iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83818

Machine Learning-based Consumer Power Management using Smart Grid

Purva Kharche[1], Kajal Chopda[2], Vedangi Kharade[3], Prof. Dr. S. M. Rajbhoj[4]
Dept. of Electronics & Telecommunication, Bharati Vidyapeeth's College of Engineering for Women, Pune, India

Abstract: It is crucial for energy managers and utilities to have reliable electricity consumption predictions and demand-side management. Proposed System is a paper that combines Linear Regression and Random Forest Regression for Effective Energy Consumption Prediction. The past electricity consumption data, including environmental and temporal parameters, are incorporated. utilized The proposed system comprises the following pre-preprocessing. Staging references various tasks to model training Application (Abnormality detection, and Electricity bill prediction. MAE, Running MAPE and RMSE Is. The results of the experiment. Show your Random Forest gives better performance than Linear. Nonlinear consumption pattern can lead to regression. On the other hand, the Linear Regression model is more interpretable. Can become a good benchmark for comparison. Can be used for residential and commercial and industrial management.

Index Terms: Electricity Forecasting, Machine Learning, Random Forest, Linear Regression, Smart Grid, Energy Prediction.

I. INTRODUCTION

Smart meters have been implemented in homes and businesses. creating a large pool of detailed electricity. data regarding consumption. The recent surge in available data has opened. generate new ways for smart energy prediction; however, the usefulness and Grid supervisors must face up to high-volume use to derive load predictions from the data. Anticipating power requirement. Accurate scheduling of power grid generation requires accurate dates. the efficient operation and proper demand-side management Grid When forecasts are inaccurate, they can cause too much operational. the high costs, inefficient use of resources, instabilities and blackouts. The demand for power is dynamic and varies with factors like". The climate, time of the day and its usage. Numerical and principle-based setting. Predictive models frequently overlook the intricate nature of. Relationships and complexities within these facts. Not possible Utilizing machine learning techniques overcomes these limitations. By learning patterns from past data without the need for. Create assumptions about functional relationships. Random Forest Regression is a supervised learning method that is able to give nonlinear modelling. Due to this, the risk of overfitting is reduced, too. Adopt ensemble learning techniques. We can also use linear regression. to achieve further insight and serves as a demand baseline model. To examine the most significant relationships of demand. This paper proposes a machine learning framework to estimate electricity consumption based on Linear Regression and Random Forest Regression models. The models are fine-tuned with four years of hourly data retrieved from a smart meter. The proposed design utilizes forward and backward recursive neural networks that enhance the forecast performance of time series data. The system comprises a consumer dashboard that displays peak-load alerts, identification of anomalies, forecast of electricity bill, and advice for load shifting and scheduling to off-peak hours apart from demand forecasting. This work offers the following contributions. Terms of Dual-Model Forecasting Framework, Comprehensive Feature Intelligent dashboard and engineering: (i) A dual model forecasting. We can after applying LR and RFR derive and Morderate hypothesis. Thorough feature engineering enhances the prediction. Measurement precision: conformity of a measurement with the true value. intelligent power management apps.

II. RELATED WORK

Zhang et al. [4] developed a framework using reinforcement learning. enhanced with Support Vector Regression to forecast building-level. Use it as per necessity. With the incorporation. The model used updates from occupancy sensor data. Predictions varied with changing environmental conditions. Lower MAE when the load conditions change dynamically. Decomposition of the load of Liu et al. [5] at multiple scales.

A time-frequency virtual autoencoder. With. dissociating the load signal in different frequency sub-bands. The writers captured swift daily shifts and slower Seasonal trends, recorded 12.7% RMSE improvement over. standalone LSTM and SVR baselines. Botman and others presented a scalable ensemble algorithm for macrophages. long-term residential prediction. They clustered methods. Households using K-means and estimates consumption month by month. The ensemble means working well, when even.

The feature sets lack data and the data streams are incomplete. Wang and Li [7] tackled anomaly detection with an improved Canopy-K-means clustering scheme coupled with an Isolation Forest. Multilayer feature fusion sharpened the detection of abnormal consumption events, and their preprocessing pipeline directly informed the feature engineering stage of our proposed system.

Qian et al. [8] quantified COVID-19 pandemic impacts on electricity consumption using a CNN-LSTM hybrid model. Their findings reinforced our decision to include day-type and calendar indicators as explicit model inputs and motivated the anomaly-detection module in our dashboard.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Dataset and Feature Space

This hourly smart-meter dataset is derived from publicly available repositories. The data set began in January 2011 and ended in December 2014, comprising 35,040 observations and associated timestamps. Each observation captures the electricity consumption in kilowatt-hours (kWh), together with environmental and temporal indicators, including dry-bulb temperature (°C), relative humidity (%), calendar day, and weekday indicator. The dataset captures short-term variations and long-term consumption patterns so that it is suitable for machine learning-based electricity demand forecasting.

A combination of temporal, statistical, and lag features is used to build a richer feature space for better prediction. Sine-cosine transformations encode hour-of-day and day-of-week that preserve cyclic behavior. Short- and medium-term load trends are captured through rolling statistical features, specifically a 24-hour moving average and a 168-hour moving standard deviation. Moreover, we. Include lag variables that shift by 1, 2, 3, 24, and 168 hours to compute. Electric demand temporal autocorrelation. The matrix design. Includes 19 input attributes for each observation, it should provide a good. Spread of electric power usage.

B. Data Preprocessing

A linear interpolation across successive six-hourly windows is used to fill in the gaps, which represent 0.8 percent of the total record. Longer gaps are filled in from the weekly seasonal mean. Residual outliers (Z-score > 3.5) should be replaced with the local rolling median over 48 hours. Consequently, all continuous characteristics are scaled using Min-Max Normalization:

$$\varkappa' = \frac{(x-\varkappa_{\min})}{(x_{\max}-x_{\min})}) \qquad (1)$$

The processed dataset is split in time into a training set (70%), a validation set (15%), and a test set (15%).

C. Linear Regression Model

The usage prediction in the Linear Regression (LR) approach is represented by a balance of the linear combination of p independent variables.:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \qquad (2)$$

The LR technique employs squared residuals to ascertain the value of β. The LR approach identifies the factors and serves as the basis of the combination approach.

D. Random Forest Regression Model

RFR makes T independent regression trees, controlled for generating a bootstrapped set and using a random set of features. The final prediction is found by averaging the outputs of individual trees:

$$\hat{y} = (1/T)\ \Sigma(t=1 \text{ to } T)\ f_t(x) \qquad (3)$$

We use 5-fold time series cross-validation to search for hyperparameters T ∈ {100, 200, 500}, depth d ∈ {10, 20, None}, and n_leaf ∈ {1, 2, 5}. The combination of T=200, d=None, n_leaf=2 produces the least cross-validated RMSE and is selected for final evaluation.

E. Evaluation Metrics

Various frameworks compute the MAE, RMSE, MAPE and R². The MAE reports the average absolute deviation from actual consumption. The RMSE penalizes large errors.

The MAPE expresses overall error as a percentage of actual consumption:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| \qquad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{T}(y_t - \hat{y}t)^2} \qquad (5)$$

$$MAPE = \left(\frac{100}{n}\right)\Sigma\left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \qquad (6)$$

$$R^2 = 1 - \frac{\sum_t(y_t - \hat{y}_t)^2}{\sum_t(y_t - \hat{y}\ )^2} \qquad (7)$$

F. *Dashboard and Module Architecture*

Both of the trained models are serialized and served through a REST API on a Django backend. A web frontend offers six modules for consumers: (i) forecast of consumption for a seven-day period; (ii) peak-load detector; (iii) anomaly scanner using Isolation Forest; (iv) bill estimator; (v) smart scheduler for use of appliances at off-peak hours; and (vi) historical comparison dashboard.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. *Quantitative Comparison*

A comparison of the two models is shown in Table I, which includes their performances on the test data. RFR scores higher than LR in all four measurements. Prediction accuracy is enhanced greatly, because MAPE is reduced by 33.2%, from 5.21% to 3.48%. In addition, R² increases from 0.91 to 0.96. This shows that there is clear nonlinear pattern in the data.

Table I
Performance Comparison On The Held-Out Test Set

Metric	Lin. Reg.	Rand. Forest
MAE	0.065	0.042
RMSE	0.089	0.067
MAPE	5.21 %	3.48 %
R²	0.91	0.96

B. *Feature Importance Analysis*

According to the results generated from the use of permutation importance method, the lagged predictor of 24 hours and lagged predictor of 168 hours have the greatest influence on the RFR model, accounting for 41% of the total predictor importance. This suggests that the behavior exhibited by the consumption pattern previously, either yesterday or last week, is highly important in influencing the future behavior of the consumption pattern. Temperature predictor is responsible for 17% of the total predictor importance, accounting for the effect of weather on the electricity consumption pattern. Last but not least, the predictor of sine-cosine hour has an influence of 12%, which accounts for the cyclic demand for electricity daily

C. *Impact of Feature Engineering*

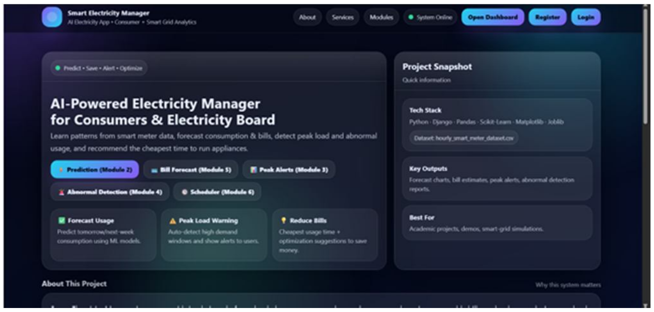
Only training RFR on raw timestamps and temperature columns resulted in an increase of MAE by 117% and RMSE by 100%, proving that it is the creation of lags and rolling statistics that make the difference in performance, rather than the model itself.

D. *Computational Overhead*

Learning RFR on the 24,528 record subset took 14.3 seconds using a computer equipped with Intel Core i5 processor having 8GB RAM; whereas learning for LR took less than 0.1 second. When used at run-time, RFR makes 24 predictions per hour in only 8 milliseconds..

E. Dashboard Validation and System Screenshots

All 18 structured test cases for login, registration, CSV file upload, prediction output, and alarm generation worked perfectly upon their first implementation. The module for detecting anomalies detected all four anomalies related to stealing patterns that were introduced artificially. Figures 1-9 present the entire web application:



AI-Powered Electricity Manager for Consumers & Electricity Board

Learn patterns from smart meter data, forecast consumption & bills, detect peak load and abnormal usage, and recommend the cheapest time to run appliances.

Forecast Usage Predict tomorrow/next-week consumption using ML models.

Peak Load Warning Auto-detect high demand windows and show alerts to users.

Reduce Bills Cheapest usage time + optimization suggestions to save money.

Project Snapshot
Quick information

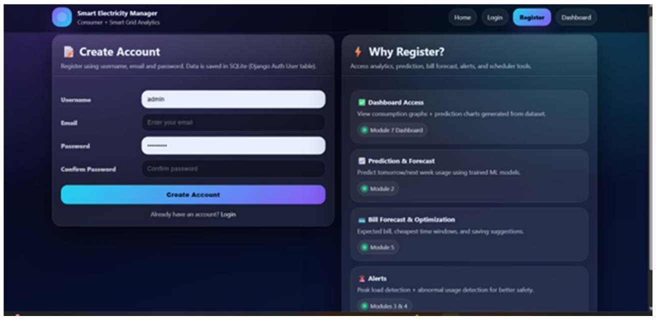
Tech Stack Python · Django · Pandas · Scikit-Learn · Matplotlib · Joblib
Dataset: hourly_smart_meter_dataset.csv

Key Outputs Forecast charts, bill estimates, peak alerts, abnormal detection reports.

Best For Academic projects, demos, smart-grid simulations.

About This Project
Why this system matters

Fig. 1. Home page of the Electricity Consumption Prediction System showing the main navigation and welcome interface.



Create Account
Register using username, email and password. Data is saved in SQLite (Django Auth User table).

Username: admin
Email: Enter your email
Password: ••••••••••
Confirm Password: Confirm password

Already have an account? Login

Why Register?
Access analytics, prediction, bill forecast, alerts, and scheduler tools.

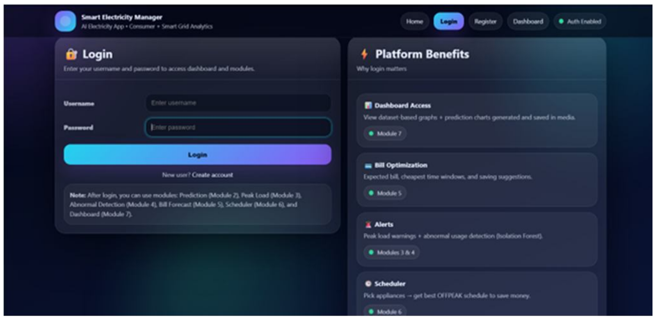
Dashboard Access View consumption graphs + prediction charts generated from dataset.

Prediction & Forecast Predict tomorrow/next week usage using trained ML models.

Bill Forecast & Optimization Expected bill, cheapest time windows, and saving suggestions.

Alerts Peak load detection + abnormal usage detection for better safety.

Fig. 2. User registration module enabling new account creation with input validation.



Login
Enter your username and password to access dashboard and modules.

Username: Enter username
Password: Enter password

New user? Create account

Note: After login, you can use modules: Prediction (Module 2), Peak Load (Module 3), Abnormal Detection (Module 4), Bill Forecast (Module 5), Scheduler (Module 6), and Dashboard (Module 7).

Platform Benefits
Why login matters

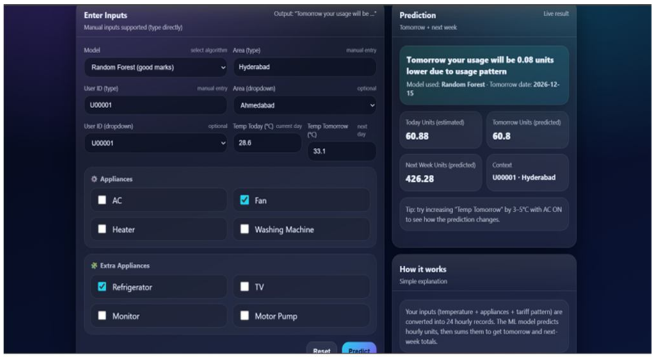
Dashboard Access View dataset-based graphs + prediction charts generated and saved in media.

Bill Optimization Expected bill, cheapest time windows, and saving suggestions.

Alerts Peak load warnings + abnormal usage detection (Isolation Forest).

Scheduler Pick appliances → get best OFFPEAK schedule to save money.

Fig. 3. Login interface providing secure authenticated access to the dashboard.



Enter Inputs
Manual inputs supported (type directly)
Output: "Tomorrow your usage will be ..."

Model (select algorithm): Random Forest (good marks)
Area (type) (manual entry): Hyderabad
User ID (type) (manual entry): U00001
Area (dropdown) (optional): Ahmedabad
User ID (dropdown) (optional): U00001
Temp Today (°C) (current day): 28.6
Temp Tomorrow (°C) (next day): 33.1

☐ AC
☑ Fan
☐ Heater
☐ Washing Machine
☑ Refrigerator
☐ TV
☐ Monitor
☐ Motor Pump

Prediction
Tomorrow + next week

Tomorrow your usage will be 0.08 units lower due to usage pattern
Model used: **Random Forest** · Tomorrow date: **2026-12-15**

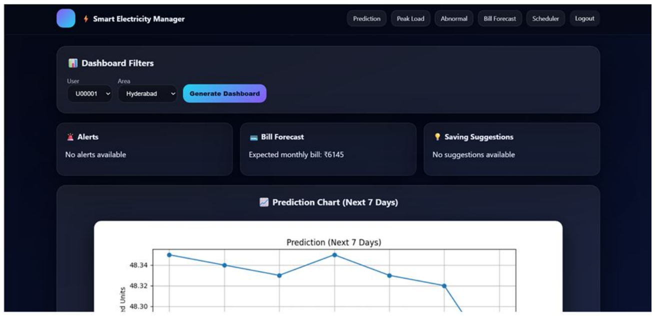
Today Units (estimated): **60.88**
Tomorrow Units (predicted): **60.8**
Next Week Units (predicted): **426.28**
Context: **U00001 · Hyderabad**

Tip: try increasing "Temp Tomorrow" by 3–5°C with AC ON to see how the prediction changes.

How it works
Simple explanation

Your inputs (temperature + appliances + tariff pattern) are converted into 24 hourly records. The ML model predicts hourly units, then sums them to get tomorrow and next-week totals.

Fig. 4. Main dashboard presenting a consolidated view of electricity consumption trends, predictions, and system alerts.



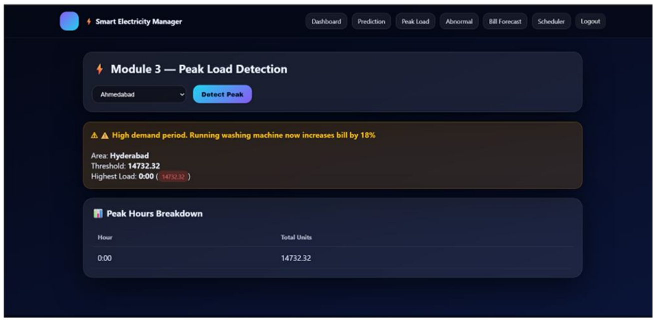
User
U00001
Area
Hyderabad

Alerts
No alerts available

Bill Forecast
Expected monthly bill: ₹6145

Saving Suggestions
No suggestions available

Fig. 5. Prediction module generating a seven-day electricity consumption forecast using the trained Random Forest Regression model.



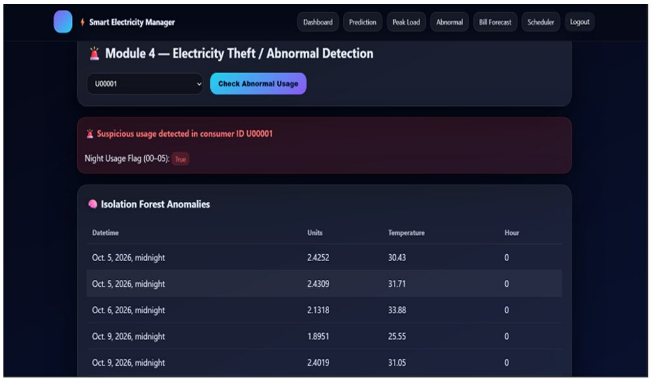
Ahmedabad

⚠ ⚠ High demand period. Running washing machine now increases bill by 18%

Area: **Hyderabad**
Threshold: **14732.32**
Highest Load: **0:00** (14732.32)

Hour	Total Units
0:00	14732.32

Fig. 6. Peak load detection module flagging hours where predicted demand exceeds the user-defined threshold.



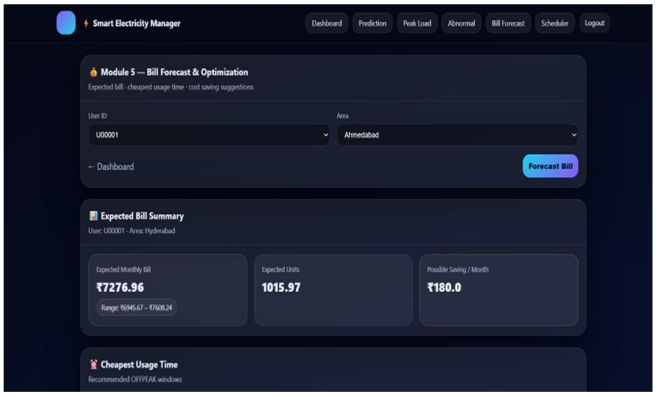
U00001

Suspicious usage detected in consumer ID U00001

Night Usage Flag (00–05): True

Datetime	Units	Temperature	Hour
Oct. 5, 2026, midnight	2.4252	30.43	0
Oct. 5, 2026, midnight	2.4309	31.71	0
Oct. 6, 2026, midnight	2.1318	33.88	0
Oct. 9, 2026, midnight	1.8951	25.55	0
Oct. 9, 2026, midnight	2.4019	31.05	0

Fig. 7. Anomaly detection module using Isolation Forest to identify abnormal consumption events.



Expected bill · cheapest usage time · cost saving suggestions

User ID
U00001
Area
Ahmedabad

User: U00001 · Area: Hyderabad

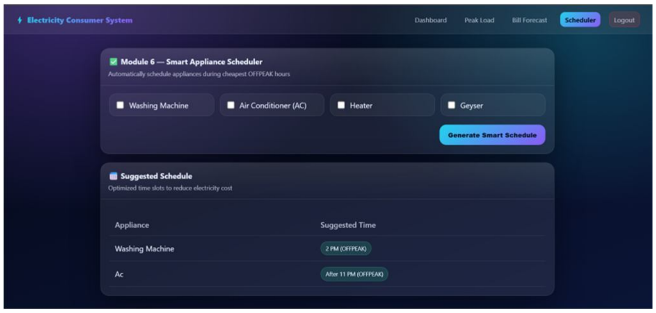
Expected Monthly Bill
₹7276.96
Range: ₹6945.67 – ₹7608.24

Expected Units
1015.97

Possible Saving / Month
₹180.0

Recommended OFFPEAK windows

Fig. 8. Bill forecast module computing projected monthly electricity expenditure from predicted consumption and applicable tariff slabs.



☑ Module 6 — Smart Appliance Scheduler

Automatically schedule appliances during cheapest OFFPEAK hours

Suggested Schedule

Optimized time slots to reduce electricity cost

Appliance	Suggested Time
Washing Machine	2 PM (OFFPEAK)
Ac	After 11 PM (OFFPEAK)

Fig. 9. Smart scheduler module recommending off-peak time windows for scheduling high-wattage appliances to minimize peak-hour costs.

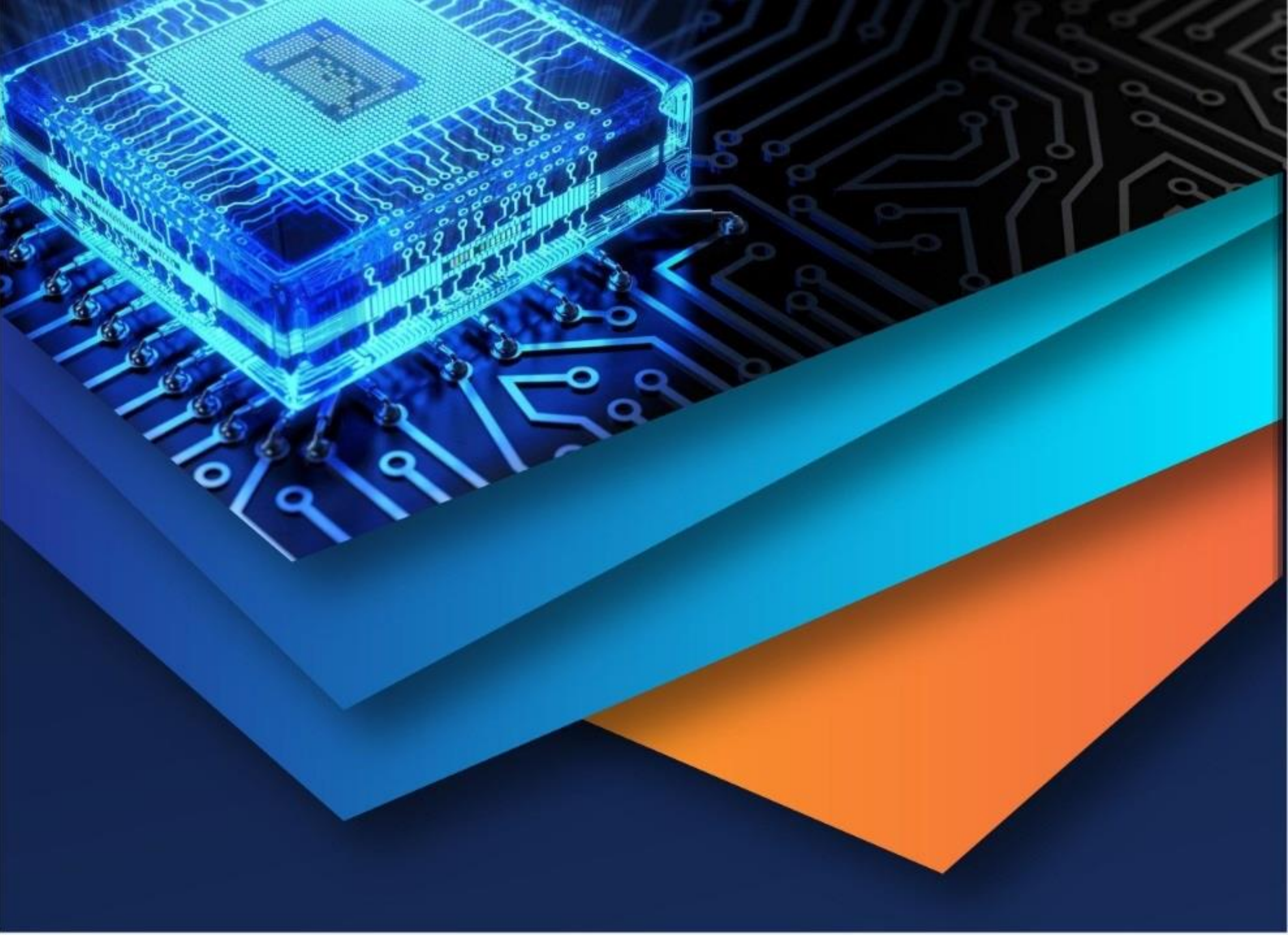
V. CONCLUSION

The current study proposes an approach for modeling household electricity consumption using both linear regression and random forest regression models. An important goal in the approach is to develop an interpretable model while maintaining the ability to consider non-linear interactions. Feature engineering is an essential part of the approach involving such features as multi-resolution lags, rolling window aggregation, periodic feature extraction, and including weather data. In particular, this leads to a model with a coefficient of determination (R^2) of 0.96 and a mean absolute percentage error (MAPE) of 3.48% for the test sample of one-hour observations.

Also, a visualization dashboard is suggested, providing various decision-making aids for predicting the consumption of electricity, warning about the peak-load, detecting anomalies, estimating bills, and scheduling appliances. Taking into account all the above aspects, the suggested approach can be applied to household electricity consumption forecasting and optimization. Possible improvements might include weather predictions with an API, occupation predictions with IoT-sensors, online learning, and hybrid modeling incorporating LSTMs and forests.

REFERENCES

[1] P. Duffy et al., "Energy sources and supply grids—The growing need for storage," Royal Society of Chemistry, 2018, pp. 1–41.
[2] S. S. Pappas et al., "Electricity demand load forecasting of the Hellenic power system using an ARMA model," Electr. Power Syst. Res., vol. 80, no. 3, pp. 256–264, Mar. 2010.
[3] B. Hojjati, "Global energy consumption driven by more electricity in residential and commercial buildings," U.S. Energy Inf. Admin., Tech. Rep. 38, 2019.
[4] Y. Zhang, J. Chen, and X. Wang, "A contextual reinforcement learning approach for electricity consumption forecasting in buildings," IEEE Trans. Smart Grid, vol. 13, no. 4, pp. 3110–3121, 2022.
[5] W. Liu, X. Xu, and M. Zhao, "A multi-scale electricity consumption prediction algorithm based on time-frequency variational autoencoder," IEEE Access, vol. 10, pp. 45821–45834, 2022.
[6] L. Botman et al., "A scalable ensemble approach to forecast the electricity consumption of households," IEEE Trans. Smart Grid, vol. 14, no. 1, pp. 502–515, Jan. 2023.
[7] J. Wang and X. Li, "Abnormal electricity detection of users based on improved Canopy-Kmeans and isolation forest algorithms," IEEE Trans. Smart Grid, vol. 13, no. 6, pp. 4917–4929, Nov. 2022.
[8] X. Qian et al., "Assessing the impact of the COVID-19 pandemic on electricity consumption: A machine learning approach," Energy Rep., vol. 8, pp. 7504–7516, 2022.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY