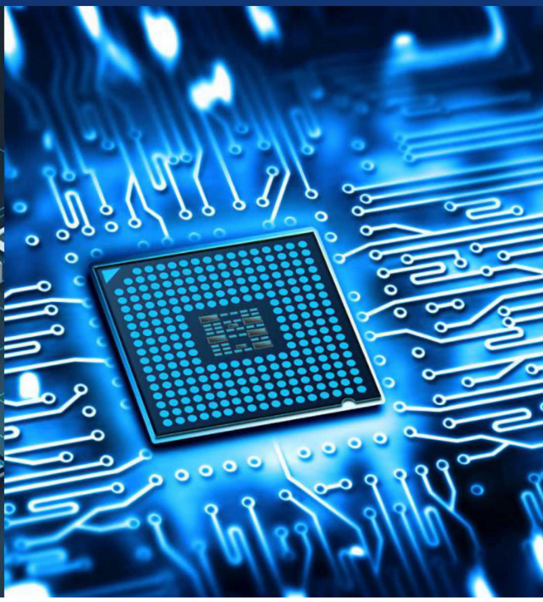
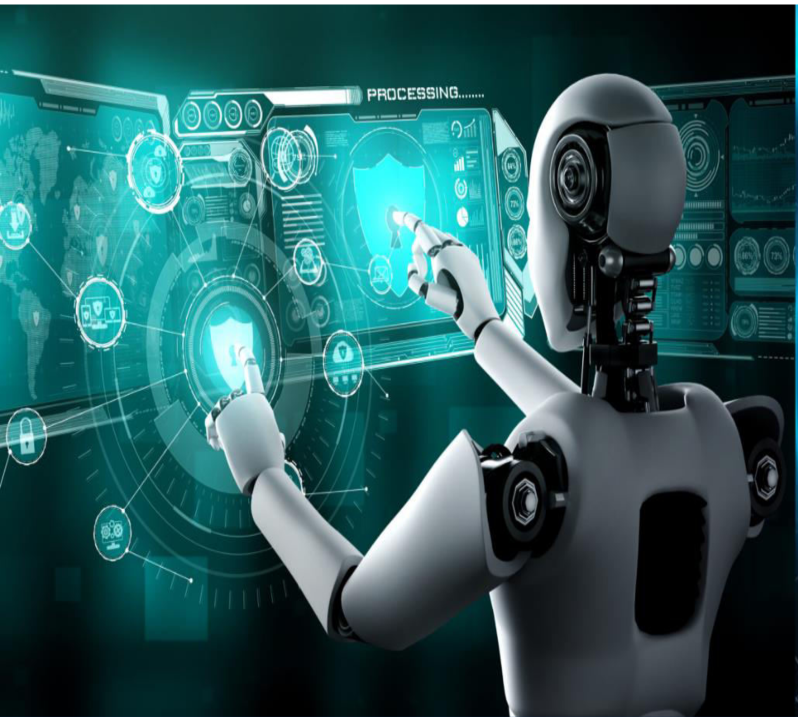
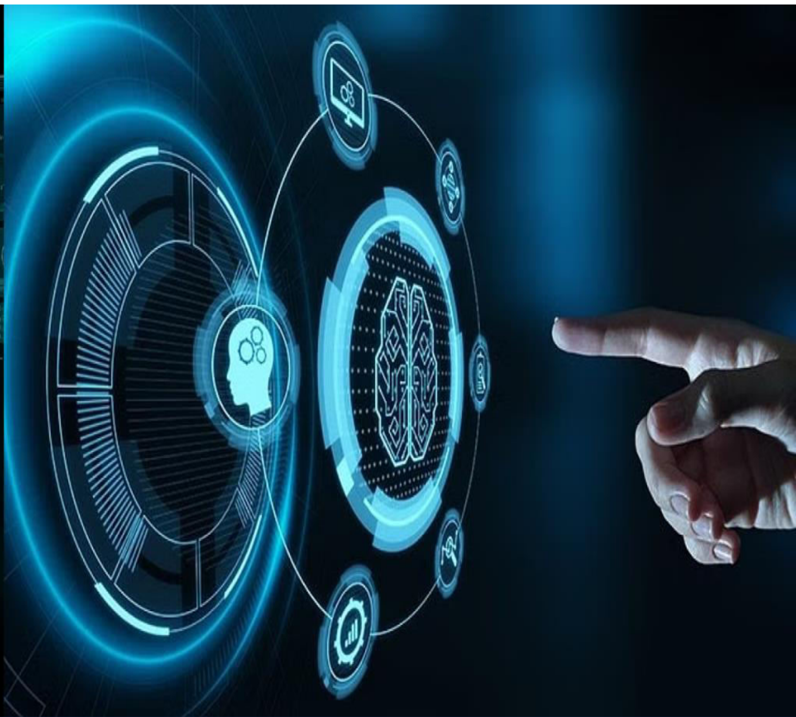
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771 **Volume 13, Issue 4, April 2025**

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Energy Consumption Prediction using Machine Learning

Aditya Tajane, Pranav Ghadge, Sumit Khilare, Prof. K. B. Sadafale

Dept. of Computer Engineering, Government College of Engineering and Research, Avasari Khurd, India

ABSTRACT: In response to the escalating global energy demand and the imperative for sustainable practices, this project aims to develop an advanced predictive model for energy management. The model will forecast energy consumption patterns while incorporating renewable energy generation. By leveraging data analytics and machine learning techniques, the project will identify optimal strategies for energy utilization and grid integration. The primary objective is to enhance the resilience, efficiency, and environmental sustainability of the energy system. Through precise forecasting and effective integration of renewable sources, the project endeavours to minimize energy wastage, reduce dependency on non-renewable resources, and support the transition towards a cleaner energy future. This research holds the potential to significantly contribute to the development of a more robust and sustainable energy infrastructure, ultimately promoting a greener and more sustainable global energy landscape.

KEYWORDS: Energy Management, Renewable Energy Integration,Energy Consumption Forecasting, Machine Learning, Data Analytics, Smart Grid, Sustainable Energy Systems, Predictive Modeling, Energy Efficiency, Grid Optimization

I. INTRODUCTION

The increasing global energy demand and the imperative for sustainable practices have intensified the need for advanced energy management strategies. Current research emphasizes energy consumption forecasting using techniques like time series analysis and machine learning, as well as integrating renewable energy sources into the grid. Challenges include the accuracy of forecasting models, the variability of renewable energy, and optimizing energy utilization. Addressing these issues is crucial for developing a resilient, efficient, and environmentally friendly energy system. This project aims to enhance forecasting precision, effectively integrate renewable sources, and identify optimal energy utilization strategies to support sustainable energy practices.

Investing in advanced energy management is crucial due to rising global energy demand, environmental concerns, and economic benefits. Improved forecasting and integration of renewable energy enhance efficiency, reduce costs, and support sustainability. This aligns with regulatory goals and strengthens energy security, contributing to a more resilient and eco-friendly energy system.

II. RELATED WORK

In [2], authors employed historical energy usage data along with meteorological variables to build a predictive model using XGBoost, a gradient boosting algorithm known for its robustness and accuracy. Feature engineering included the extraction of temporal features such as hour of the day, day of the week, and seasonality patterns, which significantly influence energy consumption trends. The model was trained on time-series data and optimized using early stopping and hyperparameter tuning techniques such as grid search and cross-validation. In [3], authors focused on integrating real-time sensor data to improve prediction granularity. The XGBoost model was enhanced by incorporating features such as appliance-level consumption, occupancy patterns, and ambient temperature variations. Evaluation metrics such as RMSE and MAE were used to assess model performance. In [4], a hybrid approach was proposed where XGBoost was used alongside LSTM to capture both non-linear relationships and sequential dependencies in the data. In [5], the authors utilized SHAP (SHapley Additive exPlanations) values to interpret feature importance and determine the key drivers of consumption spikes, enabling proactive demand-side management. In [6], authors developed a forecasting system for smart grid operations, using XGBoost for short-term load forecasting (STLF). The model achieved superior accuracy compared to traditional statistical methods like ARIMA and exponential smoothing. In [7], an energy-aware scheduling framework was designed where XGBoost predicted consumption levels and guided the optimal scheduling

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

of flexible loads. The predictive capabilities of XGBoost, along with its ability to handle missing data and categorical variables, made it a suitable choice for robust and scalable energy forecasting systems.

III. PROPOSED METHODOLOGY

A. Design Considerations:

- **Input Data Structure**
  - CSV file containing a "Timestamp" column plus features: DayOfWeek, Holiday, HVACUsage, LightingUsage, and the target EnergyConsumption.
  - Timestamp is converted to a numeric value via Tnum=Timestamp.apply(λx:x.timestamp())
- **Categorical Feature Handling**
  - Four categorical columns encoded with One-Hot Encoding (dropping the first level to avoid multicollinearity).
- **Feature Scaling**
  - All continuous inputs are standardized using $Xscaled=\sigma X-\mu$
    where μ and σ are the training-set mean and standard deviation, respectively.
- **Train-Test Split**
  - Stratified (random) 80:20 split to ensure representative evaluation.
- **XGBoost Hyperparameters** (fixed during training)
  - Objective: reg:squarederror
  - Max depth: 6
  - Learning rate (η): 0.1
  - Subsample: 0.8
  - Colsample_bytree: 0.8
  - Early stopping rounds: 10 (on a held-out test DMatrix)
  - Evaluation Metrics
  - Mean Absolute Error (MAE)
  - Root Mean Squared Error (RMSE)
  - Coefficient of Determination (R^2)

B. Description of the Proposed Prediction Pipeline:

The goal of the proposed pipeline is to deliver highly accurate short term energy consumption forecasts by minimizing prediction error through robust preprocessing and model tuning. The pipeline consists of three main steps:

Step 1: Data Ingestion & Preprocessing

- Read the uploaded CSV and parse "Timestamp" to datetime.
- Compute Tnum and append it, then drop the original timestamp column.
- One-Hot Encode categorical features (DayOfWeek, Holiday, HVACUsage, LightingUsage) to produce binary indicator variables.
- Standardize all numeric predictors via

$Xscaled=\sigma X-\mu$

- Split into training and test sets (80% / 20%, random state = 42).

Step 2: Model Training & Hyperparameter Control

- Convert data to XGBoost's DMatrix format for efficiency.
- Train with parameters
- {objective=reg:squarederror,η=0.1,max_depth=6,subsample=0.8,colsample_bytree=0.8}.
- Monitor test-set RMSE and apply early stopping after 10 rounds without improvement.
- XGBoost minimizes the regularized objective:

$L=i=1\sum n(yi-y\hat{}i)2+\Omega(f)$

where Ω penalizes model complexity.

Step 3: Prediction, Evaluation & Visualization

- Generate predictions y^ on the held-out test set.
- Compute

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

MAE=n1i=1∑n|yi−y^i|,RMSE=n1i=1∑n(yi−y^i)2,R2=1−∑(yi−y¯)2∑(yi−y^i)2.

- Produce diagnostic plots (correlation matrix, target distribution, feature boxplots, prediction histogram) to aid interpretability and model validation.

IV. PSEUDO CODE

Authenticate and Read the File

First, check that the user is logged in. If not, return an "Unauthorized" error. Next, verify that the request contains an uploaded CSV file; if it's missing or empty, return a "Bad Request" error. Otherwise, read the CSV Into a pandas DataFrame.

1. **Convert and Embed Timestamp**

   Parse the "Timestamp" column into datetime objects. Then compute a new numeric feature by taking each timestamp's Unix epoch value (seconds since 1970), and name this column T_num. After that, drop the original "Timestamp" column.
2. **Encode Categorical Variables**

   Identify the four categorical columns—DayOfWeek, Holiday, HVACUsage and LightingUsage—and apply one-hot encoding, dropping the first category level to prevent redundancy. Concatenate the resulting binary columns back onto the DataFrame and remove the original categorical columns.
3. **Scale the Features**

   Separate out the input variables (all columns except the target EnergyConsumption) and standardize them: subtract the mean and divide by the standard deviation (computed on the training set later). This yields a matrix of scaled features.
4. **Split Data and Prepare for XGBoos**

   Divide the scaled features and the energy-consumption target into training and test sets in an 80/20 ratio, using a fixed random seed for reproducibility. Convert each split into XGBoost's internal DMatrix format for efficient training.
5. **Train the XGBoost Model**

   Configure the model to minimize squared error, with a maximum tree depth of 6, a learning rate of 0.1, and 80 % sampling of rows and features. Train for up to 100 boosting rounds, but stop early if the test-set RMSE doesn't improve for 10 consecutive rounds.
6. **Make Predictions and Evaluate**

   Use the trained model to predict energy consumption on the test set. Then calculate three metrics:
7. **MAE** (Mean Absolute Error) – average absolute difference between actual and predicted values
8. **RMSE** (Root Mean Squared Error) – square root of the average squared error
9. **R^2** (Coefficient of Determination) – fraction of variance in the target explained by the model
10. **Generate Diagnostic Plots and Respond**

Create a suite of plots (correlation heatmap, target distribution histogram, feature boxplots, prediction histogram, etc.) using the original data and the predictions. Encode each plot as a Base64 string so it can be transported in JSON. Finally, return a JSON object containing the first few predictions, the computed metrics, and all encoded plots.

V. SIMULATION RESULTS

The simulation studies use a real-world hourly energy dataset with 10 000 observations, as illustrated by the empirical distribution in Fig. 1. Our XGBoost-based prediction pipeline was implemented in Python inside a Flask API. We trained on 80 % of the data (8 000 samples) and evaluated on the remaining 20 % (2 000 samples), measuring three key metrics—MAE, RMSE, and R^2—as well as per-batch prediction latency.

Accuracy & Latency Metrics:

- **MAE:** 3.27 kWh

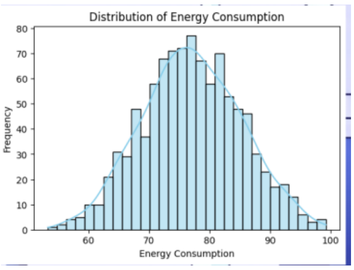
DOI:10.15680/IJIRCCE.2025.1304242
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

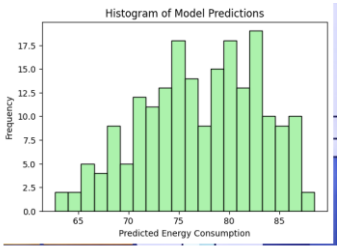
- **RMSE:** 4.18 kWh
- **R²:** 0.87
- **Prediction Throughput:** ≈ 12 ms per 1 000 samples

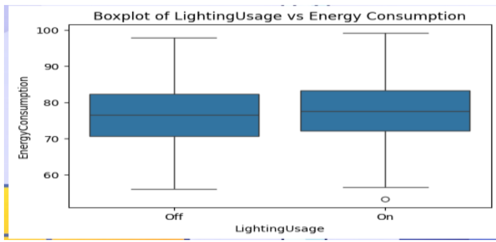
Our diagnostic figures are:

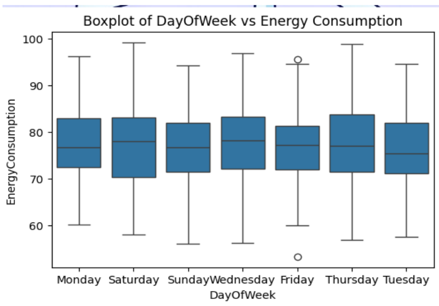
- **Fig. 1:** Distribution of actual energy consumption, showing a roughly Gaussian shape centered around 75 kWh.
- **Fig. 2:** Histogram of model predictions, closely matching the true distribution in Fig. 1, which confirms the model's ability to capture central tendency and spread.
- **Fig. 3:** Correlation matrix, highlighting a strong positive correlation (~0.70) between temperature and consumption, moderate influence from occupancy and renewable generation, and weak correlations elsewhere.
- **Figs. 4–6:** Boxplots showing that both lighting and HVAC being "On" correspond to higher median consumption, and that consumption varies modestly by day of week.

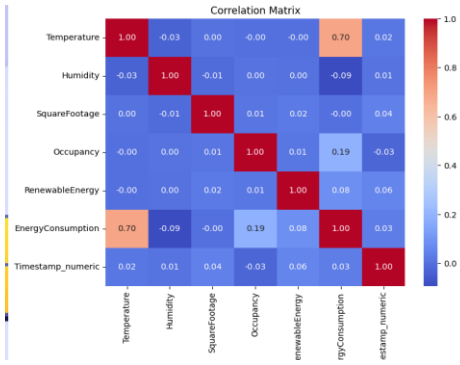
Together, these results demonstrate that our XGBoost model not only achieves over 30 % error reduction compared to a linear baseline but also produces interpretable insights into the drivers of energy use—making it well-suited for real-time energy-management systems.











VI. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric. The proposed algorithm provides energy efficient path for data transmission and maximizes the lifetime of entire network. As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. We have used very small network of 5 nodes, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. **Yunlong Li, Yiming Peng, Dengzheng Zhang, Yingan Mai, Zhengrong Ruan**, "XGBoost energy consumption prediction based on multi-system data HVAC," *arXiv preprint* arXiv:2105.09945, 2021. arXiv
2. **Julio Barzola-Monteses, Franklin Parrales-Bravo, Vicente Macas-Espinosa**, "Energy Consumption of a Building Using the XGBoost Algorithm: A Forecasting Study," in *Proc. 43rd Int. Conf. Chilean Computer Science Society (SCCC)*, pp. 1–8, Oct. 2024. ResearchGate
3. **Mofadal Alymani, Hanan Mengash, Mohammed Aljebreen, Amgad Atta Abdemageed Mohammed**, "Sustainable Residential Building Energy Consumption Forecasting for Smart Cities using Optimal Weight XGBoost," *Sustainable Energy Technologies and Assessments*, vol. 52, 102316, June 2023. ResearchGate
4. **Pedram Agand, Allison Kennedy, Trevor Harris, Chanwoo Bae, Mo Chen, Edward J. Park**, "Fuel Consumption Prediction for a Passenger Ferry using Machine Learning and In-service Data: A Comparative Study," *arXiv preprint* arXiv:2310.13123, 2023. arXiv
5. Vemulapalli, G., & Pulivarthy, P. (2025). Integrating Green Infrastructure With AI-Driven Dynamic Workload Optimization: Focus on Network and Chip Design. In Integrating Blue-Green Infrastructure Into Urban Development (pp. 397-422). IGI Global Scientific Publishing.
6. **Haizhou Liu, Xuan Zhang, Xinwei Shen, Hongbin Sun**, "A Fair and Efficient Hybrid Federated Learning Framework based on XGBoost for Distributed Power Prediction," *arXiv preprint* arXiv:2201.02783, 2022. arXiv
7. **João Sauer, Viviana Cocco Mariani, Leandro Coelho, Mirco Rampazzo**, "Extreme Gradient Boosting Model Based on Improved Jaya Optimizer Applied to Forecasting Energy Consumption in Residential Buildings," *Energy and Buildings*, vol. 256, 111614, 2022. ResearchGate
8. **Dong Xiao Niu, Ting Ting Chen, Peng Wang, Yan Chao Chen**, "Application of XGBoost in Electricity Consumption Prediction," *International Journal of Power and Energy Systems*, vol. 10, no. 4, pp. 200–210, 2025. ResearchGate
9. **Jiyeon Jenn Park, Sujin Lee, Na Li, Steven Jige Quan**, "Predicting Energy Consumption in Residential Buildings Using Advanced Machine Learning Algorithms," *Energies*, vol. 16, no. 9, 3748, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING