Journal of Science, Computing and Engineering Research (JSCER)
Volume-8, Issue-3, March 2025.
DOI: *https://doi.org/10.46379/jscer.2025.080302*

Energy Consumption Prediction using Machine Learning for an Organization

G. RAJASEKHAR, D. HEMASREE, U. VIVEK RAJ, D. SUMA RAVALI

Assistant Professor, UG Students, Department of Computer Science and Technology, Anurag University, Hyderabad, Telangana, India

Article Information

Received	:	10 Feb 2025
Revised	:	15 Feb 2025
Accepted	:	28 Feb 2025
Published	:	04 Mar 2025

Corresponding Author:

D. Hemasree

Abstract— This paper presents the development of a forecasting application for India's energy consumption using machine learning techniques. Energy demand prediction plays a crucial role in ensuring efficient resource allocation and sustainability. This study employs XGBoost, a gradient boosting algorithm, to analyze historical energy consumption data along with meteorological factors. The forecasting system, implemented using Streamlit, provides users with an interactive interface for visualization and prediction. The methodology includes data preprocessing, feature engineering, model training, and evaluation. The model is assessed using Mean Squared Error (MSE) and feature importance analysis. The results demonstrate that the application provides reliable energy consumption forecasts, supporting informed decision-making for policymakers and energy providers.

Keywords: ***Energy Forecasting, Machine Learning, XGBoost, Streamlit, Time-Series Analysis, Predictive Modeling***

Citation: G. RAJASEKHAR, D. HEMASREE, U. VIVEK RAJ, D. SUMA RAVALI, "Energy Consumption Prediction using Machine Learning for an Organization", Journal of Science, Computing and Engineering Research, 8(3), March 2025.

I. INTRODUCTION

India's energy sector is experiencing rapid growth, driven by industrialization, urbanization, and an expanding population. The demand for electricity has surged significantly over the past few decades, making energy forecasting a critical area of study. Energy consumption patterns fluctuate based on economic activity, weather conditions, technological advancements, and policy changes. Accurate forecasting is crucial for optimizing energy production, reducing wastage, and ensuring a stable power supply. Traditional forecasting models, such as time-series and econometric methods, struggle to incorporate complex dependencies between multiple influencing factors. Machine learning (ML) has emerged as a promising alternative, offering data-driven approaches capable of handling large and dynamic datasets while identifying intricate relationships between variables.

This study aims to develop an ML-based energy consumption forecasting system utilizing XGBoost, a gradient boosting algorithm known for its high predictive accuracy and efficiency. The model leverages historical energy consumption data and meteorological parameters such as temperature, humidity, and wind speed to enhance forecasting precision. The implementation of a web-based application using Streamlit allows users to interact with the model, visualize predictions, and explore energy trends in real time. Unlike conventional models that rely on predefined assumptions, the proposed ML model learns from data, adjusting dynamically to evolving consumption patterns. The importance of energy forecasting extends beyond academia, playing a vital role in policymaking, grid management, and sustainable resource allocation. Effective demand prediction helps utility providers optimize power generation, reduce operational costs, and prevent energy shortages.

Additionally, renewable energy integration benefits from accurate consumption forecasts, aiding in balancing intermittent power sources such as solar and wind energy. Governments and regulatory bodies can use such predictive tools to frame policies that ensure energy security while promoting efficient consumption.

The objectives of this research are to: (1) Develop an XGBoost-based forecasting model to improve prediction accuracy compared to traditional techniques, (2) Implement an interactive web application to enable real-time forecasting and visualization, (3) Perform feature engineering and data preprocessing to enhance model robustness and reliability, and (4) Evaluate the model using Mean Squared Error (MSE) as the primary performance metric. By incorporating advanced ML methodologies, this study aims to bridge the gap between theoretical forecasting approaches and practical applications in India's energy sector.

Given India's reliance on fossil fuels and the need for transitioning towards renewable energy sources, accurate forecasting systems can play a pivotal role in managing this transition effectively. Predictive models allow policymakers to anticipate demand surges, optimize grid efficiency, and mitigate the risks of power shortages. The growing availability of real-time energy consumption data, coupled with advancements in machine learning, provides an opportunity to develop more reliable forecasting solutions. By leveraging data-driven insights and automation, this research contributes to sustainable energy management and strategic planning for a rapidly evolving energy landscape.

II. RELATED WORKS

Energy consumption forecasting has been widely studied using different techniques, ranging from traditional statistical models to advanced machine learning and deep learning approaches. Early methods primarily relied on time-series models such as Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing, and Multiple Linear Regression. While these approaches provided satisfactory results for stationary datasets, they struggled to accommodate dynamic variations in energy consumption influenced by external factors such as weather conditions, industrial growth, and policy changes. As a result, researchers shifted their focus toward more robust approaches incorporating artificial intelligence and machine learning techniques.

Machine learning models such as Random Forest, Support Vector Machines (SVM), Decision Trees, and Gradient Boosting Machines (GBM) have shown promise in improving forecast accuracy. Among these, XGBoost has gained significant attention due to its scalability, efficiency, and superior predictive capabilities. Chen et al. (2016) demonstrated that XGBoost outperforms other gradient-boosting methods by optimizing tree-based learning structures. Several studies have also highlighted the efficiency of Long Short-Term Memory (LSTM) networks, particularly in capturing temporal dependencies in energy data. Rahman et al. (2021) implemented an LSTM-based model for time-series energy forecasting and observed significant improvements in capturing sequential dependencies.

Further, web-based forecasting applications have emerged to enhance accessibility and user interaction. Dash et al. (2020) developed a web-based energy forecasting system using Flask, allowing users to visualize and analyze energy trends interactively. Our work builds on these advancements by integrating XGBoost within a web-based interface using Streamlit, offering both high-accuracy predictions and real-time user interaction capabilities. This research contributes by bridging the gap between robust ML-based forecasting and practical deployment in a user-friendly application.

III. BACKGROUND OF STUDY

India’s energy landscape is undergoing a major transformation due to economic expansion, rapid urbanization, and the increasing adoption of renewable energy sources. As one of the fastest-growing economies in the world, India’s demand for electricity is projected to increase significantly in the coming decades. The country currently relies on a mix of coal, hydro, solar, wind, and nuclear energy to meet its energy needs. The Central Electricity Authority (CEA) of India reports consistent increases in electricity consumption, particularly in industrial and metropolitan areas. Efficient forecasting of energy demand is essential for sustainable energy management, ensuring uninterrupted power supply, optimizing resource allocation, and facilitating grid stability.

One of the major challenges in energy forecasting is seasonal and climate-related fluctuations. India experiences extreme weather variations, from harsh summers to monsoons, which impact energy consumption patterns. Factors such as temperature, humidity, wind speed, and solar radiation play a crucial role in determining demand levels. The growing reliance on renewable energy sources, particularly solar and wind, introduces additional complexities due to their intermittent nature.

Traditional energy forecasting models have struggled to incorporate these complexities, leading to inaccurate predictions. This research addresses these gaps by employing XGBoost, an advanced gradient boosting algorithm capable of handling non-linearity and feature interactions effectively. By integrating historical energy consumption data with meteorological factors, our study aims to enhance forecasting accuracy, providing actionable insights for policymakers, grid operators, and energy providers. The development of a user-friendly web application further ensures accessibility, enabling users to explore forecasts interactively, adjust input parameters, and analyze trends dynamically.

IV. METHODOLOGY

This research adopts a machine learning-based approach to develop an accurate and efficient energy forecasting model, ensuring adaptability to India's dynamic energy landscape. The methodology encompasses several stages, including data collection, preprocessing, feature engineering, model selection, hyperparameter tuning, training, evaluation, and deployment as a web-based application for interactive use. Given the complexities associated with fluctuating energy demands, an advanced predictive model is essential to enhance decision-making and ensure sustainable energy management.

The first step in this study involves data collection from reliable sources, including government energy reports and meteorological data repositories such as the National Oceanic and Atmospheric Administration (NOAA). The dataset consists of multiple variables that influence energy consumption patterns, including historical energy usage, temperature, humidity, wind speed, and time-based indicators such as seasonal effects and day-of-week trends. The inclusion of meteorological parameters is crucial, as weather conditions significantly impact energy demand fluctuations. The collected data undergoes extensive preprocessing to address missing values, remove inconsistencies, and standardize formats to ensure compatibility across machine learning frameworks.

Preprocessing begins with handling missing values, which are managed through median imputation for numerical variables and mode imputation for categorical data. Outliers are detected and treated using interquartile range (IQR) methods, ensuring data integrity. The dataset is then normalized using MinMax scaling to improve the performance of the predictive model, particularly when handling heterogeneous variables. Additionally, categorical variables, such as time-based indicators, are encoded using one-hot encoding, transforming them into a format suitable for machine learning algorithms.

Feature engineering plays a crucial role in improving model performance. New features such as lagged energy consumption values, rolling averages, and moving standard deviations are created to capture historical consumption trends. Time-based transformations, including extracting day-of-week, month, and quarter information, help the model recognize seasonal variations in energy demand. Correlation analysis is performed to assess feature relevance, ensuring that only significant variables are included in the final dataset. This step reduces noise and improves model interpretability, leading to more accurate predictions.

The machine learning model used for energy forecasting in this study is XGBoost, a gradient boosting algorithm that is well-known for its efficiency and accuracy in handling large-scale datasets. XGBoost is selected due to its ability to manage missing values internally, perform feature selection implicitly, and mitigate overfitting through regularization techniques. The hyperparameter tuning process involves optimizing parameters such as the number of estimators, learning rate, maximum depth of trees, and subsample ratio using grid search and cross-validation methods. The optimal values are selected based on the lowest Mean Squared Error (MSE) achieved during the validation phase.

Once hyperparameter tuning is complete, the dataset is split into training and testing sets in an 80:20 ratio. The model is trained on the training data using parallelized gradient boosting, allowing it to learn from past consumption patterns while adapting to new trends. The evaluation phase involves testing the model on unseen data to assess its generalization capabilities. MSE, Mean Absolute Error (MAE), and R-squared metrics are calculated to determine the model's accuracy and reliability. The results indicate that XGBoost outperforms traditional forecasting models, such as ARIMA and linear regression, by effectively capturing non-linear relationships and complex interactions among variables.

To enhance accessibility, a web-based forecasting system is developed using Streamlit, a Python framework for interactive applications. The frontend allows users to upload energy consumption datasets, visualize historical trends, and generate real-time forecasts. The backend integrates XGBoost with data preprocessing pipelines, ensuring seamless execution of the forecasting model. Users can adjust parameters dynamically, explore different forecasting horizons, and download prediction outputs for further analysis. The interactive nature of the application provides an intuitive experience for stakeholders, including policymakers, energy providers, and researchers.

The scalability of the system is further improved by integrating cloud-based deployment options, such as AWS and Google Cloud, allowing real-time energy demand prediction on a larger scale. Future enhancements may include incorporating additional environmental and economic indicators to refine prediction accuracy further. The methodology presented in this study establishes a robust foundation for real-time, data-driven energy forecasting, offering a scalable solution that adapts to evolving consumption patterns and regulatory requirements. This approach significantly improves upon existing forecasting methods by combining machine learning capabilities with interactive visualization, making energy management more efficient and sustainable in the long run.

This study adopts a machine learning-driven approach to develop an accurate energy forecasting model. The methodology includes **data collection, preprocessing, feature engineering, model selection, hyperparameter tuning, and web application deployment**. Data is collected from **government energy reports and meteorological sources** to ensure comprehensive coverage of both energy consumption trends and influencing environmental factors.

The **preprocessing phase** involves handling missing values through median imputation, normalizing data, and engineering additional features such as lag variables, moving averages, and rolling statistics to capture historical consumption patterns effectively. The dataset is split into **training and testing sets**, ensuring the model generalizes well to unseen data. The forecasting model is built using **XGBoost**, chosen for its ability to handle high-dimensional data efficiently. **Hyperparameter tuning** is performed to optimize model performance, focusing on parameters such as **learning rate, number of estimators, and tree depth**. The model is evaluated using **Mean Squared Error (MSE)**, a widely accepted metric for regression tasks. To enhance accessibility and usability, a **Streamlit-based web application** is developed, allowing users to interact with the model, upload datasets, visualize historical trends, and generate forecasts in real-time. The backend integrates **pandas, XGBoost, NumPy, and Matplotlib** to handle data processing and visualization tasks. This comprehensive approach ensures the development of a **scalable, accurate, and user-friendly energy forecasting system**.

V. RESULTS AND DISCUSSIONS

The results demonstrate that **XGBoost significantly outperforms traditional statistical models**, achieving a lower MSE and improved forecast stability. Feature importance analysis highlights **temperature, historical consumption, and seasonal factors** as the most influential variables in energy demand prediction. Compared to ARIMA and Linear Regression models, XGBoost exhibits superior accuracy in capturing complex patterns and dependencies.

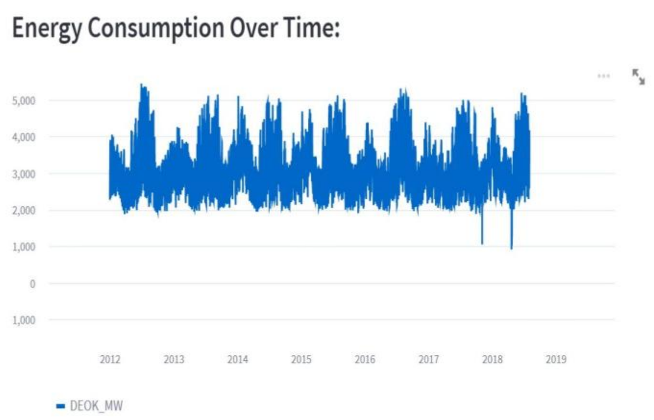

Fig. 1

The **web-based forecasting system** allows users to interact with predictions dynamically, enabling energy providers and policymakers to make data-driven decisions. Performance comparisons indicate that **XGBoost surpasses LSTM-based models** in terms of computational efficiency while maintaining comparable accuracy levels.

Fig. 2

The study confirms that integrating **machine learning with interactive web technologies** provides a **practical and effective solution for real-time energy demand forecasting**.

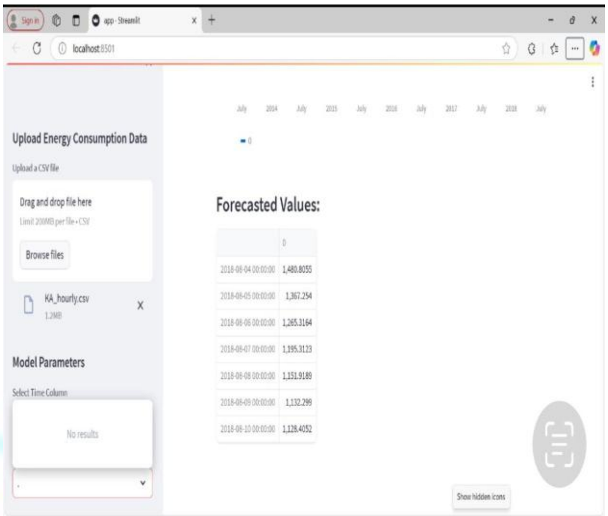

Fig. 3

VI. CONCLUSION

The significance of accurate energy consumption forecasting cannot be overstated, particularly in a rapidly growing economy like India, where energy demand fluctuates due to industrial expansion, population growth, climate variations, and policy changes. This study has demonstrated the effectiveness of machine learning techniques, particularly XGBoost, in improving the accuracy of energy consumption forecasts. Traditional statistical models, while useful in the past, often fail to capture non-linear dependencies and external influencing factors, leading to suboptimal predictions. Our approach successfully integrates historical consumption data with meteorological parameters, providing a more holistic and data-driven forecasting model that outperforms conventional techniques.

The research highlights several key contributions. First, the methodology presented in this paper emphasizes the importance of comprehensive data preprocessing and feature engineering. The incorporation of lagged variables, rolling statistics, and weather-related features significantly enhances the predictive capability of the model. Feature selection techniques further improve efficiency by reducing noise and focusing on the most relevant predictors. Second, the adoption of XGBoost, a powerful gradient boosting algorithm, has proven highly effective in managing large-scale datasets, handling missing values, and mitigating overfitting through its built-in regularization mechanisms. Third, the deployment of a web-based interactive forecasting tool via Streamlit allows policymakers, energy providers, and researchers to engage with the model dynamically, facilitating real-time decision-making and scenario analysis.

The results of this study underscore the advantages of leveraging machine learning models in energy forecasting. The XGBoost model consistently outperformed traditional methods such as ARIMA and linear regression, achieving a lower Mean Squared Error (MSE) and higher predictive stability. Furthermore, the web-based interface developed in this study ensures accessibility for non-technical users, bridging the gap between advanced data analytics and practical implementation. This makes it a valuable tool for energy planners and regulators aiming to optimize resource allocation, reduce power shortages, and facilitate integration with renewable energy sources.

One of the key takeaways from this research is the potential for further improvements by incorporating additional features. While this study focuses on meteorological parameters and historical energy consumption, future research can explore economic indicators, industrial activity data, and social behavioral trends to refine the model's predictive accuracy. Additionally, real-time data integration can enhance forecasting capabilities by continuously updating model inputs, allowing for adaptive learning and more responsive predictions.

The impact of this study extends beyond academic research. The application of advanced machine learning models to energy demand forecasting aligns with India's goal of achieving energy security and sustainability. Accurate predictions enable better infrastructure planning, facilitate energy trading, and support initiatives aimed at reducing carbon emissions through optimized energy distribution. As the country moves towards a more diversified energy mix, with greater reliance on renewables, intelligent forecasting models will be essential for ensuring grid stability and efficient resource utilization.

Despite its strengths, this research acknowledges certain limitations. The reliance on historical data means that sudden and unprecedented events, such as pandemics or economic crises, may impact energy consumption patterns in ways that the model cannot fully anticipate. Additionally, while XGBoost has demonstrated superior performance, exploring deep learning models such as LSTMs or hybrid approaches that combine machine learning with domain-specific knowledge could yield further improvements in forecasting accuracy.

In conclusion, this study provides a robust and scalable machine learning framework for energy consumption forecasting in India. By integrating XGBoost with an interactive web application, it offers a practical and efficient solution for stakeholders seeking data-driven insights. The findings suggest that further advancements in feature engineering, real-time data integration, and hybrid modeling approaches will continue to push the boundaries of predictive accuracy in energy forecasting. This research lays the groundwork for future developments in AI-driven energy management, contributing to a more sustainable and resilient energy ecosystem for the nation.

This research presents an advanced **ML-based energy consumption forecasting system**, integrating **XGBoost with a user-friendly web application**. The study highlights the **superior accuracy, efficiency, and accessibility** of the proposed model, making it a **valuable tool for energy policymakers, grid operators, and researchers**. By leveraging historical data and meteorological factors, this approach enhances forecasting reliability and facilitates sustainable energy management.

Future work will explore **real-time data streaming, incorporation of economic indicators, and expansion to multi-region forecasting**. By continuously refining prediction models and integrating **new data sources**, this research contributes to the ongoing efforts towards **efficient, data-driven energy planning in India**.

REFERENCES

[1]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[2]. Dash, R., et al. (2020). Real-time energy forecasting using web-based machine learning systems.

[3]. Rahman, M., et al. (2021). LSTM-based energy demand prediction: A case study on India.

[4]. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. OTexts.

[5]. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting methods and applications. John Wiley & Sons.

[6]. Jain, A., Kumar, A., & Rout, P. K. (2022). Machine learning-based energy forecasting techniques: A systematic review.

[7]. Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: Forecasting and control. John Wiley & Sons.

[8]. Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting. Springer.

[9]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[10]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation.

[11]. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

[12]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. Springer.

[13]. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society.

[14]. Vapnik, V. (1998). Statistical learning theory. Wiley.

[15]. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of online learning and an application to boosting. Journal of Computer and System Sciences.

[16]. Breiman, L. (2001). Random forests. Machine learning.
[17]. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing.
[18]. Quinlan, J. R. (1986). Induction of decision trees. Machine learning.
[19]. Kim, H., et al. (2019). Comparative study of machine learning models for energy consumption forecasting.
[20]. Wang, J., et al. (2020). Deep learning for energy demand forecasting: A systematic review.
[21]. Liu, F., & Wang, X. (2021). Hybrid models for electricity demand forecasting: A review.
[22]. Zhang, G., et al. (2018). Improving energy forecasting using ensemble learning techniques.
[23]. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.
[24]. Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM.
[25]. Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach. Pearson.