto-Electronic Smart Home: Heterogeneous Optical Sensors Approaches and Artificial Intelligence for Novel Paradigms in Remote Monitoring. *IEEE Internet of Things Journal*, 1–1. <https://doi.org/10.1109/JIOT.2023.3323481>
- [2] Abdelhamid, A. A., El-Kenawy, E.-S. M., Alrowais, F., Ibrahim, A., Khodadadi, N., Lim, W. H., Alruwais, N., & Khafaga, D. S. (2022). Deep Learning with Dipper Throated Optimization Algorithm for Energy Consumption Forecasting in Smart Households. *Energies*, 15(23), Article 23. <https://doi.org/10.3390/en15239125>
- [3] Felber, N. A., Tian, Y. J. (Angelina), Pageau, F., Elger, B. S., & Wangmo, T. (2023). Mapping ethical issues in the use of smart home health technologies to care for older persons: A systematic review. *BMC Medical Ethics*, 24(1), 24. <https://doi.org/10.1186/s12910-023-00898-w>
- [4] Gilani, S. M. M., Usman, M., Daud, S., Kabir, A., Nawaz, Q., & Judit, O. (2023). SDN-based multi-level framework for smart home services. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-023-15678-2>
- [5] Kulurkar, P., Dixit, C. kumar, Bharathi, V. C., Monikavishnuvarthini, A., Dhakne, A., & Preethi, P. (2023). AI based elderly fall prediction system using wearable sensors: A smart home-care technology with IOT. *Measurement: Sensors*, 25, 100614. <https://doi.org/10.1016/j.measen.2022.100614>
- [6] El-Kenawy, E.-S. M., Eid, M. M., Abdelhamid, A. A., Ibrahim, A., Takieldein, A. E., & Elkhaliq, S. H.

DOI: <https://doi.org/10.54216/JAIM.060201>

Received: May 13, 2023 Revised: August 02, 2023 Accepted: December 03, 2023

- A. (2022). Hybrid Particle Swarm and Gray Wolf optimization for Prediction of Appliances in Low-Energy Houses. 2022 International Telecommunications Conference (ITC-Egypt), 1–5. <https://doi.org/10.1109/ITC-Egypt55520.2022.9855690>
- [7] Li, J., Herdem, M. S., Nathwani, J., & Wen, J. Z. (2023). Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management. *Energy and AI*, 11, 100208. <https://doi.org/10.1016/j.egyai.2022.100208>
- [8] Cascone, L., Sadiq, S., Ullah, S., Mirjalili, S., Siddiqui, H. U. R., & Umer, M. (2023). Predicting Household Electric Power Consumption Using Multi-step Time Series with Convolutional LSTM. *Big Data Research*, 31, 100360. <https://doi.org/10.1016/j.bdr.2022.100360>
- [9] Priyadarshini, I., Alkhayyat, A., Gehlot, A., & Kumar, R. (2022). Time series analysis and anomaly detection for trustworthy smart homes. *Computers and Electrical Engineering*, 102, 108193. <https://doi.org/10.1016/j.compeleceng.2022.108193>
- [10] Bhoj, N., & Singh Bhadoria, R. (2022). Time-series based prediction for energy consumption of smart home data using hybrid convolution-recurrent neural network. *Telematics and Informatics*, 75, 101907. <https://doi.org/10.1016/j.tele.2022.101907>
- [11] Rocha, H. R. O., Honorato, I. H., Fiorotti, R., Celeste, W. C., Silvestre, L. J., & Silva, J. A. L. (2021). An Artificial Intelligence based scheduling algorithm for demand-side energy management in Smart Homes. *Applied Energy*, 282, 116145. <https://doi.org/10.1016/j.apenergy.2020.116145>
- [12] Chou, J.-S., & Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, 165, 709–726. <https://doi.org/10.1016/j.energy.2018.09.144>
- [13] Zhou, B., Li, W., Chan, K. W., Cao, Y., Kuang, Y., Liu, X., & Wang, X. (2016). Smart home energy management systems: Concept, configurations, and scheduling strategies. *Renewable and Sustainable Energy Reviews*, 61, 30–40. <https://doi.org/10.1016/j.rser.2016.03.047>
- [14] Smart Home Dataset. (n.d.). [dataset]. Retrieved December 12, 2023, from <https://kaggle.com/code/offmann/smart-home-dataset>
- [15] Cheng Fan, Meiling Chen, Xinghua Wang, Jiayuan Wang, and Bufu Huang. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9, March 2021.
- [16] El-Sayed M. El-kenawy, Abdelhameed Ibrahim, Nadjem Bailek, Kada Bouchouicha, Muhammed A. Hassan, Mehdi Jamei, and Nadhir Al-Ansari. Sunshine duration measurements and predictions in Saharan Algeria region: an improved ensemble learning approach. *Theoretical and Applied Climatology*, 147(3-4):1015–1031, November 2021.
- [17] Rob J. Hyndman and George Athanasopoulos. Forecasting with exponential smoothing: The state of the art. *Journal of Forecasting*, 37(3):334–354, 2018.
- [18] Spyros Makridakis, Michel Hibon, and David Harvey. The accuracy of major economic forecasts. *Journal of Business Economics and Statistics*, 42(1):1–11, 2020.
- [19] Theodore Spyropoulos and Eddie Maharaj. The role of arima models in time series forecasting: Recent developments and applications. *Journal of Business Research*, 145:828–847, 2021.
- [20] Peter J. Brockwell and Richard A. Davis. *Introduction to time series and forecasting*. Springer, 2016.
- [21] John D. Cryer and Kung Sik Chan. *Time series analysis with applications in R*. Springer, 2008.
- [22] El-Sayed M. El-Kenawy, Seyedali Mirjalili, Abdelaziz A. Abdelhamid, Abdelhameed Ibrahim, Nima Kho-dadadi, and Marwa M. Eid. Meta-heuristic optimization and keystroke dynamics for authentication of smartphone users. *Mathematics*, 10(16):2912, August 2022.