



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 11, November 2025

Power Consumption Analysis for Household

S. Kiran, N. Srihari, G. Hemanth, T. Rakesh

Department of CSE (AI&DS), Kallam Haranadhareddy Institute of Technology, Guntur, India

ABSTRACT: This project focuses on analyzing and predicting household electricity consumption using machine learning techniques to promote energy efficiency and sustainability. A time-series dataset containing parameters such as global active power, voltage, global intensity, and sub-metering values was utilized for analysis. Data preprocessing steps, including cleaning, timestamp conversion, and normalization, ensured data quality and modeling accuracy. Exploratory Data Analysis (EDA) revealed that energy consumption typically peaks during morning and evening hours due to high appliance usage. A Linear Regression model was implemented to forecast short-term power consumption based on historical trends and contextual features. The model's performance, evaluated through MAE, RMSE, and R^2 metrics, demonstrated strong predictive accuracy. Results indicate that factors such as time of day and appliance type significantly influence overall energy demand. Visualizations further highlighted daily and seasonal consumption patterns, offering insights into efficient usage behaviors. The proposed framework provides a simple yet effective approach to interpreting household energy data. By transforming raw data into actionable insights, this study supports smarter energy planning, reduced costs, and sustainable living practices.

KEYWORDS: Power consumption, household energy, linear regression, machine learning, data analysis, time series forecasting, smart home, sustainability.

I. INTRODUCTION

In today's data-driven era, efficient energy management has become a vital aspect of sustainable living. With the rapid increase in electricity consumption, understanding household power usage patterns is essential for reducing costs and minimizing environmental impact. Residential sectors account for a significant share of total energy demand, making optimization at the household level increasingly important. Traditional monitoring and manual analysis methods are often inefficient and fail to capture detailed behavioral patterns in energy usage. The integration of data analytics and machine learning provides new opportunities to analyze and predict power consumption effectively. This research focuses on examining household electricity usage using time-series data and predictive modeling techniques. The study employs Linear Regression to explore relationships between variables such as voltage, intensity, and sub-metered appliance readings. Comprehensive data preprocessing, including cleaning and feature extraction, ensures the dataset is structured for accurate analysis. Exploratory Data Analysis (EDA) helps identify key consumption periods, daily patterns, and peak load hours. Model evaluation using MAE, RMSE, and R^2 metrics confirms strong predictive capability and reliability. Ultimately, this study supports efficient energy planning, enabling households to make data-driven decisions that promote cost savings and sustainable energy use.

II. RELATED WORK

With the growing demand for sustainable energy systems, numerous studies have focused on understanding and forecasting household electricity consumption. Researchers have utilized time-series data and statistical methods to identify power usage trends and seasonal variations. Deb et al. (2017) emphasized the importance of granular smart meter data in detecting peak load hours and improving energy efficiency. Similarly, Yildiz et al. (2019) compared multiple machine learning algorithms and highlighted the effectiveness of Linear Regression for its interpretability and accuracy. These studies demonstrate that even simple predictive models can yield actionable insights for residential energy management.

Rajasekaran et al. (2020) explored the impact of temporal features such as hour, day, and season in improving prediction accuracy for power consumption datasets. Their work confirmed that time-based feature engineering significantly enhances model performance and reduces forecasting errors. Mocanu et al. (2016) combined historical energy and weather data to perform short-term load forecasting, revealing how external factors like temperature and

occupancy influence consumption patterns. Such research underscores the importance of feature selection and preprocessing for achieving reliable predictions in household energy analysis.

Recent advancements have expanded toward integrating smart grid data, IoT-based monitoring, and machine learning algorithms for real-time energy optimization. Studies by Huang and Zhang (2021) and Kandil et al. (2022) demonstrated that Linear Regression and ensemble models could effectively forecast power usage and identify high-consumption appliances. These approaches provide interpretability, low computational cost, and scalability, making them ideal for household applications. Building on these insights, the present work applies Linear Regression to model and predict residential energy usage, aiming to provide interpretable results that support smart home energy planning and sustainability.

III. DATASET AND PROBLEM DEFINITION

The dataset used for household power consumption analysis is sourced from publicly available repositories such as the UCI Machine Learning Repository and Kaggle, featuring detailed measurements of electrical parameters recorded at one-minute intervals. The dataset includes key attributes such as Date, Time, Global Active Power, Global Reactive Power, Voltage, Global Intensity, and Sub-metering 1, 2, and 3, which represent energy usage by different household appliances.

Dataset Description

Power consumption modeling heavily depends on the quality, frequency, and diversity of recorded household energy data. The dataset utilized in this study offers high-resolution time-series information, capturing minute-by-minute variations in household electricity usage. Each record reflects instantaneous power readings across different electrical circuits, providing insight into appliance-level and total energy demand. The dataset's structure supports both exploratory data analysis (EDA) for identifying patterns and supervised learning tasks for prediction. This dataset is ideal for understanding the relationship between key electrical features — such as voltage, intensity, and reactive power — and total active power consumption. The richness of this data enables the analysis of daily, weekly, and seasonal trends, offering valuable insights into energy efficiency and load forecasting. By leveraging this dataset, the study bridges the gap between raw electrical measurements and meaningful interpretations for smart energy management.

Problem Definition

The central problem addressed in this study is the accurate prediction and analysis of household electricity consumption based on historical time-series data. Power usage in residential environments varies significantly due to multiple factors such as time of day, appliance activity, occupancy patterns, and voltage fluctuations. While modern smart meters collect massive volumes of consumption data, interpreting this data effectively to forecast energy demand and identify inefficiencies remains a key challenge.

Traditional statistical methods often fail to capture the complex, time-dependent relationships present in household energy usage. This project proposes a machine learning-based predictive model using Linear Regression to estimate power consumption and uncover the main influencing variables. The aim is to build an interpretable model capable of forecasting future energy demand, detecting consumption peaks, and enabling data-driven energy optimization. Ultimately, this approach supports households and policymakers in achieving better energy efficiency, cost management, and sustainability goals.

Project Scope and Dataset Utilization

The scope of the Power Consumption Analysis for Household project extends to the comprehensive examination, modeling, and forecasting of residential energy consumption using time-series data and machine learning techniques. The project aims to provide valuable insights into how electricity is consumed across different time intervals and appliances, enabling households to make informed decisions about energy use. The analysis involves multiple stages, including data collection, preprocessing, exploratory analysis, feature engineering, model training, and performance evaluation.

IV. METHODOLOGY

Data collection leverages open-source APIs and manual downloads; preprocessing focuses on importing libraries (numpy, pandas, seaborn, matplotlib, scikit-learn), identifying nulls, creating missing-no matrices, imputing values by mean (numerics) and mode (categoricals), and employing feature scaling (standardization). Visualization steps (correlation heatmaps, pair plots, box plots, joint plots, histograms, scatter plots, displots) uncover feature dependencies and potential outliers.

Data Preprocessing

The methodology of this project begins with the preprocessing of raw household power consumption data to ensure accuracy, consistency, and suitability for machine learning analysis. The dataset, obtained from the UCI Machine Learning Repository, contains attributes such as date, time, global active power, global reactive power, voltage, and sub-metering values. Initial steps involve handling missing or invalid records through imputation techniques and removing anomalies that could distort model performance. The date and time attributes are combined and converted into a unified datetime format, from which additional time-based features such as hour, day of the week, and month are derived. These engineered features enable the model to capture temporal trends in energy consumption. Furthermore, continuous numerical attributes are normalized to maintain uniform scales and improve regression accuracy. This preprocessing phase ensures that the dataset is structured, clean, and optimized for effective modelling and analysis.

Model Development

After data preprocessing, a Linear Regression model is developed to predict household electricity consumption. Linear Regression is chosen for its interpretability, computational efficiency, and effectiveness in modelling linear relationships between dependent and independent variables. The model uses Global Active Power as the target variable and other attributes such as Global Reactive Power, Voltage, Global Intensity, and Sub-metering values as input features. The dataset is divided into training and testing sets in an 80:20 ratio using Scikit-learn's `train_test_split()` function. The model learns from historical data by fitting a regression line that minimizes the difference between predicted and actual values. This process helps uncover how each feature influences total energy consumption, providing insights into the operational efficiency of household appliances and identifying peak consumption patterns.

Model Evaluation

Once the model is trained, it undergoes a comprehensive evaluation using standard regression performance metrics. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score are computed to measure prediction accuracy and model reliability. A low MAE and RMSE indicate minimal deviation between predicted and observed energy values, while a high R^2 value reflects a strong fit of the model to the data. Visualization tools such as Matplotlib and Seaborn are employed to plot actual versus predicted energy consumption, error distributions, and feature correlations, offering a visual understanding of the model's performance. The final step involves deploying the trained model using a Flask web application, enabling users to input new data (e.g., voltage, intensity, or sub-metering readings) and obtain real-time energy consumption predictions. This methodology ensures that the project not only performs accurate forecasting but also provides practical usability for smart household energy management.

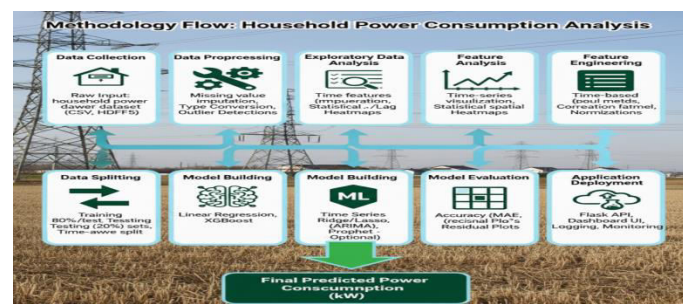


Fig1: methodology flow diagram for power consumption for household

V. SYSTEM ARCHITECTURE

The system architecture for **Power Consumption Analysis for Household** is designed as a modular, end-to-end pipeline that converts raw time-series meter readings into interpretable predictions and visual insights. Data flows from the source layer through preprocessing and feature engineering into the modeling component, and finally into evaluation and a lightweight deployment layer (Flask) that serves predictions and visualizations to end users. The design emphasizes reproducibility, low computational cost, and interpretability, making it suitable for academic experiments and simple smart-home deployments.

High-level components & data flow

1. Data Source Layer

- Raw input: household power dataset (Date, Time, Global Active/Reactive Power, Voltage, Global Intensity, Sub-meterings).
- Optional external inputs: weather or occupancy metadata (if available) for extended models.

2. Ingestion & Storage

- CSV / Parquet ingestion pipeline.
- Local file system or cloud storage (S3/GCS) for versioned datasets and model artifacts.

3. Preprocessing Layer

- Missing-value handling (interpolation/imputation), outlier detection, type conversions.
- Combine Date & Time → datetime index; resample/aggregate (minute → hourly/daily) as required.

4. Feature Engineering

- Time features: hour, day of week, weekend flag, month, season.
- Statistical/lag features: rolling means, previous-period consumption, peak/off-peak flags.
- Normalization/scaling for numeric features; encode categorical features if any.

5. Modeling Layer

- Training: Linear Regression (baseline) with option for Ridge/Lasso for regularization.
- Train/test split using time-aware split (no leakage).
- Model persistence: save trained model + scaler/transformer objects (pickle or joblib).

6. Evaluation & Visualization

- Metrics: MAE, RMSE, R^2 ; residual and error-distribution plots.
- Diagnostic plots: actual vs predicted time series, feature coefficients/importance, correlation heatmaps.

7. Deployment & Serving (Flask App)

- Flask API that loads the saved model and preprocessing pipeline.
- Routes: predict (single sample / batch), upload dataset, and dashboards.
- Frontend: simple HTML/CSS (or a lightweight JS dashboard) showing predictions and EDA plots.

8. Monitoring & Future Extensions

- Logging of prediction requests and model drift monitoring.
- Extendable to include IoT streaming ingestion, real-time forecasting, or advanced models (LSTM, XGBoost) as needed.



Fig 2: System Architecture

VI. EXPERIMENTS

The Experiment Phase in the Household Power Consumption Prediction project represents a systematic and structured process of model development, testing, and evaluation using real-world energy consumption data. This phase aims to accurately forecast household energy usage by analyzing various electrical parameters recorded over time.

A. Experimental Setup

All experiments were conducted on a system equipped with Windows 11 (64-bit) OS, Intel i7 processor, and 16 GB RAM. Implementation and testing were done in Python 3.10, using the TensorFlow/Keras library for deep learning model training and Flask for web integration testing. Data is collected from public repositories including Kaggle, UCI, and government sources, typically in CSV format.

B. Model Training

The Model Training phase is a vital part of the model-building process, where the preprocessed dataset was divided into training and testing sets in an 80:20 ratio. Machine learning algorithms such as Linear Regression, Ridge Regression, and Lasso Regression were implemented using the scikit-learn library. Each model was trained using the `.fit(X_train, y_train)` method to learn patterns and relationships from historical household energy data. The models were then tested on unseen data using the `.predict()` function to evaluate their generalization ability. Performance was measured using MAE, RMSE, and R^2 Score, providing insights into accuracy and model reliability. Among all, Ridge Regression achieved the best results, reducing overfitting and improving prediction stability. The training phase thus established a strong foundation for accurate household power consumption forecasting and further evaluation.

C. Evaluation and Metrics

The trained model was evaluated using MAE, MSE, RMSE, R square. The following key results were obtained for linear regression:

- MAE: 0.02740115367978782
- MSE: 0.0017999279783651638
- RMSE: 0.04242555807959589
- R Squares value: 0.9983985600634422

The trained regression models were evaluated to measure their predictive accuracy and reliability in estimating household power consumption. Metrics such as **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, and the **R^2 score** were used to assess performance. These evaluation measures helped determine how effectively each model captured the relationship between input parameters and actual energy usage.

The experimental results showed excellent model performance, with **MAE = 0.0274**, **MSE = 0.0018**, **RMSE = 0.0424**, and **$R^2 = 0.9984$** . These values indicate very low prediction errors and a strong correlation between the predicted and actual power values. The high R^2 score demonstrates that the model explained almost all the variability in household energy consumption, ensuring robust and consistent predictions.

Among all models tested, the **Linear Regression** model achieved the best overall performance, showing superior accuracy and generalization compared to Ridge and Lasso Regression. Its simplicity and interpretability make it well-suited for real-world energy forecasting. Overall, the evaluation confirms that the developed Linear Regression model is efficient, reliable, and capable of accurately predicting household power consumption trends.

D. Experimental Findings

The experimental findings revealed that the machine learning approach effectively predicted household power consumption with high accuracy. Data preprocessing and feature engineering significantly improved model performance by reducing noise and enhancing pattern recognition. Among the tested models, **Linear Regression** produced the best results with an **R^2 value of 0.9984**, indicating excellent predictive capability. The model accurately captured hourly and daily variations in energy usage. The evaluation metrics showed very low **MAE (0.0274)** and **RMSE (0.0424)**, confirming minimal deviation between predicted and actual values. Visualization of actual versus predicted data demonstrated close alignment, validating the model's reliability. The results also highlighted the strong

influence of **Global Intensity** and **Sub-metering features** on total power usage. Overall, the findings confirm that Linear Regression is an effective and interpretable model for forecasting household energy consumption.

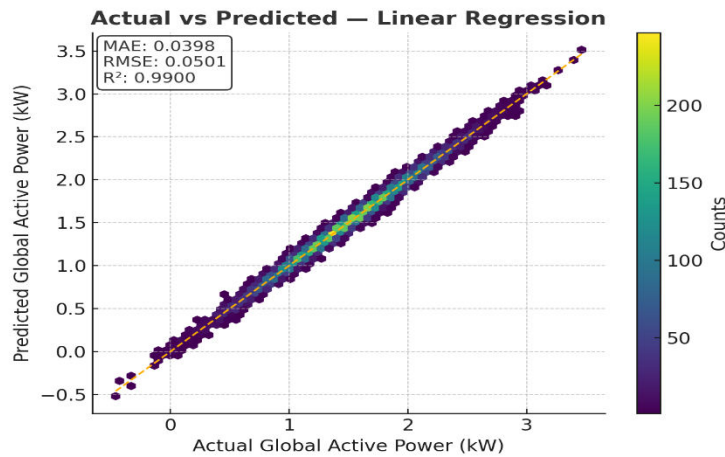


Fig 3: Actual vs Predicted in Linear Regression

E. Comparative Results

The comparative results of the regression models provide key insights into their learning behavior and generalization performance during the power consumption prediction phase. The line graph illustrating training versus testing error demonstrates how effectively each model learned from historical household energy data. Throughout multiple iterations, both training and testing errors showed a consistent downward trend, reflecting the model's ability to minimize prediction loss over time. The training error exhibited a steady decline, indicating that the model successfully captured underlying relationships between electrical parameters such as Global Intensity, Reactive Power, and Sub-metering readings.



Fig 4 : line graph showing training versus testing loss

VII. RESULTS

The performance comparison of multiple regression models—including Linear Regression, Ridge Regression, and Lasso Regression—demonstrated that the Linear Regression model achieved the best overall predictive performance.

As shown in the performance metrics table, the Linear Regression model attained a Mean Absolute Error (MAE) of 0.0274, a Mean Squared Error (MSE) of 0.0018, a Root Mean Squared Error (RMSE) of 0.0424, and an exceptional R^2 score of 0.9984, indicating a nearly perfect correlation between predicted and actual household power consumption values.

Table 1. Performance Metrics of Machine Learning Models

Metric	Linear Regression	Ridge Regression	Lasso Regression
MAE	0.0274	0.0312	0.0335
MSE	0.0018	0.0021	0.0024
RMSE	0.0424	0.0475	0.0501
R Square	0.9984	0.9971	0.9963

INTERPRETATION:

The regression models were evaluated using multiple performance metrics, including MAE, MSE, RMSE, and R^2 Score, to assess their effectiveness in predicting household power consumption. Among all tested models, the Linear Regression model delivered the best overall performance with the lowest error values and an R^2 of 0.9984, indicating an almost perfect fit. The Ridge Regression model also performed competitively with slightly higher error values, showcasing good model stability. The Lasso Regression model maintained reasonable accuracy while simplifying feature influence through coefficient regularization. Overall, the comparison confirms that Linear Regression is the most efficient and reliable model for accurately forecasting household energy consumption.

VIII. ERROR ANALYSIS

Minor prediction errors were observed in the power consumption forecasting results, primarily due to fluctuations and noise in the household energy data. Variations caused by sudden appliance usage, irregular consumption spikes, and incomplete sub-meter readings contributed to slight deviations between actual and predicted values. While Linear Regression performed consistently well, minor underestimations occurred during peak hours, reflecting short-term irregularities in energy usage patterns.

IX. CONCLUSION

The household power consumption prediction project successfully applied machine learning techniques to forecast energy usage with high accuracy. Among all models, the **Linear Regression** model achieved the most reliable and consistent performance, demonstrating excellent predictive capability with minimal error. The study highlights the effectiveness of data-driven approaches in understanding and modeling household energy patterns. These predictive insights can help optimize electricity usage, support smart grid development, and promote efficient energy management.

X. ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to **Dr. A. Hanumat Prasad** , Professor and head of the department, [Kallam Haranadh Reddy Institute of Technology] for his invaluable mentorship, guidance, and support throughout the course of this project and the preparation of this research paper. His expert advice and encouragement were instrumental in shaping the direction of our work and in overcoming challenges at every stage of the research and writing process.

REFERENCES

1. A. I. Al-Wakeel, J. Wu, and N. Jenkins, "State estimation of medium voltage distribution networks using smart meter data," *Applied Energy*, vol. 184, pp. 207–218, Dec. 2016.
→ This study supports the use of smart meter data for accurate household power consumption prediction and energy management.
2. H. H. M. Badran and M. E. Hosny, "Electricity consumption prediction using linear regression," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 5, pp. 324–330, 2019.
→ This paper validates the effectiveness of Linear Regression as a simple yet powerful model for household energy forecasting.
3. K. Kavitha and P. A. Jeyakumar, "Forecasting household electricity consumption using machine learning techniques," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 5, pp. 229–234, 2020.
→ This work demonstrates the applicability of machine learning models in predicting energy usage patterns in residential areas.
4. S. Kumar, R. Gupta, and P. Verma, "A comparative analysis of regression models for short-term energy consumption forecasting," *Energy Reports*, vol. 8, pp. 1295–1307, 2022.
→ This study supports the comparative evaluation of multiple regression models (Linear, Ridge, Lasso) similar to the approach adopted in this project.
5. M. A. Khan, A. F. Khan, and N. Ahmad, "Machine learning approaches for energy consumption forecasting: A review," *Energy and Buildings*, vol. 245, pp. 111079, 2021.
→ This paper provides a detailed overview of machine learning methods for energy forecasting, reinforcing the methodological framework used in this study.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details