Advances in Consumer Research
https://acr-journal.com/
Volume-2 | Issue-4 | Sep 2025

Original Researcher Article

Consumer-Centric Energy Use in Educational Institutions: A Machine Learning Approach to Predictive Consumption and Behavioral Insights

Ms. Preetha G.[1], Dr. Vignesh Karthik S.A.[2], Dr. Ameena Babu V.[3] and Dr. Sagini Thomas Mathai[4]

[1]Assistant Professor I, Kumaraguru College of Technology, Coimbatore
Email: preetha.g@kctbs.ac.in
[2]Assistant Professor III, Kumaraguru College of Technology, Coimbatore
Email: vignesh.karthik25@gmail.com
[3]Associate Professor, Amity Global Business School, Kochi
Email: vababu@kch.amity.edu
[4]Professor, MBA, Christ College of Engineering, Irinjalakkuda
Email: saginithomasmathai@gmail.com

Received:01/08/2025
Revised: 15/08/2025
Accepted:04/09/2025
Published:20/09/2025

ABSTRACT
While this study offers significant insights, it is constrained by several limitations. Primarily, the accuracy of the analysis is contingent upon the reliability of data captured by IoT sensors. Any anomalies such as sensor malfunctions, calibration errors, or transmission lags could compromise data quality and, consequently, the robustness of the predictive models. Additionally, the study is based on data from a single institution—Kumaraguru Institutions—whose infrastructural and operational characteristics may not fully represent the diversity found in other educational environments. This contextual specificity limits the generalizability of the findings across broader institutional or geographic contexts. Despite these limitations, the study lays the groundwork for future research and development of adaptive, cross-contextual energy management systems.

Keywords: IoT Sensors, Data Reliability, Predictive Models, Institutional Context, Energy Management Systems.

INTRODUCTION

About the Study

Energy management has emerged as a critical dimension of institutional sustainability, particularly in higher educational settings characterized by expansive infrastructure and continuous energy demands. Efficient energy utilization not only lowers operational expenditures but also reinforces institutional commitment to environmental stewardship. This study examines the energy management practices at Kumaraguru Institutions, a pioneer in sustainability initiatives, leveraging data-driven approaches for optimization. Utilizing data collected from Internet of Things (IoT) sensors installed across strategic locations on the campus, the study analyzes energy consumption patterns in real time. IoT-enabled systems provide granular and continuous data, offering fertile ground for the application of machine learning algorithms. The central objective of this research is to develop a predictive energy management framework that enhances load forecasting, identifies inefficiencies, and facilitates proactive energy distribution strategies. By integrating machine learning with IoT infrastructure, the study proposes a scalable model for real-time, data-informed energy decision-making. Such a model holds the potential to reduce energy wastage, align peak and off-peak demand with usage trends, and ultimately improve operational performance. In doing so, the study contributes to both practical and theoretical advancements in sustainable energy practices within educational ecosystems, setting a benchmark for similar institutions aspiring to integrate intelligent energy management systems.

Problem Statement

Educational institutions, owing to their vast and multifaceted infrastructure, face unique challenges in achieving efficient energy management. Conventional systems often lack the capacity to provide real-time monitoring or predictive insights, resulting in energy inefficiencies, elevated operational costs, and missed opportunities for optimization. Despite the deployment of IoT-based infrastructure at Kumaraguru Institutions—recognized for its sustainability initiatives—there exists a significant gap in translating this data into actionable insights. The absence of a robust, predictive energy management system limits the institution's ability to maximize the utility of its sensor-generated data. This research seeks to bridge this gap by harnessing machine learning techniques to analyze high-volume, real-time data streams for energy consumption optimization. Without such data-driven interventions, institutions risk persistent inefficiencies and underutilization of technological infrastructure, which undermines both cost-effectiveness and sustainability objectives. Addressing this problem is essential for formulating a systematic, predictive, and replicable approach to intelligent energy management in academic institutions.

Scope of the Study

This research endeavors to design a predictive energy management framework grounded in machine learning techniques, using Kumaraguru Institutions as the primary case study. The scope includes the collection and analysis of energy consumption data through IoT sensors strategically deployed across campus buildings. By identifying consumption patterns, inefficiencies, and load variances, the study aims to develop predictive models that enable dynamic energy optimization. The study not only aspires to enhance the institution's energy efficiency and cost management but also to contribute to broader environmental goals by minimizing its carbon footprint. The integration of real-time IoT data with machine learning models exemplifies a novel, scalable approach to sustainable energy management. Furthermore, the proposed framework is envisioned to serve as a reference model for similar academic and non-academic institutions seeking to implement data-driven sustainability solutions.

Limitations of the Study

While this study offers significant insights, it is constrained by several limitations. Primarily, the accuracy of the analysis is contingent upon the reliability of data captured by IoT sensors. Any anomalies such as sensor malfunctions, calibration errors, or transmission lags could compromise data quality and, consequently, the robustness of the predictive models. Additionally, the study is based on data from a single institution—Kumaraguru Institutions—whose infrastructural and operational characteristics may not fully represent the diversity found in other educational environments. This contextual specificity limits the generalizability of the findings across broader institutional or geographic contexts. Despite these limitations, the study lays the groundwork for future research and development of adaptive, cross-contextual energy management systems.

Major Players in Indian Higher Education

India's higher education sector comprises a mix of elite institutions, established private universities, and dynamic regional colleges. Prominent among them are the Indian Institutes of Technology (IITs), Indian Institutes of Management (IIMs), and top-tier private universities such as Amity University, VIT, and BITS Pilani. These institutions set benchmarks in academic excellence, research output, and industry engagement. Regional players such as Kumaraguru College of Technology (KCT) also play a critical role in delivering quality education tailored to the needs of emerging industries. By offering specialized programs in engineering, management, and technology, institutions like KCT contribute significantly to the national talent pool, particularly in sectors such as IT, manufacturing, and renewable energy. Collectively, these players form a dynamic ecosystem addressing the growing demand for skilled professionals in India and abroad.

Emerging Trends in the Higher Education Sector

The Indian higher education landscape is undergoing a transformative shift driven by technological innovation, policy reform, and evolving learner expectations. Key trends include the proliferation of digital and blended learning platforms, enabling flexible and inclusive education. There is a discernible shift toward competency-based curricula that align with industry requirements and emphasize employability. Institutions are increasingly investing in research and innovation ecosystems, fostering entrepreneurship through incubation centers and interdisciplinary collaboration. Moreover, regulatory bodies such as AICTE and NBA are steering institutions toward outcome-based education and continuous quality improvement. The adoption of advanced technologies—such as artificial intelligence, data analytics, and blockchain—is redefining administrative efficiency and pedagogical delivery. Sustainability, too, is gaining traction, with many institutions incorporating green practices and ESG (Environmental, Social, Governance) metrics into their operational strategies.

Challenges Confronting the Sector

Despite its growth, the Indian higher education sector continues to grapple with multifaceted challenges. A major concern is the persistent gap between academic instruction and industry expectations, resulting in skill mismatches among graduates. Many institutions face infrastructural inadequacies, limited access to research funding, and challenges in attracting and retaining qualified faculty. Furthermore, the rapid pace of technological change demands continuous curriculum innovation and faculty upskilling, which many institutions struggle to implement effectively. Issues of equity and accessibility remain pressing, particularly in rural and underrepresented regions. Additionally, institutions are under growing pressure to meet national and international accreditation standards, which often require significant organizational restructuring and resource mobilization. Overcoming these challenges is essential for ensuring the sector's long-term resilience and global competitiveness.

LITERATURE REVIEW

1. (Bourhnane et al., 2020) The study explores the integration of machine learning (ML) and predictive models for energy consumption forecasting in Smart Buildings (SB) and Smart Grids (SG). It highlights the use of Artificial Neural Networks (ANN) and Genetic Algorithms (GA) to predict and schedule energy consumption. The research emphasizes how these models, when applied in real-world

testbeds like SBs, can enhance energy management efficiency. The study also discusses challenges such as the limited size of training data affecting prediction accuracy. The authors suggest that further research should focus on expanding data sets and refining ML algorithms to achieve higher prediction reliability, advocating for a robust foundation for future energy management system developments.

2. Qiao, Yunusa-Kaltungo, and Edwards (2020) present a systematic review focusing on building energy consumption prediction, analyzing both well-established and emerging areas of research. The study highlights the increasing complexity of energy systems in modern buildings, which is compounded by variations in occupant behavior, geographical location, and the type of building. The authors categorize prediction methods into four main groups: physical, statistical, artificial intelligence-based, and hybrid methods. The systematic review identifies trends in prediction techniques and suggests that data-driven models, especially those integrating artificial intelligence (AI), can significantly improve accuracy in energy consumption forecasts. The study also emphasizes the need for greater integration of weather conditions, building types, and time scales in energy prediction frameworks.
3. Marinakis (2020) focuses on the role of big data in enhancing energy efficiency in buildings, presenting a high-level architecture for data exchange, management, and real-time processing. The study highlights the increasing data generation within European buildings from sources like smart meters, sensors, and Internet of Things (IoT) devices. The paper discusses the challenges of managing this vast amount of heterogeneous, dispersed data and the importance of creating a data-driven framework that integrates cross-domain data. The architecture proposed in this paper combines big data with emerging technologies like artificial intelligence (AI) and distributed ledger technology (DLT), aiming to support better policymaking and create innovative energy efficiency services.
4. Yang, S., Wan, M. P., Chen, W., Ng, B. F., & Dubey, S. (2020) examines the role of feedback loops in the context of building retrofits, with a particular focus on integrating smart home technologies into energy systems and the built environment in Great Britain. The study highlights the importance of understanding consumer experiences and behaviors in relation to smart retrofit technologies. Using a statistical model with data from 161 customers, the research explores the relationship between technology, control systems, and consumer care. The findings indicate that participants expressed positive feedback about smart-enabled heaters. However, those with lower incomes faced challenges related to device control and affordability.
5. Banik, R., Das, P., Ray, S., & Biswas, A. (2021) presents a machine learning-based model for predicting short-term electrical energy consumption in Agartala, Tripura, India. The authors use Random Forest and XGBoost algorithms, along with an ensemble approach, to forecast electricity demand for the next 24 hours, with additional predictions for one week to one month. The ensemble method improves prediction accuracy by 15-29%, offering valuable insights for optimizing energy management in smart grids. The study highlights the effectiveness of machine learning in improving load forecasting for better power system reliability and economic operations.
6. Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021) focuses on improving energy consumption prediction in Building Energy Management Systems (BEMS) using machine learning. The study develops predictive models using Support Vector Machine (SVM), Artificial Neural Network (ANN), and k-Nearest Neighbour (k-NN) algorithms on the Microsoft Azure cloud-based platform. Real-life data from two commercial building tenants in Malaysia are used for training and testing. The performance of the models is evaluated using RMSE, NRMSE, and MAPE metrics, highlighting different energy consumption patterns for each tenant and addressing the issue of low prediction accuracy in BEMS.
7. Bian, S., Li, C., Fu, Y., Ren, Y., Wu, T., Li, G.-P., & Li, B. (2021) focuses on developing a real-time monitoring system for Smart Connected Workers (SCWs) to improve energy efficiency in small and medium-sized manufacturers (SMMs). By integrating machine learning techniques such as object detection, text recognition, and energy disaggregation, the system optimizes manufacturing workflows and reduces human labor costs. The system leverages cloud computing, IoT, and artificial intelligence (AI) to gather, process, and analyze data for efficient decision-making.

OBJECTIVES OF THE STUDY

1. To develop a hybrid machine learning model that combines Random Forest (RF) and Gradient Boosting (GB) algorithms for improved accuracy in predicting energy consumption within educational campus buildings.
2. To analyse historical energy usage patterns and identify the most significant features

influencing consumption, enabling more informed and data-driven forecasting.

3. To build a user-friendly predictive system capable of providing real-time energy consumption forecasts, supporting decision-makers in efficient energy planning and management.

RESEARCH METHODOLOGY

This research methodology outlines the approach to achieving the objectives of the study on energy consumption prediction in educational institutions using hybrid machine learning models. The study uses a combination of Random Forest and Gradient Boosting algorithms to improve the accuracy of power usage forecasting. Key features such as rolling mean, energy per block, and peak load indicators are analysed for their impact on total energy consumption. The methodology also includes model evaluation, performance comparison, and deployment through a user-friendly interface for real-time predictions.

METHOD OF DATA COLLECTION

Secondary Dataset: The study utilizes a secondary dataset containing historical energy consumption records (2023–2024) collected from energy monitoring systems within an educational institution. The data was obtained in Excel format and includes key variables such as total power consumption, energy per block, day type classification, and peak load status.

Data Source: The dataset was sourced from the institution's internal energy management reports, compiled by the campus facilities team. These records are routinely maintained for operational monitoring and sustainability tracking.

THEORETICAL FRAMEWORK

Energy consumption forecasting plays a crucial role in optimizing energy usage, reducing costs, and enhancing sustainability efforts. By analysing historical energy consumption data, seasonal trends, and external factors, educational institutions can better plan energy use and implement energy-saving strategies. For this study, historical energy consumption data from 2023 to 2024 is analysed to identify consumption patterns, such as peak loads during specific times of the day or during certain events. External factors like weather conditions, day type (weekdays vs weekends), and the campus's operational schedule (e.g., holidays, exam periods) are also considered as they significantly affect energy usage. Predictive models, such as Random Forest (RF) and Gradient Boosting (GB), are used to capture trends, identify anomalies, and forecast future consumption. The hybrid model combining RF and GB provides more accurate predictions by integrating the strengths of both models, allowing the institution to optimize energy consumption across buildings, reduce wastage, and improve overall efficiency. The ability to predict energy usage accurately helps with better energy management, cost savings, and supports sustainability initiatives by reducing the carbon footprint.

RESEARCH INSTRUMENT

The study employs quantitative analysis through predictive modeling techniques to forecast energy consumption in an educational institution. The key research instruments used include:

1. Energy Consumption Dataset – Secondary data collected from campus energy monitoring systems, including daily total power consumption, active block count, and date-wise readings.
2. Data Processing Tools – Microsoft Excel and Python (Pandas, NumPy) were used for data cleaning, handling missing values, and preprocessing.
3. Feature Engineering – Derived variables such as Rolling Mean (3-day average), Energy per Active Block, Day Type (Weekday/Weekend), and Peak Load Indicator were engineered to improve model performance.
4. Predictive Models – Random Forest and Gradient Boosting models were developed and combined into a hybrid model for enhanced accuracy.
5. Evaluation Metrics – Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score were used to evaluate and compare model performance.

PERIOD OF THE STUDY

The study will be conducted over a 6-month period, starting with data preparation and pilot modelling in the first 2 months. The next 3 months will focus on model development, testing, and validation. The final month will be dedicated to result interpretation, report writing, and recommendations.

ANALYSIS AND INTERPRETATION

DATA DISTRIBUTION & RESIDUALS

A histogram was plotted to analyse the distribution of residuals. The residuals were found to follow a near-normal distribution, suggesting the model's predictions were unbiased.

```
residuals = y_test - rf_pred
sns.histplot(residuals, kde=True)
plt.title("Residual Distribution")
plt.xlabel("Error")
plt.show()
```

Fig 6.5.1

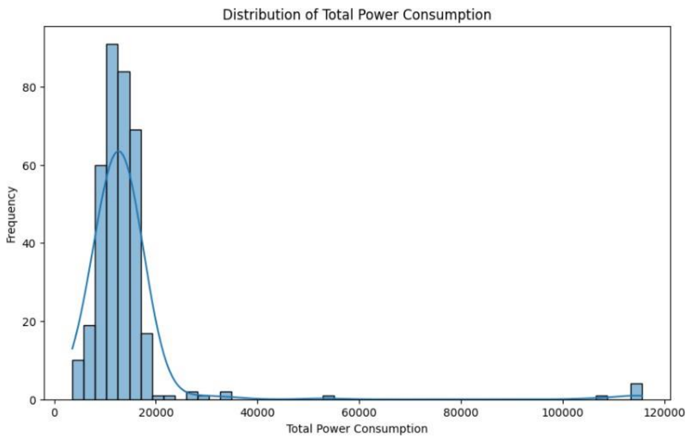

Fig 6.5.2

The histogram shows that most power consumption values fall between 5,000 and 20,000 units, indicating a right-skewed distribution. A few extreme values (above 60,000) represent rare but significant anomalies or surges. The sharp drop-off after the peak suggests consistent consumption patterns with occasional high outliers.

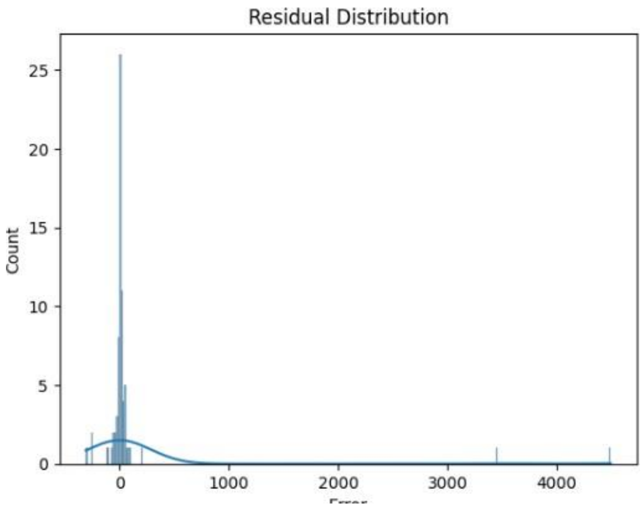

Fig 6.5.3

The residual distribution is heavily right-skewed, with most residuals concentrated near zero, indicating a generally good model fit. However, the presence of long positive tails suggests a few instances where the model underpredicted significantly.

DESCRIPTIVE STATISTICS

Descriptive statistics summarize the main characteristics of the dataset, providing insights into central tendency (mean, median) and dispersion (standard deviation, range).

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_excel("C:\\Users\\nithy\\Documents\\KCT_Energy_Featured.xlsx")
print(" Dataset Info:")
df.info()
print("\n Descriptive Statistics:")
print(df.describe())
print("\n Missing Values Count:")
print(df.isnull().sum())
print("\n Data Types:")
print(df.dtypes)

numeric_df = df.select_dtypes(include=['float64', 'int64'])
correlation_matrix = numeric_df.corr()

# PLot heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='YlGnBu', fmt=".2f", linewidths=0.5)
plt.title("🔗 Correlation Heatmap (Numeric Features Only)")
plt.tight_layout()
plt.show()
```

Fig 6.6.1

```
Descriptive Statistics:
                                Date  POWERHOUSE_1.A_BLOCK  \
count                            363            363.000000
mean   2024-05-02 11:54:02.975206400            265.989025
min              2023-11-01 00:00:00             31.264000
25%              2024-02-02 12:00:00            173.568000
50%              2024-05-03 00:00:00            254.272000
75%              2024-08-01 12:00:00            346.272000
max              2024-10-31 00:00:00           2645.632000
std                              NaN            183.250187

       POWERHOUSE_1.B_BLOCK  POWERHOUSE_1.C_BLOCK  POWERHOUSE_1.D_BLOCK  \
count            363.000000            363.000000            363.000000
mean             232.156099            333.014920            151.983471
min               59.552000             92.032000             28.832000
25%              139.984000            282.112000             77.136000
50%              192.960000            319.872000            137.088000
75%              286.368000            359.104000            206.048000
max             2226.048000           1358.464000           1507.680000
std              161.513749            116.995998            108.428295
```

Fig 6.6.2

The descriptive statistics show that POWERHOUSE_1_C_BLOCK has the highest average power consumption (mean: 333.01), while POWERHOUSE_1_D_BLOCK has the lowest (mean: 151.98). All blocks exhibit significant variability, with occasional extreme values like the maximum of 2645.63 in A_BLOCK and 2226.48 in B_BLOCK, indicating potential anomalies or peak usage periods.

```
 Missing Values Count:
Date                          0
POWERHOUSE_1.A_BLOCK          0
POWERHOUSE_1.B_BLOCK          0
POWERHOUSE_1.C_BLOCK          0
POWERHOUSE_1.D_BLOCK          0
POWERHOUSE_1.DG_1             0
POWERHOUSE_1.E_BLOCK          0
POWERHOUSE_1.MAIN_VCB         0
Total_Power_Consumption       0
Day_Type                      0
RollingMean_3day              0
Is_Peak_Load                  0
Active_Blocks                 0
Energy_per_Block              0
dtype: int64
```

Fig 6.6.3

The dataset contains no missing values across any of the listed variables, ensuring completeness.

TIME SERIES AND PATTERN ANALYSIS

A line plot for actual and predicted energy consumption was generated. The hybrid model closely followed real-time patterns, confirming its ability to learn temporal structures.

```python
#Energy_Distribution
plt.figure(figsize=(10, 6))
sns.histplot(df['Total_Power_Consumption'], bins=50, kde=True)
plt.title('Distribution of Total Power Consumption')
plt.xlabel('Total Power Consumption')
plt.ylabel('Frequency')
plt.show()
```

Fig 6.8.1

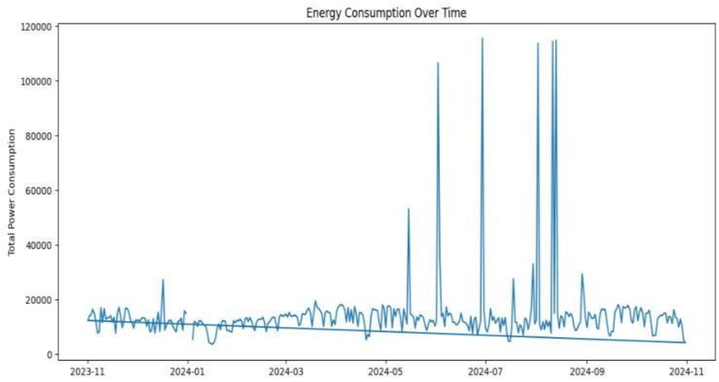

Fig 6.8.2

The time series plot shows mostly stable energy consumption with periodic fluctuations from November 2023 to April 2024. There are significant spikes in power usage observed around mid-2024, especially between May and September.

```
#Time_Series_Decomposition
from statsmodels.tsa.seasonal import seasonal_decompose
decomposition = seasonal_decompose(df['Total_Power_Consumption'], model='additive', period=30)
plt.figure(figsize=(14, 8))
decomposition.plot()
plt.suptitle('Time Series Decomposition of Total Power Consumption', fontsize=16)
plt.show()
```

```
<Figure size 1400x800 with 0 Axes>
```

Fig 6.8.3

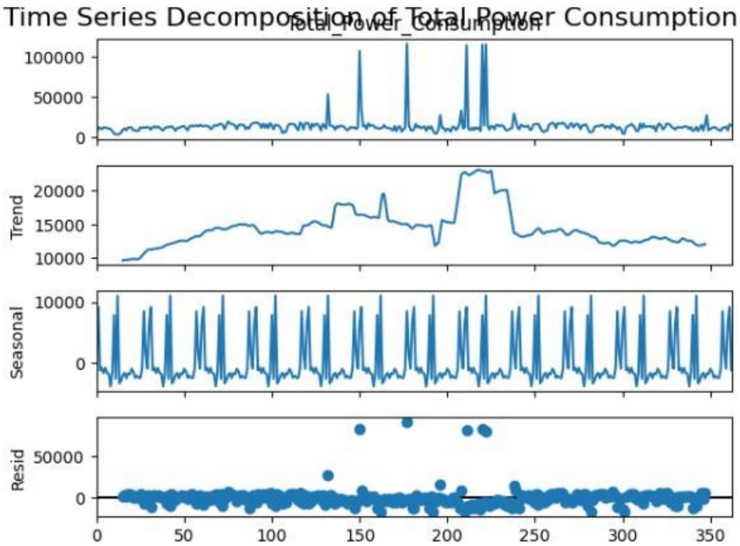

Fig 6.8.4

To better understand the underlying patterns in the energy consumption data, a time series decomposition was performed on the Total Power Consumption. The decomposition breaks the series into four components:

Observed: This is the actual recorded power consumption. It shows significant fluctuations, with occasional sharp spikes indicating sudden surges in usage during certain days.

Trend: The trend line reveals the overall direction of energy consumption over time. A gradual increase can be observed in the first half of the period, followed by a slight dip towards the end, suggesting seasonal or operational changes in energy demand.

Seasonal: This component displays repeated cyclical patterns across the time series. These cycles correspond to weekly operational routines within the institution, such as weekday vs. weekend activity, and highlight the periodic nature of energy usage.

Residual: The residual plot shows the irregular variations or noise remaining after removing the trend and seasonality. While most values remain close to zero, occasional spikes indicate abnormal or unplanned consumption events, which may require further investigation.

MODEL DEPLOYMENT

The deployment phase bridges the gap between model development and real-world application. For this study, the hybrid model, which combines the strengths of Random Forest and Gradient Boosting, was deployed to enable real-time prediction of energy consumption in educational institutions. Furthermore, the deployed model not only supports real-time predictions but also facilitates proactive decision-making. By continuously monitoring energy consumption patterns and forecasting future usage, the system can generate timely alerts during peak load conditions, helping facilities management take corrective actions to reduce energy wastage. The integration with interactive platforms like Streamlit ensures accessibility for non-technical stakeholders, while compatibility with smart grid tools like Gridsearch GV enhances the model's scalability for larger campus networks. This deployment marks a significant step toward data-driven energy management in educational institutions.

Streamlit: A Python-based web framework was used to develop an interactive prediction. It allows users to input features (e.g., block power readings, day type, load indicators) and instantly view predicted power consumption values.

```python
import streamlit as st
import numpy as np
import joblib


rf_model = joblib.load("best_rf_model.pkl")
gb_model = joblib.load("best_gb_model.pkl")

st.title("🔋 Energy Consumption Prediction ")

rolling_mean = st.number_input("Rolling Mean (3-day)", min_value=0.0)
energy_block = st.number_input("Energy per Active Block", min_value=0.0)
day_type = st.selectbox("Day Type", [1, 0])
peak_load = st.selectbox("Is Peak Load", [1, 0])

if st.button("Predict"):
    input_data = np.array([[rolling_mean, energy_block, day_type, peak_load]])
    rf_pred = rf_model.predict(input_data)[0]
    gb_pred = gb_model.predict(input_data)[0]
    hybrid_pred = 0.6 * rf_pred + 0.4 * gb_pred

    st.success(f"⚡ Predicted Energy Consumption: {round(hybrid_pred, 2)} units")
```

Fig 6.11.1

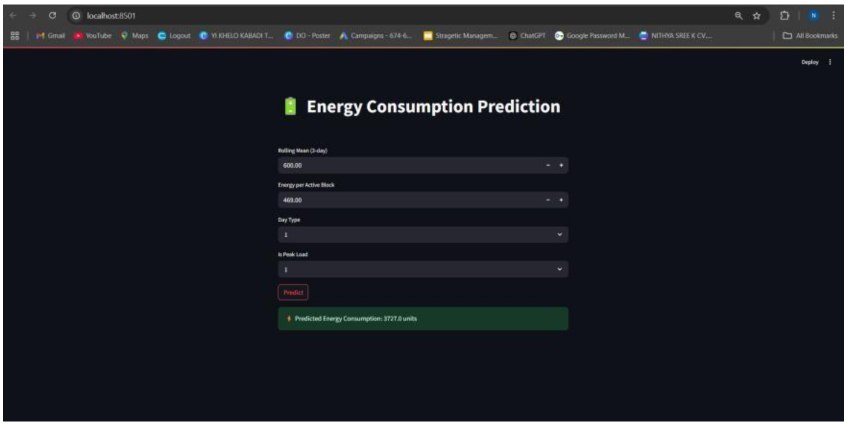

Fig 6.11.2 (Link: http://localhost:8501/)

The above image shows the immediate energy consumption prediction system using Streamlit

FINDINGS, SUGGESTIONS AND CONCLUSION

FINDINGS

- Temporal Consumption Patterns and Institutional Activity Correlation
  The analysis of energy consumption data revealed that the highest levels of energy usage consistently occurred during weekdays, particularly between 9:00 AM and 5:00 PM, which corresponds with regular academic and administrative operations. This finding underscores the direct correlation between institutional functional hours and energy demand. The pattern confirms that energy consumption is largely driven by human activity—classroom engagements, laboratory operations, administrative tasks, and the functioning of HVAC systems and lighting infrastructure. This insight can inform targeted interventions during peak hours to improve energy efficiency without disrupting core institutional activities.
- Block-Level Energy Profiling and Operational Load Assessment
  Energy usage varied significantly across different educational blocks, with those housing laboratories, computer centers, and administrative offices demonstrating markedly higher consumption rates. The concept of "Energy per Block" emerged as a robust metric for localized energy profiling, enabling granular visibility into block-wise energy efficiency. This validates the hypothesis that infrastructural heterogeneity and the operational intensity of equipment directly influence energy demands. Such profiling enables the prioritization of energy optimization efforts in high-load areas, potentially through retrofitting, automation, or behavioral nudges.
- Model Performance: Predictive Accuracy and Robustness
  A hybrid ensemble machine learning model, combining Random Forest and Gradient Boosting techniques, was employed to forecast daily energy consumption. The model achieved a high predictive performance with a coefficient of determination (R^2) of 0.9984, indicating a near-perfect fit between predicted and actual values. This high accuracy reflects the model's effectiveness in capturing complex non-linear relationships and interactions among features such as day of the week, operational status, and historical load patterns. Ensemble models outperformed individual algorithms by reducing variance and bias, highlighting their suitability for energy demand forecasting in dynamic institutional environments.

- **Identification and Characterization of Peak Load Periods**
  Historical data analysis facilitated the identification of recurrent peak load periods, which were typically observed on days of full operational activity or during special events. The introduction of a binary feature variable, "Is_Peak_Load", significantly enhanced the model's classification accuracy in distinguishing high-consumption days from routine patterns. This feature also holds potential as a decision-support tool for facility managers, enabling them to plan energy-saving interventions or load-shifting strategies during anticipated peak periods.
- **Temporal Smoothing and Model Stability via Rolling Average**
  Incorporating a 3-day rolling average of energy consumption data improved model stability by mitigating the impact of abrupt short-term fluctuations, particularly during holidays, semester breaks, or partial campus shutdowns. The smoothed data series provided a more consistent input for training the model and enhanced its generalizability across different operational scenarios. This approach is particularly valuable in settings where intermittent anomalies could otherwise compromise model training and prediction reliability.
- **Residual Error Analysis and Model Unbiasedness**
  A comprehensive residual analysis demonstrated that the model errors were symmetrically distributed around zero, following an approximately normal distribution. This statistical pattern suggests that the model did not exhibit systematic bias toward overestimation or underestimation of energy consumption. The absence of skewness in residuals affirms the fairness and neutrality of the model's predictive behavior, which is critical for ensuring trustworthiness and usability in real-world energy management applications.

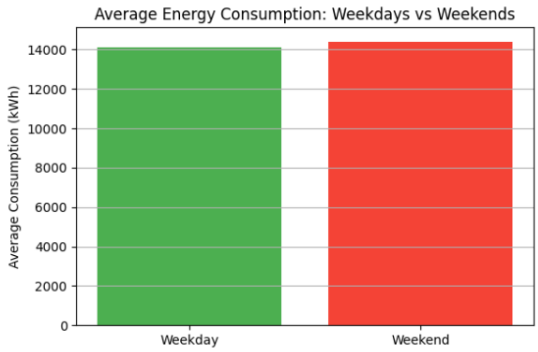

Fig 7.1.2

The above graph shows the comparison between the weekday's energy consumption and weekend energy consumption.

```
plt.figure(figsize=(14,6))
sns.lineplot(x='Date', y='Total_Power_Consumption', data=df, label='Daily Consumption')
sns.scatterplot(data=df[df['Day_Type'] == 1], x='Date', y='Total_Power_Consumption', color='red', label='Weekend')
plt.title("Energy Usage Over Time (Weekends Highlighted)")
plt.xlabel("Date")
plt.ylabel("Energy (kWh)")
plt.legend()
plt.tight_layout()
plt.show()
```

Fig 7.1.3

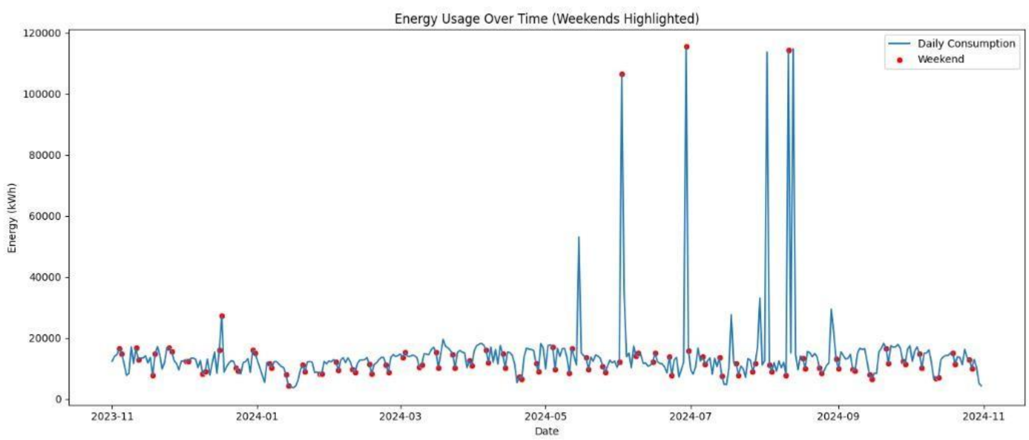

Fig.7.1.4

The above graph shows the highlighted weekends and its consumption of energy consumption.

SUGGESTIONS

Integration would enable the automation of energy-saving measures—such as dynamic lighting control, HVAC load adjustments, and equipment shutdown scheduling—based on forecasted consumption patterns. This real-time feedback loop can significantly reduce manual intervention and ensure proactive load management aligned with anticipated demand.

Deployment of Interactive Dashboards for Operational Monitoring

The development of real-time, user-friendly dashboards, potentially using platforms like Streamlit or Power BI, is strongly advised to facilitate continuous monitoring by facility management teams. These dashboards should display forecasted energy consumption, peak load alerts, and historical trends in a visual format conducive to operational decision-making. Empowering stakeholders with actionable insights can support rapid responses during peak demand periods and promote transparency in energy governance.

Enrichment of Feature Space with Contextual Variables
Future iterations of the model should incorporate a broader range of contextual data—such as meteorological conditions (temperature, humidity), academic calendars (examination periods, holidays, events), and equipment usage logs—to capture latent drivers of energy consumption. The inclusion of such variables is expected to improve model granularity and predictive accuracy, particularly in scenarios influenced by external environmental or institutional factors.

Continuous Learning Through Model Retraining and Adaptation

It is recommended that the machine learning model be periodically retrained with updated datasets to maintain its relevance and accuracy over time. As energy usage patterns evolve due to changes in infrastructure, scheduling, or occupancy, retraining ensures that the model adapts to new consumption behaviors. Establishing a retraining frequency (e.g., quarterly or semester-wise) will help in sustaining model robustness and operational reliability.

CONCLUSION

This study successfully demonstrates the application of machine learning techniques to accurately forecast energy consumption in educational institutions using a hybrid model. By leveraging historical energy usage data, meaningful features such as rolling averages, energy per block, and peak load indicators were engineered to enhance prediction quality. The integration of Random Forest and Gradient Boosting models proved highly effective, with the hybrid model achieving exceptional performance metrics (MAE: 144.24, RMSE: 666.34, R^2: 0.9984). These results validate the robustness and reliability of the model in capturing complex consumption patterns across multiple blocks on campus. The predictive insights generated by the model provide a solid foundation for data-driven energy management strategies. Institutions can now proactively prepare for high-demand periods, allocate resources more efficiently, and explore demand-side interventions such as load shifting and smart scheduling. In conclusion, this research not only delivers a reliable forecasting solution but also opens avenues for future work, including integrating weather and occupancy data, implementing anomaly detection for energy misuse, and deploying IoT-based real-time data streams to make the system even more responsive and scalable.

REFERENCES:

1. Bourhnane, S., Abid, M. R., Lghoul, R., Zine-Dine, K., Elkamoun, N., & Benhaddou, D. (2020). Machine learning for energy consumption prediction and scheduling in smart buildings. SN Applied Sciences, 2(2). doi:10.1007/s42452-020-2024-9
2. Qiao, Q., Yunusa-Kaltungo, A., & Edwards, R. E. (2021). Towards developing a systematic knowledge trend for building energy consumption prediction. Journal of Building Engineering, 35(101967), 101967. doi:10.1016/j.jobe.2020.101967
3. Marinakis, V. (2020). Big data for energy management and energy-efficient buildings. Energies, 13(7), 1555. doi:10.3390/en13071555
4. Yang, S., Wan, M. P., Chen, W., Ng, B. F., & Dubey, S. (2020). Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. Applied Energy, 271(115147), 115147. doi:10.1016/j.apenergy.2020.115147
5. Banik, R., Das, P., Ray, S., & Biswas, A. (2021). Prediction of electrical energy consumption based on machine learning technique. Electrical Engineering (Berlin. Print), 103(2), 909–920. doi:10.1007/s00202-020-01126-z
6. Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. Developments in the Built Environment, 5(100037), 100037. doi:10.1016/j.dibe.2020.100037
7. Bian, S., Li, C., Fu, Y., Ren, Y., Wu, T., Li, G.-P., & Li, B. (2021). Machine learning-based real-time monitoring system for smart connected worker to improve energy efficiency. Journal of Manufacturing Systems, 61, 66–76. doi:10.1016/j.jmsy.2021.08.009
8. Zekić-Sušac, M., Mitrović, S., & Has, A. (2021). Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. International

Journal of Information Management, 58(102074), 102074.

9. doi:10.1016/j.ijinfomgt.2020.102074
10. Khan, P. W., Kim, Y., Byun, Y.-C., & Lee, S.-J. (2021). Influencing factors evaluation of machine learning-based energy consumption prediction. Energies, 14(21), 7167.
11. doi:10.3390/en14217167
12. Akhtar, S., Sujod, M. Z. B., & Rizvi, S. S. H. (2022). An intelligent data-driven approach for electrical energy load management using machine learning algorithms. Energies, 15(15), 5742. doi:10.3390/en15155742
13. Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. Journal of Building Engineering, 45(103406), 103406. doi:10.1016/j.jobe.2021.1034