Predicting peak demand for electricity consumption using time series data and machine learning model

Suriya S., Agusthiyar R.
Department of Computer Science and Applications, SRM Institute of Science and Technology, Chennai, India

Article Info

Article history:

Received Jul 12, 2024
Revised Oct 3, 2024
Accepted Nov 10, 2024

Keywords:

ARIMA
Lasso
Multivariate
Ridge
Univariate
XGR regressor

ABSTRACT

Energy consumption is influenced by various factors, including the proliferation of electronic devices, technological advancements, economic growth, agricultural development, and population increase. Each of these factors contributes to the rising demand for energy. This paper addresses the challenge of predicting peak energy demand (ED) by utilizing historical time series data from the past five years, combined with temperature data from Tamil Nadu's official sources. We employed feature engineering techniques to prepare the data for machine learning models, specifically XGBoost regressor, lasso, and ridge regression. The time series data was then analyzed using both univariate and multivariate models, including auto regressive integrated moving average (ARIMA) and vector autoregressive (VAR) models. The results show that our models can effectively forecast ED, providing critical insights for policymakers and stakeholders involved in energy planning and resource management.

Corresponding Author:

Suriya S.
Department of Computer Science and Applications, SRM Institute of Science and Technology
Chennai, India
Email: suriyas@srmist.edu.in

1. INTRODUCTION

Accurate forecasting of peak energy demand (ED) is crucial for effective power resource management and infrastructure planning. In Tamil Nadu, like many other regions, ED fluctuations can significantly impact the efficiency of power supply systems. Inaccurate predictions can lead to over or under-utilization of resources, affecting both operational costs and energy reliability. The challenge lies in developing predictive models that can accurately capture complex patterns and trends in energy consumption.

Existing solutions for forecasting peak ED typically involve time series models and machine learning techniques. Univariate models, such as the auto-regressive (AR) integrated moving average (ARIMA) model, are commonly used due to their simplicity and ability to capture historical patterns. However, these models often struggle with irregularities and complex patterns, such as those induced by economic changes, policy impacts, or sudden disruptions. On the other hand, multivariate models like the vector autoregressive (VAR) model offer a more holistic approach by incorporating multiple variables, but they can be computationally intensive and require careful parameter tuning. The major constraints in forecasting peak demand include data quality issues, such as missing values and inconsistencies, and the challenge of capturing external factors that influence energy consumption.

One of the critical aspects of energy management is the prediction of peak ED, i.e., the maximum amount of energy required at any given time [1]-[7]. Accurate forecasting of peak demand enables utilities to optimize resource allocation, mitigate supply-demand imbalances, and avoid potential grid failures or blackouts. However, predicting peak demand is a complex task influenced by a myriad of factors, including weather patterns, socio-economic trends, and technological advancements. In this paper, we address the challenge of predicting peak ED in Tamil Nadu by leveraging advanced data analytics techniques, specifically machine learning and time series analysis [8]-[16]. Our approach involves integrating historical time series data on energy consumption with meteorological data obtained from the Tamil Nadu government's official websites. By combining these datasets, we aim to capture the intricate relationship between energy usage and weather conditions, which play a significant role in shaping consumption patterns. The integration of historical energy consumption data provides insights into past trends and patterns, allowing us to identify recurring seasonal variations, long-term growth trends, and other relevant factors influencing ED [17]-[25]. The use of Himawari-8 time series data for meteorological applications as in [26] highlights the importance of high-frequency and accurate weather data in predictive analytics. Similarly, the work in [27] demonstrate the efficacy of machine learning approaches in analyzing time series data within financial markets. Concurrently, the inclusion of meteorological data, particularly minimum and maximum temperatures, enables us to account for weather-induced fluctuations in energy consumption, such as increased demand for heating or cooling during extreme weather conditions. A key aspect of our methodology is feature engineering, wherein we transform raw data into meaningful features that can be effectively utilized by machine learning algorithms. Through careful selection and manipulation of features, we aim to enhance the predictive capability of our models, ensuring that they can capture the underlying patterns and dynamics inherent in the data. To explore the predictive potential of different modeling approaches, we employ a variety of machine learning algorithms, including XGBoost regressor, lasso, and ridge regression. These algorithms are chosen for their ability to handle regression tasks effectively and their capacity to accommodate complex relationships between input features and target variables. In addition to machine learning models, we also explore the realm of time series analysis, recognizing the temporal nature of our data. Specifically, we apply popular time series models such as ARIMA and VAR models to capture temporal dependencies and forecast future ED based on historical patterns.

2. LITERATURE REVIEW

Understanding how consumers' behavior influences their power usage is crucial, especially given the lack of comprehensive research in this area illuminated by the shortage of big data analytics (BDA) exploration focusing on the consumer side of the energy sector. By employing descriptive and insightful analysis on mined consumer data, the model can effectively illustrate, characterize, and forecast the power usage of individual households based on their profiles. Evaluation results have demonstrated that the model plays a pivotal role in providing essential information to help consumers track and monitor their energy consumption. Additionally, it aids them in effectively managing their energy usage, thus contributing to the maintenance of data accuracy and efficiency.The layers are information sources, information pre-arranging, information assessment, and information assignment centre and information show. The principle layer is data arrangement from both AMI advanced metering infrastructure (AMI) and Customers informational indexes. The customer informational collection involves coordinated data and AMI contains both coordinated and unstructured data. This will convert the data into the needed format for analysing purposes. Then Data cleansing and analysis need to be done with data analytics using both descriptive and predictive analysis. Now the processed data underwent to Data warehouse. The most used information distribution centre for enormous information execution is Hadoop disseminated document framework (HDFS). HDFS have enormous scope datasets across various hubs repetitively to handle failover and flawed occasions. A non-social data set. Openings for the model execution of the BDA that is focusing on the purchaser side of the energy region are abundant. The model, which utilizes clear and prescriptive examination targets furnishing the customers with the examination results that can help them to more readily deal with their family power utilization. Taking into account the evaluation performed, it tends to be induced that the proposed BDA model for family power use following and noticing is viewed as usable and prepared to achieve its objective [1]. In this paper, they planned one decision support system (DSS) framework that works inside an IoT biological system. This gives a progressed investigation of smart meter for upgrading the nature of information works on the expense forecast, utilization and that incorporates choice with noteworthy suggestions. The dataset has been taken from the commercial organization. The methodology given here is with Bayesian organization alongside three AI classifiers for expectation as Random backwoods, Naïve Bayes, and choice tree. Results display that our procedure delivers quantifiably basic evaluations and that the DSS will chip away to the detriment capability of ESM network exercises and support. Accuracy assesses the limit of explicit classifiers to accurately arrange unlabelled data. It presents the extent between the different requested data (for both misguided and write requests) and the amount of given data. Preliminaries performed on the dataset showed the efficiency and the adequacy of the proposed approach [2]. The geographically distributed data centerserve as an armature for cloud services. Which consumes a huge amount of electricity this leads to huge costs for operational services and it is a typical challenge for cloud computing. On the off chance that the power costs of server farms are anticipated ahead of time, the cloud supplier can diminish energy costs. An effective power value expectation is required to limit the power bill of GDCs. This paper proposes power value expectations for GDCs in multi-district power markets. The test is led on genuine power value informational collections with AI calculations. By relatively surveying the forecast precision of the models, the most exact one is chosen. Analysis results show that the forecast model can give promising precision. They presented the electricity bill prediction for nearly three sectors in the United States using some machine learning algorithm. This consists of the dataset for three years in each sector. It produces accurate results compared to the decision table and linear regressions. The analyses express the viability and execution accomplished by the chosen models considered for both machine learning calculation and yearly prepared informational index. They chose models to give promising precision to foreseeing the power cost of GDCs in multi-area power markets. The models can't foresee well the value spikes and instability. Later on, enhancements to the models, for example, the issue of predictability in the cost will be tended to [4]. Due to the huge extension of cities, and industrial, and commercial growth energy consumption is in high demand. Because of IoT and smart meters energy metering is possible in all places. For forecasting purposes, the machine learning algorithm will be used. This paper utilizes recurrent neural networks (RNN) to catch time conditions and proposes an original energy load determining system based for example age and sequence-to-sequence (S2S) profound learning calculation. The S2S design that is regularly utilized for language interpretation was adjusted for energy load determination. Analyses center around gated recurrent unit (GRU) based S2S models and long short-term memory (LSTM) based S2S models [3].

In this paper, the author proposed a deep-learning prediction method that considers the weather and the progressive rate. In this paper, they propose a deep learning-based hypothetical system for environmentally friendly single-family energy rates. They used the weighted mean square error to set up the proposed model for better accuracy. Finally, according to the proposed model, the expected energy consumption is converted into monthly energy charges [5].

In this paper, a hardware module has been generated, which is used to check the amount of energy consumed by each contraption in a house and, as a result, compares the month to the monthly power bill. This framework utilizes a calculation that sets the seeing speed of a connected contraption to the association is dependent upon the kind of gadget and subsequently. The ACS712 current sensor was used in this piece, and the microcontroller used to operate the stuff units was also an Arduino Ide. The force supply to the differentiating contraptions and regulators, among many other things, is furnished by a boost converter [6].

The application is sent to verify electric bills. The framework takes the commitment of relevant data and forecasts grid inspections for the upcoming months. It offers presumptions for objects to be used over an apparent timeframe based on its force usage and other factors [6]-[8]. The existing literature on energy consumption prediction often falls short in addressing the nuanced impact of consumer behaviour, relying on simplistic models and overlooking the complexities of individual decision-making processes. These studies often struggle with data availability, quality, and scalability, hindering their practical applicability and effectiveness.

3. METHOD

This study employs a combination of advanced statistical and data analysis techniques to forecast peak ED in Tamil Nadu. To achieve accurate and reliable predictions, we utilize the ARIMA model, the VAR model, and exploratory data analysis (EDA) to process and analyze the data. The ARIMA model is selected for its proven capability in time series forecasting, particularly for capturing underlying trends and seasonal patterns in energy consumption. Complementing this, the VAR model is employed to explore the complex interplay between multiple economic and environmental factors affecting ED, providing a holistic view of the influences driving consumption patterns.

3.1. ARIMA model

The ARIMA model serves as a fundamental predictive tool for analyzing time series data, particularly in forecasting peak ED in Tamil Nadu. ARIMA is adept at capturing underlying trends and patterns present in historical data, making it well-suited for analyzing energy consumption patterns over time. However, it may encounter challenges when confronted with intricate patterns such as those observed during pandemics or irregular data fluctuations.

The ARIMA model comprises three essential components: AR, integration, and moving average (MA). The AR component identifies patterns within the data series, aiding in the detection of recurring trends and behaviors. The Integration component detects underlying trends in the data, allowing for the identification of long-term patterns or shifts in energy consumption. Finally, the MA component deals with noise between data points, ensuring the accurate capture of the underlying signal amidst fluctuations.

The ARIMA model was employed as a prediction model for time series data due to its simplicity and utility in working with trends and patterns in historical data. However, it is limited in handling complex patterns such as pandemics and struggles with irregular data fluctuations. The model's formulation includes three key parts:

- AR component:

$$AR(p) = \varphi 1 * Y(t-1) + \varphi 2 * Y(t-2) + \ldots + \varphi p * Y(t-p)$$

where φ1, φ2, ..., φp are the AR coefficients, and Y(t-1), Y(t-2), ..., Y(t-p) are the lagged values of the time series.

- Integration (I) component:

$$I(d) = (1 - B)^d * Y(t)$$

where d is the differencing parameter, B is the backshift operator, and Y(t) is the original time series.

- MA component:

$$MA(q) = \theta 1 * \varepsilon(t-1) + \theta 2 * \varepsilon(t-2) + \ldots + \theta q * \varepsilon(t-q)$$

where θ1, θ2, ..., θq are the MA coefficients, and ε(t-1), ε(t-2), ..., ε(t-q) are the lagged residuals.

Both additive and multiplicative models are utilized, with additive models decomposing the time series into seasonal, trend, and recent components $(Y(t) = S(t) + T(t) + R(t))$, while multiplicative models provide an alternative perspective by multiplying these components $(Y(t) = S(t) * T(t) * R(t))$. Despite its limitations, the ARIMA model serves as a powerful analytical tool for capturing temporal dependencies and seasonality within energy consumption patterns. By leveraging both additive and multiplicative approaches, the ARIMA model facilitates a thorough examination of the various factors influencing ED, thereby enhancing the accuracy and robustness of predictive forecasts. Through a comprehensive analysis of historical data using ARIMA, stakeholders can gain valuable insights into energy consumption dynamics, enabling them to develop strategies for efficient resource allocation and infrastructure planning in Tamil Nadu.

3.2. VAR model

The VAR model serves as a pivotal tool in forecasting ED, particularly in regions like Tamil Nadu, India, where complex interactions between various economic and environmental factors influence energy consumption patterns. Unlike univariate models, which focus solely on a single variable, the VAR model incorporates multiple variables into the analysis, providing a more comprehensive understanding of the underlying dynamics driving ED. By comparing recent observations of a variable with previously observed values and historical observations of other variables within the system, the VAR model captures the intricate relationships between different factors influencing energy consumption.

VAR model formulation: the VAR model can be expressed as:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + ApY_t - p + \epsilon_t$$

where:

- Y_t is a vector of the variables being forecasted at time ttt,
- $A_1, A_2, \ldots, Ap$ are matrices of coefficients to be estimated,
- p is the number of lags in the model,
- ϵ_t is a vector of error terms, assumed to be white noise.

In the context of ED forecasting, the VAR model enables analysts to explore the interplay between energy consumption and various socio-economic indicators, such as GDP growth, population dynamics, industrial output (IO), and environmental factors. For example, changes in GDP growth rates can impact ED, as higher economic activity often translates into increased energy usage across sectors like manufacturing, transportation, and residential consumption. Additionally, policy decisions, such as changes in interest rates or government incentives for renewable energy adoption, can also influence ED patterns. By incorporating these diverse variables into the analysis, the VAR model offers a holistic framework for understanding the complex dynamics of energy consumption.

Consider a VAR model with three variables: ED, GDP, and IO. The model can be written as:

$$\begin{vmatrix} ED_t \\ GDP_t \\ IO_t \end{vmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{vmatrix} ED_t \\ GDP_t \\ IO_t \end{vmatrix} + \begin{vmatrix} \epsilon\ ED_t \\ \epsilon\ GDP_t \\ \epsilon\ IO_t \end{vmatrix}$$

Here, a_{ij} represents the coefficients that capture the relationships between the variables. The error terms $\epsilon\ ED_t$, $\epsilon\ GDP_t$, and $\epsilon\ IO_t$ represent shocks or innovations to the variables at time t.

3.3. EDA

EDA plays a crucial role in the research on forecasting peak ED in Tamil Nadu using the VAR model. EDA involves the initial exploration and visualization of the multivariate time series data to gain insights into the underlying patterns, relationships, and potential outliers. In the context of this research, EDA focuses on understanding the characteristics of the variables included in the VAR model, identifying any trends or seasonality, and assessing the correlation between energy consumption and other relevant factors such as GDP growth, population dynamics, IO, and environmental factors.

$$Correlation\ Coefficient(r) = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 (\sum x)^2)(n\sum y^2 - (\sum y)^2}}$$

It includes identifying and handling outliers or anomalies in the data that may affect the model's performance. Outlier detection techniques such as box plots, z-scores, and clustering algorithms help researchers identify observations that deviate significantly from the expected patterns.

$$z - score = \frac{(X-\mu)}{\sigma}$$

It involves examining the presence of missing values in the data and implementing strategies for imputing missing values if necessary. Missing value imputation techniques such as mean imputation, median imputation, or interpolation help ensure the completeness of the dataset.

$$Mean\ Imputation = \frac{\sum X}{n}$$

EDA explores the need for data transformation techniques such as normalization or standardization to ensure the comparability of variables with different scales or units. Data transformation techniques help improve the performance and interpretability of the VAR model.

$$Standardization = \frac{(X-\mu)}{\sigma}$$

The analysis concludes with exploratory visualization of the relationships between energy consumption and other variables using techniques such as scatter plots, line plots, and correlation matrices. Exploratory visualization helps researchers identify potential patterns, trends, and dependencies that can inform the VAR model specification and interpretation.

3.4. Feature engineering

Feature engineering is a crucial step in the research on forecasting peak ED in Tamil Nadu using the VAR model. Feature engineering involves transforming raw data into informative features that enhance the predictive performance of the model. In the context of this research, feature engineering focuses on extracting relevant variables or creating new features from the multivariate time series data to capture important patterns, trends, and relationships related to energy consumption and other factors influencing demand. Below are key aspects of feature engineering in this research:

Variable selection: feature engineering begins with selecting the variables to be included in the VAR model based on their relevance to ED forecasting. Variables such as GDP growth, population dynamics, IO, environmental factors, and policy indicators are considered based on their potential impact on energy consumption patterns.

$$X = \{GDP, Population, Industrial\ Output, Environmental\ Factors, Policy\ Indicators\}$$

- Normalization and standardization: feature engineering involves normalizing or standardizing the variables to ensure that they are on the same scale and have comparable magnitudes. Normalization techniques such as min-max scaling or standardization techniques such as z-score normalization are applied to improve the numerical stability and convergence of the VAR model.

$$X' = \frac{X - min(X)}{max(X) - min(X)}$$

$$X'' = \frac{(X - \mu)}{\sigma}$$

- Feature selection: feature engineering includes selecting the most relevant features to be included in the final VAR model. Feature selection techniques such as forward selection, backward elimination, or recursive feature elimination are employed to identify the subset of features that contribute most to the model's predictive performance.

$$X_{selected} = Select\ Features(X, criterion)$$

- Feature transformation: feature engineering may involve transforming the variables using mathematical transformations such as logarithmic transformation or square root transformation to improve their distributional properties and capture nonlinear relationships more effectively.

$$X_{transformed=}\sqrt{X}$$

4. RESULT AND DISCUSSION

The experimental setup encompasses data collection, preprocessing, feature engineering, model selection, and evaluation. The dataset, sourced from the official website of the Tamil Nadu government, spans from 2018 to 2022 and includes variables such as energy consumption, weather patterns, and economic indicators. The dataset undergoes preprocessing to eliminate null values, impute missing data, and standardize the features. The results represent the year-wise peak demand fluctuation along with the upper subdivision peak demand analysis, detailing a significant drop in peak demand to 6,000 units in the year 2018 and again in the sixth month of 2022. This analysis highlights critical periods of reduced demand, suggesting potential factors such as economic changes, policy impacts, or external disruptions. Based on historical trends and current data patterns, it is likely that the predicting peak demand may increase in the upcoming years, necessitating proactive measures in resource planning and infrastructure development. Figure 1 illustrates the year-wise fluctuation in peak demand, with a notable decrease to 6,000 units in 2018 and a similar drop in the sixth month of 2022. This drop indicates periods of reduced ED, which could be attributed to various factors such as economic changes, policy impacts, or external disruptions. The analysis suggests that while there have been significant decreases, trends from the historical data imply that peak demand might increase in the future. This insight is crucial for planning resources and developing infrastructure to meet future ED efficiently.

Peak demand is shown to be influenced by minimum temperatures, with higher temperatures leading to increased power consumption. The graphical representation highlights the relationship between temperature and peak demand, demonstrating that as temperatures rise, so does power usage. The SMA helps smooth out short-term fluctuations and highlight longer-term trends in ED. By calculating the SMA, we can better understand the underlying trends and predict future demand more accurately. Figure 2 presents a comprehensive visualization of peak demand data across various features.

Table 1 shows the VAR model analysis, presenting various performance metrics. The mean absolute error (MAE) is 150.25, indicating the average magnitude of errors between predicted and actual values. The root mean squared error (RMSE) is 180.42, reflecting the model's prediction accuracy by penalizing larger errors more heavily. The mean absolute percentage error (MAPE) is 7.82%, illustrating the prediction errors relative to actual values. Additionally, the 95% confidence interval (CI) ranges from 14802 to 18411.7, indicating the range within which the true peak demand values are expected to fall, providing a measure of the model's reliability and precision.

As shown in Table 1, the VAR model's MAE is 150.25, indicating the average magnitude of errors amongst forecast and real values. The RMSE of 180.42 measures the regular degree of errors, by lower values signifying improved model performance. The MAPE of 7.82% quantifies the percentage difference between predicted and actual values, providing a relative measure of accuracy. Additionally, the 95% CI ranges from 14802 to 16607 to 18411.7, indicating the range within which the true peak demand values are likely to fall with 95% confidence.

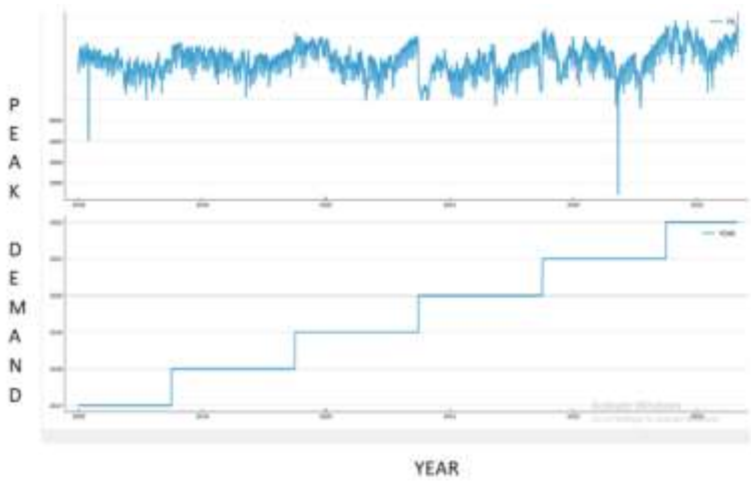

Figure 1. Year wise peak demand

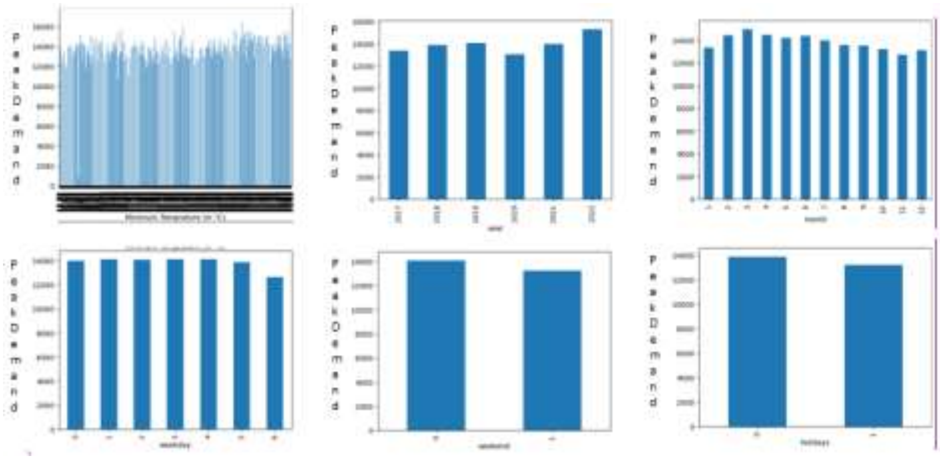

Figure 2. Visualising data using various features

Table 1. VAR model analysis

Metric	Value
MAE	150.25
RMSE	180.42
MAPE	7.82%
Confidence interval (95%)	[14802, 16607, 18411]

Table 2 shows the ARIMA model analysis, highlighting key performance metrics. The MAE is 120.50, indicating the average error magnitude between predicted and actual values. The RMSE is 140.67, showing the model's prediction accuracy by giving more weight to larger errors. The MAPE is 6.28%, reflecting the prediction errors as a percentage of actual values. The 95% CI ranges from 14500 to 17800.8, suggesting the range within which the true values are likely to fall, thus providing insight into the model's reliability and precision.

Table 2. ARIMA model analysis

Metric	Value
MAE	120.50
RMSE	140.67
MAPE	6.28%
Confidence interval (95%)	[14500, 16200, 17800.8]

Similarly, for the ARIMA model as in Table 2, the MAE is 120.50, representing the average magnitude of errors between predicted and actual values. The RMSE of 140.67 measures the typical magnitude of errors, by lower values representing improved model performance. The MAPE of 6.28% provides a relative measure of accuracy. The 95% CI ranges from 14500 to 16200 to 17800.8, indicating the range within which the true peak demand values are likely to fall with 95% confidence.

5. CONCLUSION

In this study, we employed advanced time series forecasting models, including VAR and ARIMA, to predict peak ED in Tamil Nadu. Through meticulous data preprocessing, feature engineering, and model training, our aim was to provide accurate and reliable forecasts to support energy planners and policymakers in efficient resource allocation and infrastructure planning. The experimental evaluation demonstrated the efficacy of both the VAR and ARIMA models in predicting peak ED. While both models showed promising results, the ARIMA model exhibited superior performance with lower MAE, MAPE, and RMSE values compared to the VAR model. Additionally, the ARIMA model produced tighter CI, indicating higher accuracy and precision in peak demand forecasting. Furthermore, the CI generated by the VAR model, ranging from 14802 to 16607 to 18411.7 with 95% accuracy, provided valuable insights into the uncertainty associated with the predictions. This information is crucial for decision-makers to assess the reliability of forecasts and make informed choices regarding energy resource management. Moreover, the integration of various machine learning models with time series data holds significant potential in predicting future peak demand and addressing challenges such as excess power supply and power theft. By leveraging advanced modeling techniques and data-driven approaches, stakeholders can optimize energy distribution networks, reduce inefficiencies, and promote sustainability in the energy sector.

REFERENCES

[1] S. Suriya and D. Agusthiyar, "Comparative study based on the consumption of electricity bill for the Indian states with various machine learning algorithms," *Journal of the Balkan Tribological Association*, vol. 29, no. 2, pp. 211-218, 2023.

[2] R. Dollah and H. Aris, "A big data analytics model for household electricity consumption tracking and monitoring," *2018 IEEE Conference on Big Data and Analytics (ICBDA)*, Malaysia, 2018, pp. 44-49, doi: 10.1109/ICBDAA.2018.8629769.

[3] J. Siryani, B. Tanju and T. J. Eveleigh, "A machine learning decision-support system improves the internet of things' smart meter operations," *in IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 1056-1066, 2017, doi: 10.1109/JIOT.2017.2722358.

[4] M. M. Than and T. Thein, "Electricity price prediction for geographically distributed data centers in multi-region electricity markets," *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*, Nagoya, Japan, 2018, pp. 89-93, doi: 10.1109/CCOMS.2018.8463272.

[5] S. A. Awate, R. Chaurasia, P. Kirolkar, M. More, P. Ingale, "Electricity consumption using big data analytics," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 3, no. 2, 2023.

[6] J. Park, M. Kim, S. Hong, Y. Jeung and I. -H. suh, "Prediction of individual household energy bills using deep learning," *2020 35th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Nagoya, Japan, 2020, pp. 500-504.

[7] A. Pandoh, A. S. Wasekar, S. Sarkar and A. B. Thakur, "Smart electricity meter monitoring and prediction using iSocket," *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2021, pp. 211-216, doi: 10.1109/ICACCS51430.2021.9442011.

[8] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.

[9] A. Paszke *et al*., "Automatic differentiation in PyTorch," *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017.

[10] M. K. M. Shapi, N. A. Ramli, and L. J. Awalin, "Energy consumption prediction by using machine learning for smart building: Case study in Malaysia," *Developments in the Built Environment*, vol. 5, p. 100037, 2021, doi: 10.1016/j.dibe.2020.100037.

[11] S. R. Rallapalli, S. Ghosh, "Forecasting monthly peak demand of electricity in India—A critique," *Energy Policy*, vol. 45, pp. 516-520, 2012, doi: 10.1016/j.enpol.2012.02.064.

[12] S. Amosedinakaran, K. Mala, A. Bhuvanesh and P. Marishkumar, "Electricity demand forecasting using differential evolution algorithm for Tamil Nadu," *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, Tirunelveli, India, 2020, pp. 143-146, doi: 10.1109/ICOEI48184.2020.9142872.

[13] A. M. Mirlatifi, F. Egelioglu, U. Atikol, "An econometric model for annual peak demand for small utilities," *Energy*, vol. 89, pp. 35-44, 2015, doi: 10.1016/j.energy.2015.06.119.

[14] S. A. Neves, A. C. Marques, and J. A. Fuinhas, "On the drivers of peak electricity demand: what is the role played by battery electric cars?," *Energy*, vol. 159, pp. 905-915, 2018, doi: 10.1016/j.energy.2018.06.209.

[15] H. Prasetyo, Y. T. Tularsih, F. Pandansari, S. B. Prasetya, A. Priyandono and A. S. Dermawan, "Design of power monitoring system based on internet of things (IoT) with calibration interface," *2021 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia, 2021, pp. 1-5, doi: 10.1109/ICISS53185.2021.9533239.

[16] A. Ed-daoudy and K. Maalmi, "A new internet of things architecture for real-time prediction of various diseases using machine learning on big data environment," *Journal of Big Data*, vol. 104, 2019, doi: 10.1186/s40537-019-0271-7.

[17] P. Deligiannis, S. Koutroubinas and G. Koronias, "predicting energy consumption through machine learning using a smart-metering architecture," in *IEEE Potentials*, vol. 38, no. 2, pp. 29-34, 2019, doi: 10.1109/MPOT.2018.2852564.

[18] L. Bottaccioli *et al.*, “Building energy modelling and monitoring by integration of IoT devices and building information models,” *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, Turin, Italy, 2017, pp. 914-922, doi: 10.1109/COMPSAC.2017.75.

[19] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, “Big data in healthcare: management, analysis and future prospects,” *Journal of Big Data*, vol. 54, 2019, doi: 10.1186/s40537-019-0217-0.

[20] F. Firouzi, B. Farahani, M. Ibrahim and K. Chakrabarty, “Keynote paper: from EDA to IoT ehealth: promises, challenges, and solutions,” in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 12, pp. 2965-2978, Dec. 2018, doi: 10.1109/TCAD.2018.2801227.

[21] J. Siryani, B. Tanju and T. J. Eveleigh, “A machine learning decision-support system improves the internet of things’ smart meter operations,” in *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 1056-1066, 2017, doi: 10.1109/JIOT.2017.2722358.

[22] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz and J. Lopez, “A survey of iot-enabled cyberattacks: assessing attack paths to critical infrastructures and services,” in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3453-3495, Fourthquarter 2018, doi: 10.1109/COMST.2018.2855563.

[23] P. Sundaravadivel, K. Kesavan, L. Kesavan, S. P. Mohanty and E. Kougianos, “Smart-log: a deep-learning based automated nutrition monitoring system in the IoT,” in *IEEE Transactions on Consumer Electronics*, vol. 64, no. 3, pp. 390-398, Aug. 2018, doi: 10.1109/TCE.2018.2867802

[24] D.-Ly Vu, T.-Kha Nguyen, T. V. Nguyen, T. N. Nguyen, F. Massacci, P. H. Phung, “HIT4Mal: hybrid image transformation for malware classification,” *Transactions on Emerging Telecommunications Technologies*, 2019, doi: 10.1002/ett.3789.

[25] H. Zhang, Q. Zhang, J. Liu and H. Guo, “Fault detection and repairing for intelligent connected vehicles based on dynamic bayesian network model,” in *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2431-2440, 2018, doi: 10.1109/JIOT.2018.2844287.

[26] A. L. Hadiyanto, K. Wikantika, A. S. Prihatmanto, N. J. Trilaksono, and D. Irawadi, “Generating himawari-8 time series data for meteorological application,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 29, no. 2, pp. 780-787, 2023, doi: 10.11591/ijeecs.v29.i2.pp780-787.

[27] A. K. Bitto *et al.*, “CryptoAR: scrutinizing the trend and market of cryptocurrency using machine learning approach on time series data,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*,” vol. 28, no. 3, pp. 1684-1696, 2022, doi: 10.11591/ijeecs.v28.i3.pp1684-1696.

BIOGRAPHIES OF AUTHORS

Suriya S. is an assistant professor in the Department of Computer Applications at SRM Institute of Science and Technology (SRMIST). With a specialization in machine learning and the Internet of Things (IoT), Suriya brings a wealth of knowledge and expertise to her teaching and research endeavors. She is currently pursuing his Ph.D. in Computer Science at SRM University, furthering her research in cutting-edge technological advancements. Suriya holds a Master of Computer Applications (MCA) degree from Anna University, awarded in 2014, and a Bachelor of Science (BSc) in Information Technology from Manonmaniam Sundaranar University, awarded in 2011. Her academic and professional journey is marked by a commitment to innovation and excellence in the field of computer science. She can be contacted at email: suriyas@srmist.edu.in.

Dr. Agusthiyar R. is a professor and the head of the Department of Computer Applications at SRM Institute of Science and Technology, Ramapuram Campus. He specializes in data mining and artificial intelligence, areas where he has made significant contributions through his research and teaching. Dr. Agusthiyar earned his Ph.D. in Data Mining from Anna University in 2017, and his M.Phil. in Computer Science from Pride, Periyar University, Salem, in 2008. He also holds an MCA in Computer Applications from Madurai Kamaraj University, awarded in 2001, and a B.Sc. in Mathematics from Dr. Zakir Hussain College, Ilayangudi, affiliated to Madurai Kamaraj University, awarded in 1998. In his academic career, Dr. Agusthiyar has taught various courses including data mining and warehousing and software engineering. His research interests lie in data mining, artificial intelligence, and machine learning, where he continues to explore and innovate. As the Head of the Department, he plays a crucial role in shaping the curriculum and guiding the department towards academic excellence. He can be contacted at email: hod.dca.rmp@srmist.edu.in.