CHAPTER 1
INTRODUCTION

1.1. GENERAL

This project addresses the growing need for efficient energy consumption management, particularly as global energy demands continue to rise. By leveraging predictive machine learning technology, the system helps both residential and commercial users understand, monitor, and optimize their energy usage. In a world increasingly focused on environmental sustainability, the system provides data-driven insights that foster energy-efficient behaviors and reduce carbon footprints.

At its core, the system uses the XGBoost algorithm, analyzing a combination of historical energy data, weather conditions, and occupancy patterns to predict future energy consumption with high accuracy. These insights enable users to make informed decisions, reduce waste, and lower energy costs. For example, the system can detect when devices are left running unnecessarily and send alerts to users, prompting them to take action and avoid unnecessary energy consumption. This functionality helps users save money and supports broader efforts in sustainable resource management.

Beyond basic tracking, the machine learning capabilities of the system transform how energy consumption is approached. As more data is collected, the algorithm's accuracy improves, delivering even greater benefits over time. The system empowers users to manage their energy habits more effectively, contributing to responsible consumption. By promoting smarter energy use, it aligns personal cost savings with the global objective of reducing environmental impact, offering a forward-thinking solution for sustainable energy management.

1.2. NEED FOR THE STUDY

The need for this study is driven by the urgent global challenges related to energy consumption and sustainability. As energy demand continues to rise, so do greenhouse gas emissions and resource depletion, placing immense pressure on the environment and threatening future energy security. Inefficiencies in energy usage remain a significant problem worldwide; many buildings and facilities lack real-time insights into their consumption patterns, leading to substantial waste. Unused devices are often left running, heating and cooling systems operate unnecessarily, and peak-hour usage goes unmanaged—all contributing to higher costs and environmental impact.

This study explores the potential of predictive analysis in tackling these inefficiencies by anticipating energy needs and optimizing usage patterns. Predictive technologies such as machine learning can analyze vast datasets, from historical energy usage to external factors like weather and occupancy rates, to forecast energy demand accurately. With these insights, users can make informed adjustments, such as scheduling high-consumption tasks during off-peak hours or reducing energy output when not needed.

By understanding consumption trends and taking proactive steps to prevent waste, users can not only achieve cost savings but also contribute to the broader goals of reducing carbon footprints and conserving resources. The study's focus on applying predictive models like XGBoost to energy management highlights a practical and impactful way to address the growing energy crisis. With precise consumption forecasts, our project enables a new level of control, allowing both residential and commercial users to adopt more sustainable practices and better manage their energy needs in alignment with environmental priorities.

1.3 OBJECTIVES OF THE STUDY

The primary aim of this study is to create a comprehensive, user-friendly solution for energy consumption management that leverages predictive analytics and real-time insights. This will empower users to make informed, sustainable choices in their energy usage.

1. Developing a Predictive Model for Energy Consumption

The study's core objective is to design a predictive model capable of accurately forecasting energy consumption in residential and commercial buildings. Using the XGBoost algorithm, the model will analyze historical data, weather patterns, and occupancy information to predict daily and hourly energy demand.

2. Enhancing Predictive Accuracy and Providing Actionable Insights

Another objective is to enhance predictive accuracy by utilizing advanced machine learning techniques to analyze energy consumption patterns. The system will focus on processing diverse data sources, such as historical energy usage and weather conditions, to deliver actionable insights. These insights will help users understand their energy consumption trends.

3. Generating Insights for Users

By analyzing consumption patterns, the system will provide users with personalized recommendations, such as identifying devices left running or suggesting optimal usage times. These insights are designed to help users achieve better energy efficiency.

4. Supporting Sustainable Energy Management

This study aims to promote sustainable practices by encouraging users to actively monitor and adjust their energy usage. With access to predictive insights, users will be better equipped to align their consumption with sustainability goals, reducing their carbon footprint and contributing to environmental preservation.

1.4. OVERVIEW OF THE PROJECT

This project aims to address the pressing need for efficient energy consumption management by leveraging data-driven insights and predictive technology. It integrates various components to provide a comprehensive solution for optimizing energy use in both residential and commercial spaces, enabling users to make informed decisions that support energy conservation. The key elements of the project include data collection and initial exploration, feature engineering, model training and development, model evaluation and interpretation, and the generation of user insights.

1. Data Collection and Initial Exploration

The project begins by gathering data from multiple sources, including historical energy consumption records, weather data, and occupancy information. This data is crucial for building an accurate model that reflects real-world usage patterns. The initial data exploration phase involves cleaning and formatting the data, ensuring consistency, handling missing values, and conducting basic analyses to understand trends, distributions, and potential outliers. This stage is essential for identifying preprocessing steps necessary for effective feature engineering and model training.

2. Feature Exploration

After collecting and cleaning the data, the next step is feature exploration. This involves analyzing various factors that could influence energy consumption, such as time of day, temperature, humidity, and occupancy patterns. Feature engineering plays a critical role in improving the model's predictive accuracy by creating new variables. For example, combining temperature and humidity into a single feature or identifying weekend usage patterns can help the model capture complex relationships that affect energy use. These engineered features enhance the model’s ability to learn intricate consumption patterns.

3. Model Training and Development

With a well-prepared dataset, the next phase involves training the predictive model using advanced algorithms, such as XGBoost, known for its high performance in predictive analytics. The model is trained on historical energy data, incorporating the engineered features to identify patterns and make accurate predictions about future energy consumption. Hyperparameter tuning is conducted to optimize the model's performance, ensuring that it can generalize effectively across different conditions. This stage is iterative, involving adjustments to model parameters and features to refine accuracy.

4. Model Evaluation and Interpretation

Once the model is trained, it undergoes thorough evaluation using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics provide a quantitative measure of the model's accuracy and reliability. In addition to evaluation, model interpretation is vital for understanding the factors that influence energy consumption predictions. By identifying which features have the most significant impact, users and stakeholders can make more informed decisions and pinpoint areas for improvement in energy use.

5. Generating User Insights

The final step involves transforming the model's predictions into actionable, user-friendly insights. An interactive dashboard is developed to display real-time energy data, historical trends, and future forecasts. Additionally, a notification system is implemented to alert users to unusual consumption patterns or provide reminders for optimizing energy usage. By generating personalized insights, the system empowers users to make data-driven decisions that promote sustainable energy practices, reduce their energy footprint, and help them achieve cost savings.

CHAPTER 2
REVIEW OF LITERATURE

2.1. INTRODUCTION

In today's world, energy conservation is essential not only for reducing operational costs but also for addressing global environmental challenges. With rapid urbanization and increasing reliance on energy-intensive technologies, inefficient energy usage has become a critical issue in both residential and commercial sectors. This project addresses these challenges by harnessing predictive technology to optimize energy consumption, ultimately contributing to sustainable energy practices. The system utilizes advanced machine learning techniques, specifically the XGBoost algorithm, to forecast energy usage accurately based on historical data, weather conditions, and occupancy information.

The system operates by transforming raw data into actionable insights. Data from building energy consumption patterns is collected, processed, and then fed into the predictive model to assess daily and seasonal trends in energy use. By incorporating factors like weather and occupancy rates, the model captures fluctuations in demand, enabling users to understand and potentially adjust their consumption behaviors. These insights are then made accessible through an interactive dashboard, allowing users to monitor real-time data, track consumption patterns, and receive alerts regarding anomalies or excess usage.

The need for such predictive tools has grown in light of increasing energy costs and heightened awareness of environmental issues. The system not only aims to improve energy efficiency but also seeks to empower users by providing data-driven recommendations, such as identifying unused appliances that can be turned off. Ultimately, the project supports energy-saving measures and encourages sustainable behaviors, contributing to a more energy-efficient future.

2.2 LITERAURE REVIEW

S. No.	Author Name	Paper Title	Description	Journal	Year
1	L. H. M. Truong, K. H. K. Chow, R. Luevisadpaibul, G. S. Thirunavukkarasu, M. Seyedmahmoudian, B. Horan, S. Mekhilef, A. Stojcevski	Accurate Prediction of Hourly Energy Consumption in a Residential Building	This study focuses on the use of machine learning approaches to predict hourly energy consumption in residential buildings based on occupancy rates.	Applied Sciences	2021
2	Z. Wang, T. Hong, H. Li, et al.	Predicting City-Scale Daily Electricity Consumption Using Data-Driven Models	Examines the use of data-driven models to predict daily electricity consumption across an entire city, improving the accuracy of large-scale forecasts.	Adaptive Energy	2021
3	J. L. Harris and L.-M. Liu	Dynamic Structural Analysis and Forecasting of Residential Electricity Consumption	Analyzes and forecasts residential electricity consumption, using dynamic structural analysis to enhance predictive capabilities in energy management.	Economics and Forecasting	2020
4	R. K. Jain, K. M. Smith, P. J. Culligan, J. E. Taylor	Forecasting Energy Consumption of Multi-	Investigates the impact of monitoring granularity on the accuracy of support	Energy and Buildings	2019

		Family Residential Buildings Using Support Vector Regression	vector regression models in forecasting multi-family residential energy		
5	Y.-S. Kim, J. Srebric	Impact of Occupancy Rates on the Building Electricity Consumption in Commercial Buildings	Explores the effects of occupancy rates on electricity consumption in commercial buildings, highlighting the significance of occupancy patterns	Energy and Buildings	2016

Table 1. Review of Literature

2.3. FRAMEWORK OF LCA

The framework of Life Cycle Assessment (LCA) for this project is designed to evaluate the environmental impact of energy consumption throughout its entire lifecycle, with a focus on reducing energy waste and promoting sustainable practices. The first phase, goal and scope definition, outlines the primary objective of minimizing energy consumption and environmental impacts by providing accurate predictions of energy usage based on historical data, weather conditions, and occupancy patterns. This scope covers all stages of energy consumption, from data collection to the generation of actionable insights for users. The functional unit is the energy consumption of a building, with the ultimate goal being the reduction of unnecessary energy usage.

The second phase, inventory analysis, involves collecting data on energy consumption, weather variables, and occupancy patterns. This data forms the foundation for the predictive model, which analyzes consumption trends and identifies inefficiencies. During this phase, the project also captures the energy inputs, such as electricity usage and external factors like temperature, as well as the outputs, including potential energy savings and patterns of inefficiency.

The impact assessment phase evaluates the environmental consequences of energy consumption, with a focus on factors like carbon emissions and resource depletion. By analyzing energy usage patterns and inefficiencies, the model provides insights that help reduce carbon footprints and optimize energy consumption. The final phase, interpretation, takes these insights and translates them into actionable recommendations for users. These recommendations include strategies for adjusting consumption behaviors, such as turning off unused devices, ultimately contributing to reduced energy waste and fostering sustainable practices. This LCA framework supports the project's goal of promoting energy-efficient and environmentally responsible behaviors.

CHAPTER 3
SYSTEM OVERVIEW

3.1. EXISTING SYSTEM

Existing systems for energy management and forecasting often rely on traditional methods like manual monitoring or basic rule-based algorithms. These systems are typically limited in their ability to handle the complexities of modern energy consumption patterns, which can be influenced by numerous variables such as weather conditions, occupancy, and time of day. Conventional systems often focus on static data analysis and fail to provide real-time solutions.

Some existing systems use simple energy consumption tracking, where users are provided with historical consumption data or averages to monitor energy use. These systems, while useful, do not offer predictive capabilities or actionable insights for future energy consumption, limiting their effectiveness in reducing waste or costs. Additionally, many current solutions are not integrated with real-time data streams, which makes it difficult to address sudden changes in energy demand or inefficient behaviors in real-time.

More advanced systems, such as those employing machine learning algorithms, offer improved accuracy in predicting future energy usage. However, these systems often require complex configurations and may not be user-friendly, limiting their widespread adoption. Some machine learning-based systems use weather forecasts or building-specific data to predict energy consumption, but they may not incorporate a comprehensive set of factors, such as occupancy patterns or daily habits, which are key to accurate forecasting.

The need for a more holistic, real-time, and accessible energy management system remains. Current systems often fall short of providing comprehensive insights and actionable recommendations that can lead to real energy savings and sustainable practices.

3.2. PROPOSED SYSTEM

The proposed system introduces a cutting-edge approach to energy consumption forecasting by employing the XGBoost algorithm, a robust and efficient machine learning model. Traditional energy forecasting methods often rely on simple statistical techniques or basic data evaluations, which fail to capture the complexities of modern energy consumption patterns. This system addresses these limitations by integrating diverse datasets, including historical energy usage, weather conditions, and occupancy trends, to deliver more accurate predictions of daily and seasonal energy demands.

The XGBoost model is particularly well-suited for this application due to its ability to handle large datasets and uncover intricate, non-linear relationships between variables. By leveraging advanced gradient boosting techniques, the model adapts to patterns in the data, improving its forecasting accuracy over time. This dynamic adaptability ensures that the system can effectively account for fluctuations in energy usage, such as those caused by seasonal variations or changing occupant behavior. The use of XGBoost also minimizes errors, enhances computational efficiency, and provides a more reliable framework for understanding and predicting energy needs in diverse scenarios.

By combining powerful machine learning algorithms with comprehensive data analysis, the proposed system goes beyond simple predictions. It offers users actionable insights, such as identifying patterns of excessive energy usage, highlighting opportunities to optimize consumption, and recommending energy-efficient practices. This innovative solution not only enhances energy efficiency but also supports sustainability goals by promoting smarter energy use and reducing environmental impact, contributing to a greener and more sustainable future.

3.3. FEASIBILITY STUDY

The feasibility study for this project evaluates its technical, operational, and economic viability, ensuring that the proposed solution is both achievable and sustainable.

Technical Feasibility: The project is built on widely recognized machine learning techniques, specifically the XGBoost algorithm, which is known for its high accuracy and efficiency in predictive analytics. The system integrates data from various sources, such as historical energy consumption, weather conditions, and occupancy data, to forecast future energy needs. Leveraging Python and Google Colab for model development ensures that the technology stack is well-supported and scalable. The use of cloud-based platforms ensures data accessibility and processing power, making the system suitable for both residential and commercial users.

Operational Feasibility: From an operational perspective, the system's design is straightforward and user-friendly, ensuring that users can easily interact with the predictive analytics features through a simple interface, such as a dashboard. The system's ability to send real-time alerts for energy inefficiencies or excess consumption empowers users to take immediate actions. Furthermore, the model continuously improves with more data, ensuring long-term operational effectiveness. The integration of such a system into existing infrastructure can be achieved with minimal disruption.

Economic Feasibility: The system is economically viable, as it reduces operational costs for both residential and commercial users by optimizing energy consumption. By minimizing energy wastage, users can significantly reduce their utility bills. The project's initial costs are centered around data collection, model development, and dashboard design, while long-term benefits include cost savings for users and a positive impact on sustainability. With the increasing emphasis on energy efficiency, this project has the potential to attract both individual and business users, making it a sound investment.

Market Feasibility: The market for energy optimization solutions is growing rapidly due to the increasing demand for sustainable energy practices and cost-cutting measures across industries. With rising energy costs and the global push toward reducing carbon footprints, there is a strong market demand for tools that help manage energy consumption more efficiently. Both residential and commercial sectors are actively seeking ways to optimize energy use, which presents a significant opportunity for this project. The integration of machine learning algorithms and predictive analytics positions the system as a competitive and innovative solution in the market.

Legal and Regulatory Feasibility: The project aligns with current environmental regulations and energy efficiency standards, such as those outlined by the International Energy Agency (IEA) and local regulatory bodies. There are no major legal barriers to implementing this system, as it focuses on improving energy efficiency and promoting sustainability. However, data privacy and security will be important considerations, particularly in handling user data. Compliance with data protection regulations, such as the General Data Protection Regulation (GDPR) for European users, will be maintained to ensure the protection of personal and energy consumption data.

Environmental Feasibility: The system's core objective is to promote energy efficiency, reducing overall energy consumption and supporting sustainable energy practices. By helping users minimize waste and optimize usage, the project contributes to reducing the carbon footprint, making it environmentally beneficial and aligned with global sustainability goals. The system's long-term impact on energy conservation is a critical factor in its environmental feasibility.

CHAPTER 4
SYSTEM REQUIREMENTS

4.1. HARDWARE REQUIREMENTS

Data Processing and Analytics Server:

- **Processor (CPU):** High-performance CPU with multiple cores (e.g., Intel Xeon or AMD EPYC processors) to handle complex data processing tasks and large datasets in real-time.
- **Graphics Processing Unit (GPU):** A GPU (e.g., NVIDIA Tesla or RTX series) may be beneficial for accelerating machine learning tasks, especially during training phases that involve large datasets. This is particularly useful for deep learning models, if applicable in the future.
- **RAM:** 32 GB or more for handling the large volume of data and ensuring efficient operation during model training, real-time analysis, and prediction.

Network Infrastructure:

- **Router/Switch:** High-speed and reliable networking hardware, such as Gigabit Ethernet or 10Gbps switches, to handle the flow of data between sensors, cloud storage, and user interfaces with minimal latency.
- **Firewall and Security Infrastructure:** Ensuring secure communication and data exchange between cloud-based servers and client systems is crucial for privacy and data integrity.

4.2. SOFTWARE REQUIREMENTS

Operating System:

- **Windows/Linux/MacOS**: The system can be deployed on any of the major operating systems. Linux is preferred for its scalability and efficiency in handling large datasets and supporting the necessary software tools.

Programming Languages:

- **Python**: The core programming language for developing the machine learning model (XGBoost), data processing, and analysis. Python provides a wide array of libraries (like pandas, numpy, and matplotlib) for data manipulation, visualization, and model training.

Machine Learning Libraries:

- **XGBoost**: This is the primary library for implementing the machine learning model used for forecasting energy consumption. It is highly efficient for predictive tasks and is known for its scalability and performance.
- **Scikit-learn**: Useful for additional machine learning techniques, such as model evaluation, hyperparameter tuning, and preprocessing.
- **TensorFlow/Keras (optional)**: These deep learning frameworks may be considered if more complex modeling approaches are needed in the future.

CHAPTER 5

SYSTEM DESIGN

5.1. SYSTEM ARCHITECTURE

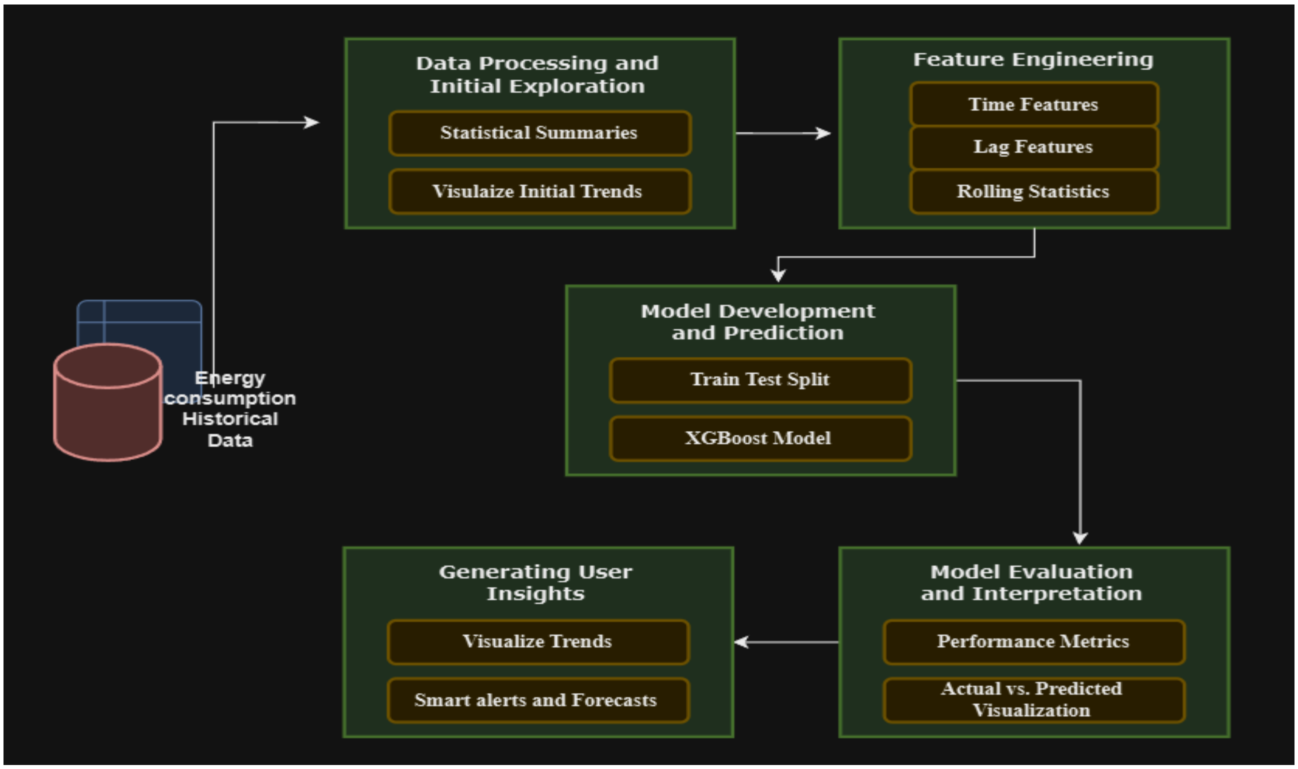

Fig. 5.1. System Architecture

The system architecture is organized into five essential modules: data collection and initial exploration, feature engineering, model development and prediction, model evaluation and interpretation, and user insights generation. The architecture first gathers and prepares data, then applies feature engineering to enhance predictive accuracy. The refined data is used to train the model, which is subsequently evaluated to ensure reliability. Finally, insights are generated for users through an accessible dashboard, empowering energy-efficient decisions.

5.2. MODULE DESCRIPTION

5.2.1. Data collection and initial exploration module

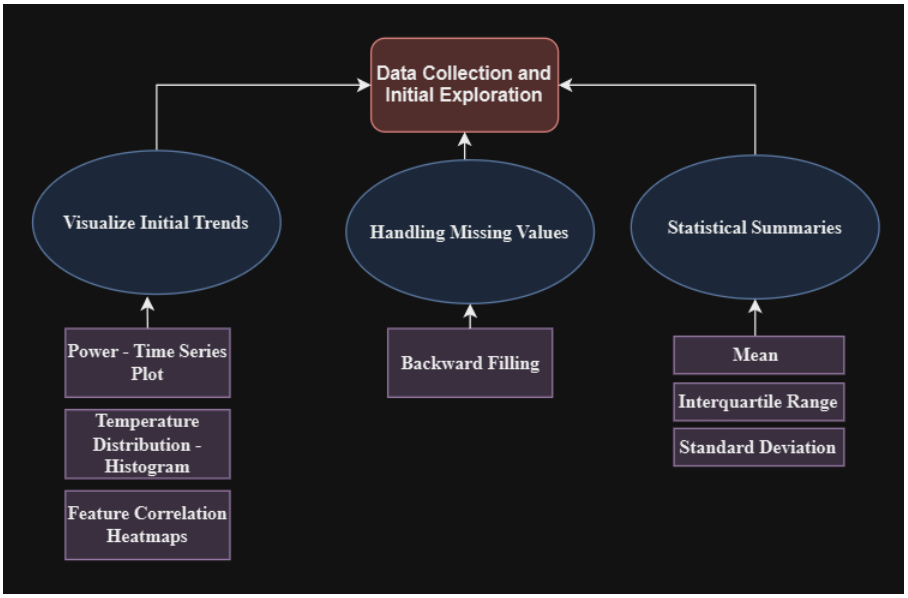

Fig. 5.2.1. Data Collection and Initial Exploration Module

The *Data Collection and Initial Exploration* module as mentioned in the **Fig. 5.2.1**, serves as a foundation for building an effective predictive model by focusing on data quality, time-based trends, and preliminary analyses. Initially, missing values are identified and resolved to ensure data integrity. With the dataset structured, relevant time-based features like hour, day, and month are extracted to capture daily, weekly, and seasonal fluctuations in energy consumption. Visualizations such as time series plots of active power, temperature histograms, and feature correlations bring key insights to the factors influencing energy usage. The correlation heatmap, helps identify variables with strong predictive power, offering direction for feature engineering. Together, these steps not only provide a comprehensive view of consumption patterns and their drivers but also enable a data-driven approach for creating a highly accurate forecasting model. This groundwork ensures that the model is robust and capable of capturing real-world energy consumption dynamics effectively.

5.2.2. Feature Engineering module

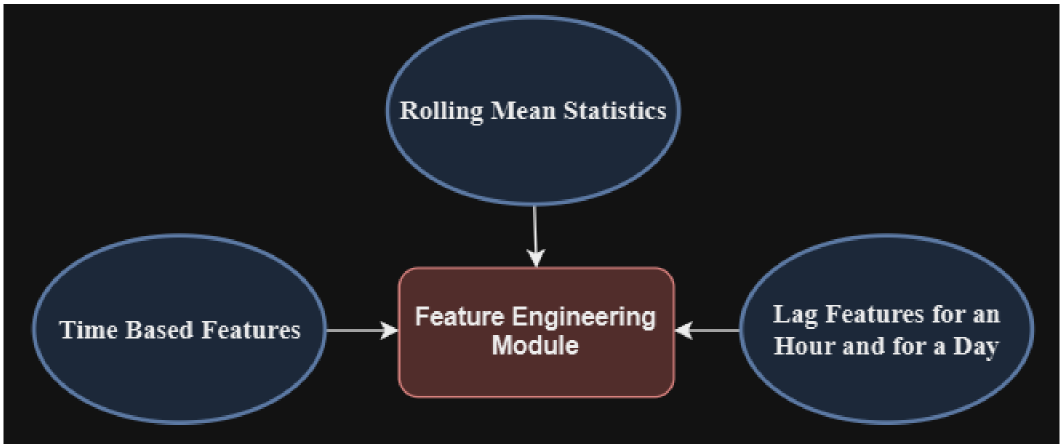

Fig. 5.2.2. Feature Engineering

The *Feature Engineering* module is crafted to transform raw data into features that significantly enhance model accuracy and insight. As mentioned in the **Fig. 5.2.2**, this module beginning with time-based patterns, hourly, daily, and monthly average consumption metrics are calculated to uncover trends in energy usage, revealing when peak demands occur. To further capture temporal dynamics, we introduce lagged variables representing previous hour and day consumption, which are critical in understanding the dependencies over time. Rolling statistics, like the 24-hour rolling mean and standard deviation, provide smoothed views of energy trends, helping detect anomalies and cyclical behaviors.

In addition, interaction features are engineered, such as the product of temperature and humidity, to better capture environmental influences on energy demand. These features are visually assessed through correlation heatmaps, highlighting relationships among variables. After engineering, missing values are backfilled to ensure data continuity, resulting in a robust and enriched dataset. This comprehensive set of features enhances the model's ability to make precise energy consumption predictions, ultimately aiding users in making informed, energy- efficient decisions.

5.2.3. Model Training and Prediction module

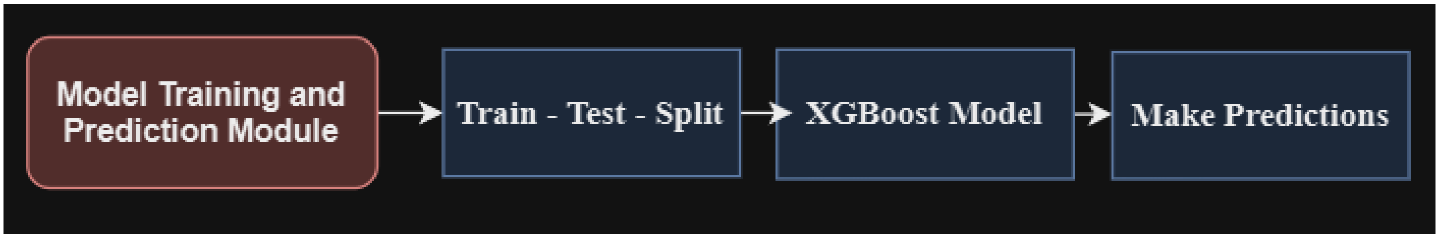

Fig. 5.2.3. Model Training and Prediction Module

As shown in the **Fig. 5.2.3**, the *Model Training and Prediction* module focuses on building and fine-tuning a predictive model using XGBoost, a powerful and efficient gradient boosting algorithm particularly suited for structured data. In this implementation, we use the XGBRegressor model, configured with the objective function set to reg:squarederror, which minimizes squared error during training— a common choice for regression tasks. Key hyperparameters, including n_estimators, learning_rate, and max_depth, are optimized to balance model complexity and accuracy.

With 100 trees (n_estimators) and a learning rate of 0.1, the model learns in increments, which helps prevent overfitting while capturing essential patterns. A max_depth of 6 controls the complexity, ensuring the model generalizes well across diverse conditions in the test set. The model is trained on historical energy data, weather, and occupancy-related features to learn correlations and dependencies in the data.

Once trained, the model is used to predict energy consumption on a separate test dataset (X_test). The resulting predictions form the basis for further analysis and provide insights into future energy consumption patterns. This module's predictive accuracy allows users to make proactive decisions to optimize their energy usage based on forecasted trends, promoting cost- efficiency and sustainability.

5.2.4. Model Evaluation and Interpretation

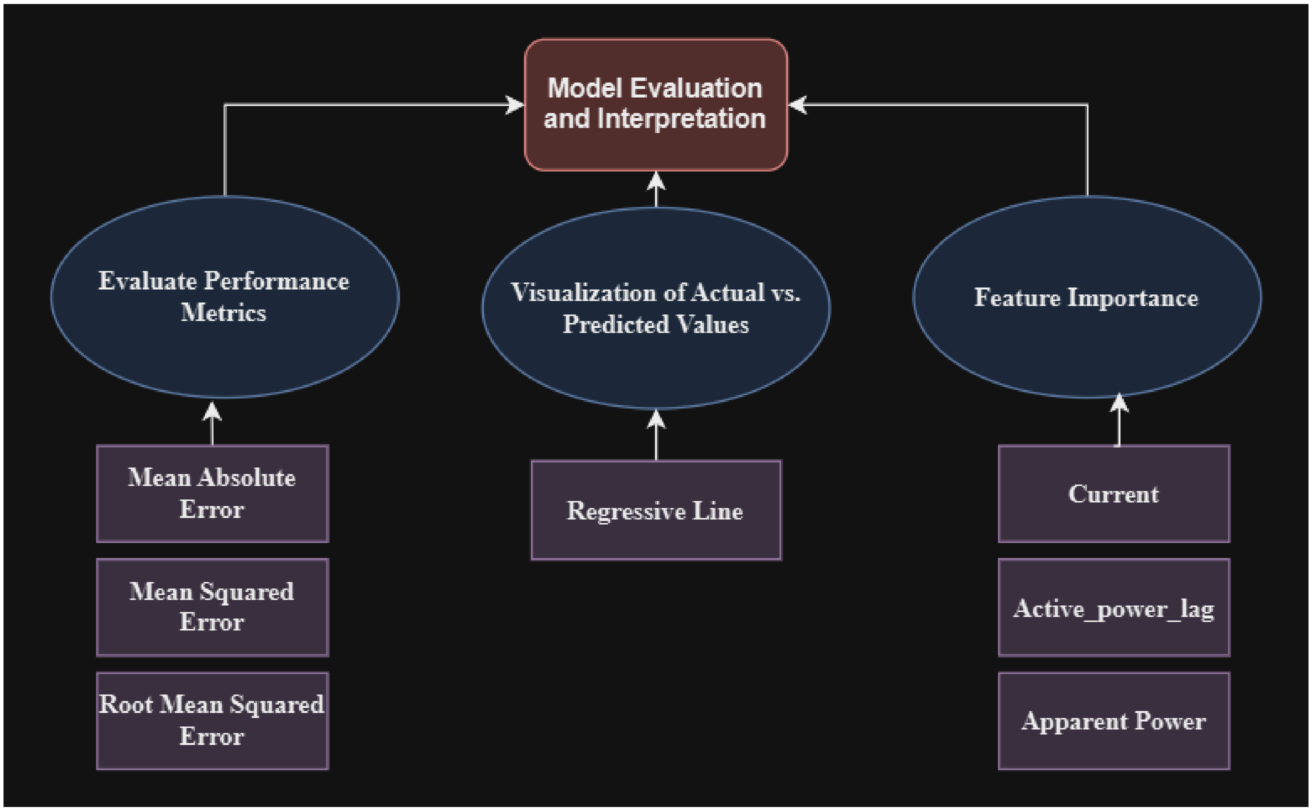

Fig. 5.2.4. Model Evaluation and Interpretation Module

The *Model Evaluation and Interpretation* module assesses the predictive performance and reliability of the trained model. As mentioned in the **Fig 5.2.4**, Key metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), offer quantitative insights into prediction accuracy. A scatter plot of predicted versus actual values, along with a residuals distribution plot, visually reveals model performance and prediction error patterns. Feature importance analysis highlights the contribution of individual features, helping users understand which factors most significantly impact energy consumption predictions.

The feature importance analysis further helps in identifying which features contribute most to the model's predictions, guiding the refinement process. By understanding the significance of each feature, the model can be adjusted for greater accuracy.

5.2.5. Generating User Insights Module

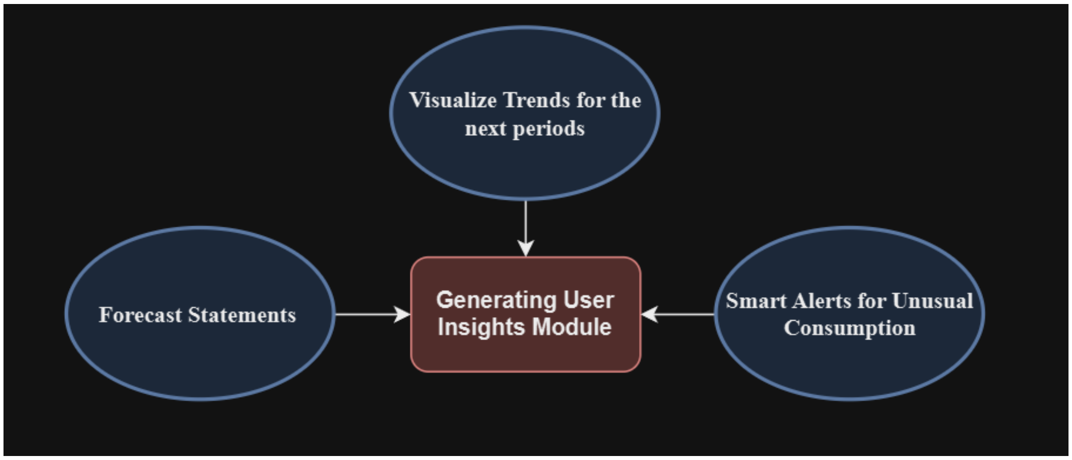

Fig. 5.2.5. Generating User Insights Module

The *Generating User Insights* module leverages the predictive model to provide valuable information to users, enhancing their ability to manage energy consumption effectively. As shown in the **Fig. 5.2.5**, the system allows users to anticipate their needs and adjust usage accordingly. It includes an alert system for high consumption, notifying users when their energy usage exceeds a specified threshold, prompting them to reduce non-essential device usage. Additionally, the module provides a detailed forecast for the next 24 periods, such as hourly or daily, giving users a clearer understanding of their consumption patterns. This empowers users to take proactive measures, ensuring better control over their energy usage while contributing to cost savings and sustainability efforts.

CHAPTER 6
RESULTS AND DISCUSSIONS

The results and discussion section of the project focuses on the effectiveness of the system in predicting and optimizing energy consumption. The system has shown promising results in analyzing energy usage patterns, identifying peak consumption periods, and offering actionable insights for users to reduce energy costs. Through various visualizations, such as time-based analysis, rolling means, and weather-related correlations, the system provides a comprehensive understanding of how energy is consumed over time.

One of the significant findings from the data analysis is the clear identification of peak energy consumption periods, which are often tied to specific hours of the day or weather conditions. By observing these trends, users can adjust their consumption habits, potentially shifting energy-intensive tasks to off-peak times, leading to cost savings. Furthermore, the inclusion of a rolling mean smoothens out short-term fluctuations, helping users identify underlying consumption trends without being misled by noise.

The system's ability to forecast energy usage for future periods was another key result, with predictions showing a strong alignment with actual consumption patterns. This indicates that the model is reliable and can be used for proactive energy management. The comparison of predicted versus actual consumption helps evaluate the accuracy of the model, highlighting areas where further improvements could be made.

Overall, the system has proven to be an effective tool for energy optimization. By offering users predictive insights, real-time monitoring, and the ability to adjust based on weather forecasts, the system not only aids in reducing energy consumption but also contributes to sustainable energy practices. Future work could involve enhancing the predictive model by incorporating additional variables and improving its scalability for larger datasets.

CHAPTER 7
CONCLUSION AND FUTURE ENHANCEMENT

7.1. CONCLUSION

In conclusion, this project successfully developed a system that leverages predictive analytics and machine learning to optimize energy consumption in residential and commercial settings. By utilizing the XGBoost algorithm, the system accurately forecasts energy usage based on historical data, weather conditions, and occupancy patterns, providing valuable insights into consumption trends. Through comprehensive data analysis and visualizations, users are empowered to make informed decisions about their energy usage, reduce waste, and lower costs. The system also offers predictive capabilities that enable proactive management, helping users adjust their habits in response to forecasted consumption patterns. Additionally, the integration of weather-related factors further enhances the model's accuracy by showing how external conditions influence energy demand. Overall, the project demonstrates the potential for machine learning to contribute to more sustainable energy practices. Future enhancements could focus on improving the model's scalability, integrating more data sources, and refining the user experience for even greater energy optimization.

7.2. FUTURE ENHANCEMENT

Future enhancements for this project can include incorporating adaptive learning techniques into the model, enabling it to continuously improve its predictions by learning from new data over time. This would allow the system to adjust to changing energy consumption patterns, user behavior, and external factors like weather conditions. Another potential improvement is integrating energy consumption data with external platforms, such as utility providers or smart home ecosystems, for better synchronization and more accurate billing forecasts.

Expanding the system's geographical scope to support diverse energy consumption patterns and climatic conditions could make it more universally applicable. Furthermore, adding gamification features, such as rewarding users for reducing energy consumption or achieving energy-saving milestones, could increase user engagement and foster a proactive approach to energy conservation. Multi-platform support, including mobile apps and voice assistants, would further enhance accessibility, allowing users to track and manage their energy use more conveniently.

Additionally, incorporating real-time energy market data could help users make informed decisions on when to use energy-intensive appliances based on fluctuating energy prices, further optimizing their energy usage and reducing costs. These enhancements would not only improve the system’s functionality and usability but also contribute to more efficient energy management, leading to greater sustainability and cost savings.

APPENDIX

A1. SAMPLE CODE

1. DATA COLLECTION AND INITIAL EXPLORATION MODULE

```
import pandas as pd
data = pd.read_csv('/content/energy_weather_raw_data.csv')
data.head()
data.info()
missing_values = data.isnull().sum()
print("Missing values per column:\n", missing_values)
data.describe()
data['date'] = pd.to_datetime(data['date'])
data['hour'] = data['date'].dt.hour
data['day'] = data['date'].dt.day
data['month'] = data['date'].dt.month
import matplotlib.pyplot as plt
plt.figure(figsize=(14,6))
plt.plot(data['date'], data['active_power'], color='orange')
plt.title('Active Power Over Time')
plt.xlabel('Date')
plt.ylabel('Active Power')
plt.show()
data['temp'].hist(bins=20)
plt.title('Temperature Distribution')
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.show()
import seaborn as sns
import matplotlib.pyplot as plt
numerical_data = data.select_dtypes(include=['number'])
plt.figure(figsize=(12, 8))
sns.heatmap(numerical_data.corr(), annot=True,cmap='YlOrRd')
plt.title('Feature Correlations')
plt.show()
```

2. FEATURE ENGINEERING MODULE

```
data['hour'] = data['date'].dt.hour
import matplotlib.pyplot as plt
import seaborn as sns
# Average power by hour of the day
hourly_avg = data.groupby('hour')['active_power'].mean()
plt.figure(figsize=(10, 6))
# Change color of the line to 'crimson' for a new look
sns.lineplot(x=hourly_avg.index, y=hourly_avg.values,
color='crimson', linewidth=2)
plt.title('Average Energy Consumption by Hour')
plt.xlabel('Hour of the Day')
plt.ylabel('Average Active Power')
plt.show()
import matplotlib.pyplot as plt
import seaborn as sns
# Average power by day of the week
weekly_avg =
data.groupby('day_of_week')['active_power'].mean()
plt.figure(figsize=(8, 5))
# Use a color palette for distinct bar colors
sns.barplot(x=weekly_avg.index, y=weekly_avg.values,
color='orange')
plt.title('Average Energy Consumption by Day of the Week')
plt.xlabel('Day of the Week')
plt.ylabel('Average Active Power')
plt.show()
data['is_weekend'] = data['day_of_week'].apply(lambda x: 1
if x >= 5 else 0)
data['month'] = data['date'].dt.month
# Average power by month
monthly_avg = data.groupby('month')['active_power'].mean()
plt.figure(figsize=(10, 6))
sns.lineplot(x=monthly_avg.index, y=monthly_avg.values,
color='red')
plt.title('Average Energy Consumption by Month')
```

```
plt.xlabel('Month')
plt.ylabel('Average Active Power')
plt.show()
data['active_power_lag1'] = data['active_power'].shift(1)
data['active_power_lag24'] = data['active_power'].shift(24)
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(8, 5))
# Scatter plot with color based on hour of the day
scatter = plt.scatter(data['active_power'],
data['active_power_lag1'],
                      c=data['hour'], cmap='YlOrRd',
alpha=0.7)
plt.colorbar(scatter, label='Hour of the Day')  # Colorbar
legend for hour
plt.title('Energy Consumption vs. Previous Hour (Lagged)')
plt.xlabel('Current Hour Power Consumption')
plt.ylabel('Previous Hour Power Consumption')
plt.show()
data['active_power_rolling_mean_24'] =
data['active_power'].rolling(window=24).mean()
data['active_power_rolling_std_24'] =
data['active_power'].rolling(window=24).std()
plt.figure(figsize=(12, 6))
plt.plot(data['date'], data['active_power'], label='Actual
Consumption', alpha=0.5, color='red')
plt.plot(data['date'], data['active_power_rolling_mean_24'],
label='24-hour Rolling Mean', color='yellow')
plt.title('Energy Consumption with Rolling Mean')
plt.xlabel('Date')
plt.ylabel('Active Power')
plt.legend()
plt.show()
data['temp_humidity_interaction'] = data['temp'] *
data['humidity']
corr = data[['active_power', 'temp', 'humidity',
'temp_humidity_interaction']].corr()
plt.figure(figsize=(8, 6))
```

```
sns.heatmap(corr, annot=True, cmap='YlOrRd', vmin=-1,
vmax=1)
plt.title('Correlation Heatmap of Weather and Energy
Features')
plt.show()
data.bfill(inplace=True)
data.tail()
```

3. MODEL TRAINING AND PREDICTION MODULE

```
from sklearn.model_selection import train_test_split
X = data.drop(['active_power'], axis=1)  # All columns
except target
y = data['active_power']                 # Target column
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
data = data.drop(columns=['date', 'main', 'description'])  #
Drop unnecessary columns if they're not useful
X_train, X_test, y_train, y_test =
train_test_split(data.drop(columns=['active_power']),
data['active_power'], test_size=0.2, random_state=42)
from xgboost import XGBRegressor
model = XGBRegressor(objective='reg:squarederror',
n_estimators=100, learning_rate=0.1, max_depth=6,
random_state=42)
model.fit(X_train, y_train)
predictions = model.predict(X_test)
```

4. MODEL EVALUATION AND INTERPRETATION

```
from sklearn.metrics import mean_absolute_error,
mean_squared_error
import numpy as np
mae = mean_absolute_error(y_test, predictions)
mse = mean_squared_error(y_test, predictions)
```

```
rmse = np.sqrt(mse
print(f"Mean Absolute Error (MAE): {mae:.2f}")
print(f"Mean Squared Error (MSE): {mse:.2f}")
print(f"Root Mean Squared Error (RMSE): {rmse:.2f}")
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(8, 6))
sns.scatterplot(x=y_test, y=predictions, alpha=0.5,
color='green')
plt.plot([y_test.min(), y_test.max()], [y_test.min(),
y_test.max()], '--', color='red')
plt.xlabel('Actual Energy Consumption')
plt.ylabel('Predicted Energy Consumption')
plt.title('Predicted vs. Actual Energy Consumption')
plt.show()
residuals = y_test - predictions
plt.figure(figsize=(8, 5))
sns.histplot(residuals, kde=True, color="purple")
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.title('Distribution of Residuals')
plt.show()
model.fit(X_train, y_train)
importances = model.feature_importances_
feature_names = X_train.columns  # Get the actual feature
names from the training data
print("Adjusted length of feature_names:",
len(feature_names))
print("Length of importances:", len(importances))
if len(importances) == len(feature_names):
    plt.figure(figsize=(10, 6))
    sns.barplot(x=importances, y=feature_names,
palette="magma")
    plt.xlabel('Importance')
    plt.title('Feature Importance in Energy Consumption
Model')
    plt.show()
else:
    print("Still a mismatch. Further verification needed.")
```

5. GENERATING USER INSIGHTS

```
next_period_prediction = model.predict([X_test.iloc[-1]])  # 
Using the last test example
print(f"Forecast: Your expected energy consumption for the 
next period is approximately {next_period_prediction[0]:.2f} 
kWh.")
high_usage_threshold = y_train.mean() + 2 * y_train.std()  # 
Example: 2 standard deviations above mean
if next_period_prediction > high_usage_threshold:
    print("Smart Alert: High Energy Consumption Detected! 
Consider reducing usage of non-essential devices to save on 
energy costs.")
forecast_periods = 24  # Example for hourly data, customize 
as needed
future_predictions = 
model.predict(X_test.tail(forecast_periods))
plt.figure(figsize=(10, 6))
plt.plot(range(forecast_periods), future_predictions, 
label='Predicted Consumption', color="crimson")
plt.xlabel("Time Period (e.g., hours)")
plt.ylabel("Energy Consumption (kWh)")
plt.title("Forecasted Energy Consumption for Next Periods")
plt.legend()
plt.show()
```

A2. OUTPUT SCREENSHOTS

Data Processing and Initial Exploration:

The statistical summary of the dataset reveals key insights, such as the completeness of data entries and the range of values for each variable as shown in **Fig. A2.1**, (i.e.) mean, standard deviation and quartiles.

	active_power	current	voltage	reactive_power	apparent_power	power_factor	temp
count	605260.000000	605260.000000	605260.000000	605260.000000	605260.000000	605260.000000	605260.000000
mean	286.019377	2.587303	125.417300	132.538007	321.839271	0.854202	19.525256
std	189.545683	1.591395	4.390612	71.030254	191.743962	0.114629	6.607351
min	24.400000	0.300000	107.600000	4.730000	37.140000	0.201800	-5.560000
25%	159.800000	1.610000	122.600000	74.070000	203.580000	0.754600	14.810000
50%	250.300000	2.290000	124.400000	128.500000	286.440000	0.891700	19.490000
75%	365.400000	3.220000	129.500000	177.652500	401.470000	0.950000	24.250000
max	2900.000000	24.410000	135.500000	1293.580000	2931.640000	1.000000	39.370000

Fig. A2.1. Statistical Summaries of the data

The line plot as shown in **Fig. A2.2**, serves as an important step in understanding how active power consumption varies over time, potentially correlating with external factors such as weather conditions or occupancy levels, which could further enhance predictive modeling efforts in our project.

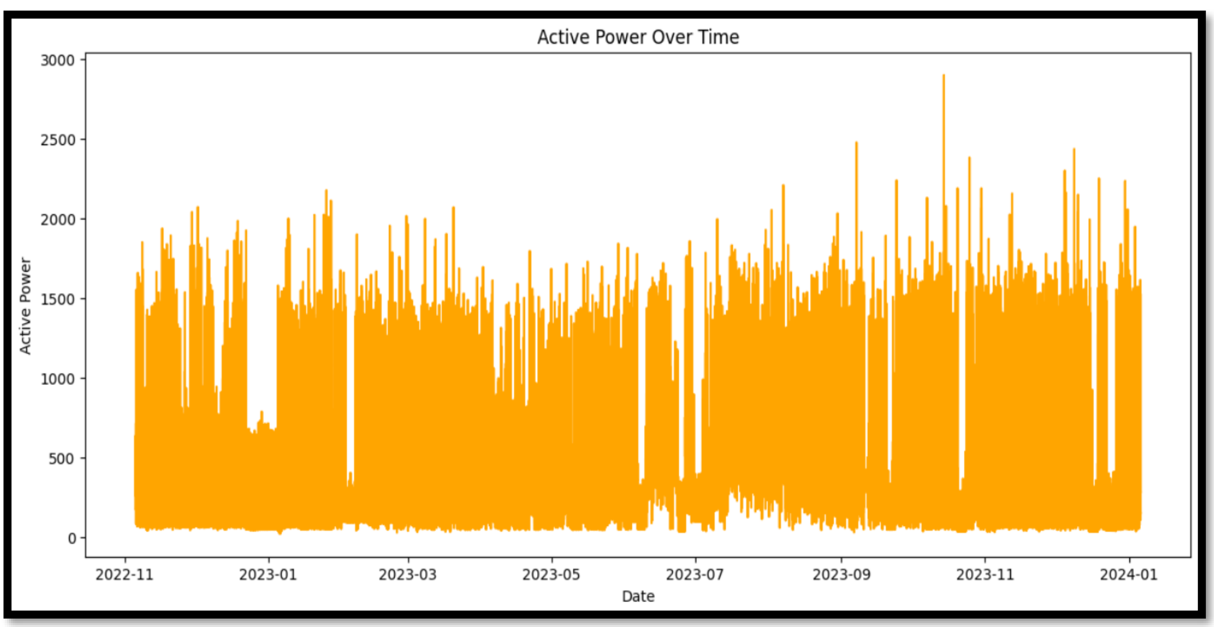

Fig. A2.2. Active Power over Time Series Plot

The histogram is an important exploratory data analysis step, as shown in **Fig. A2.3**, the underlying distribution of temperature values before incorporating them into our forecasting model.

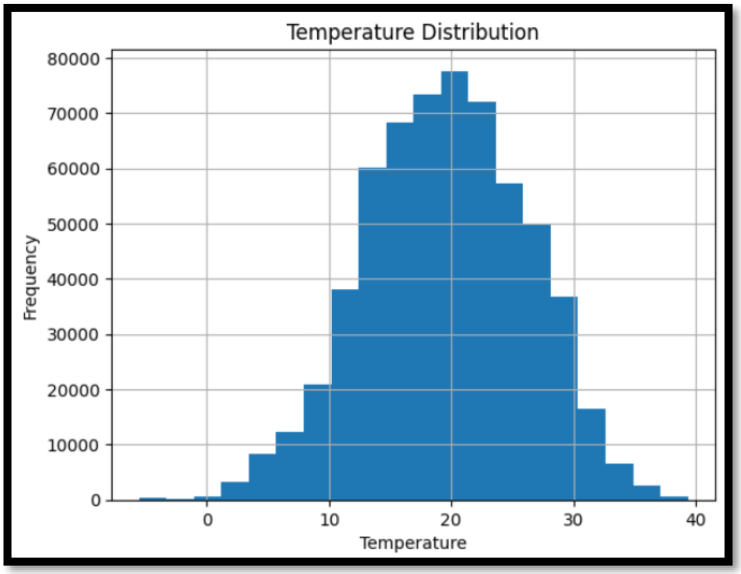

Fig. A2.3. Temperature Distribution – Histogram

As mentioned in **Fig. A2.4**, Exploring correlations identifies relationships between factors like weather and energy use, helping select key features and improving model accuracy by capturing dependencies and patterns in the data.

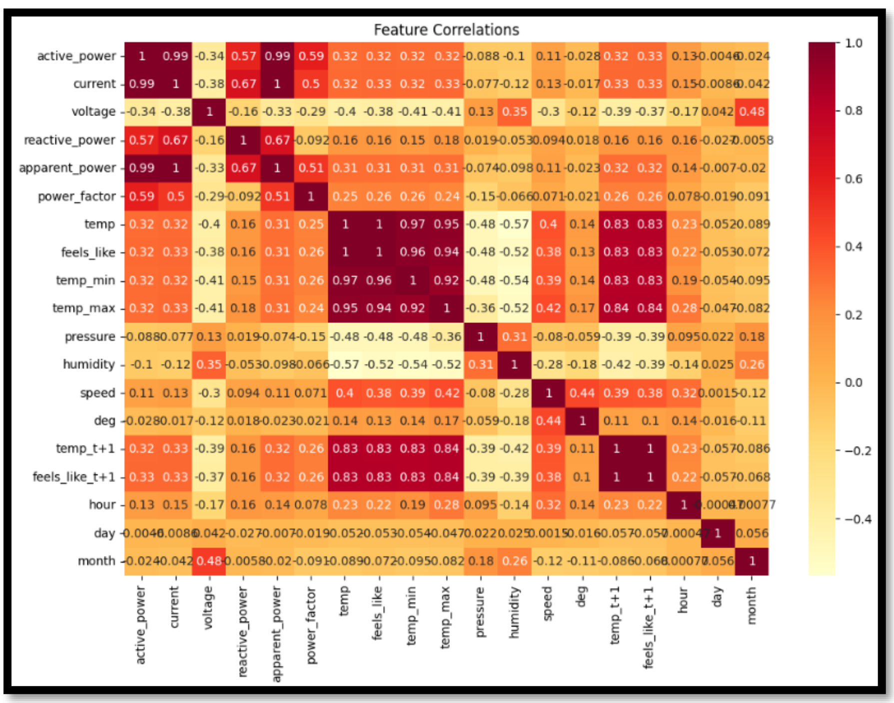

Fig. A2.4. Feature Correlation Heatmap

Feature Engineering:

The Line Plot as mentioned in the Fig. A2.5, captures daily cycles, as energy demand often varies at different times (e.g. peak usage in the evenings

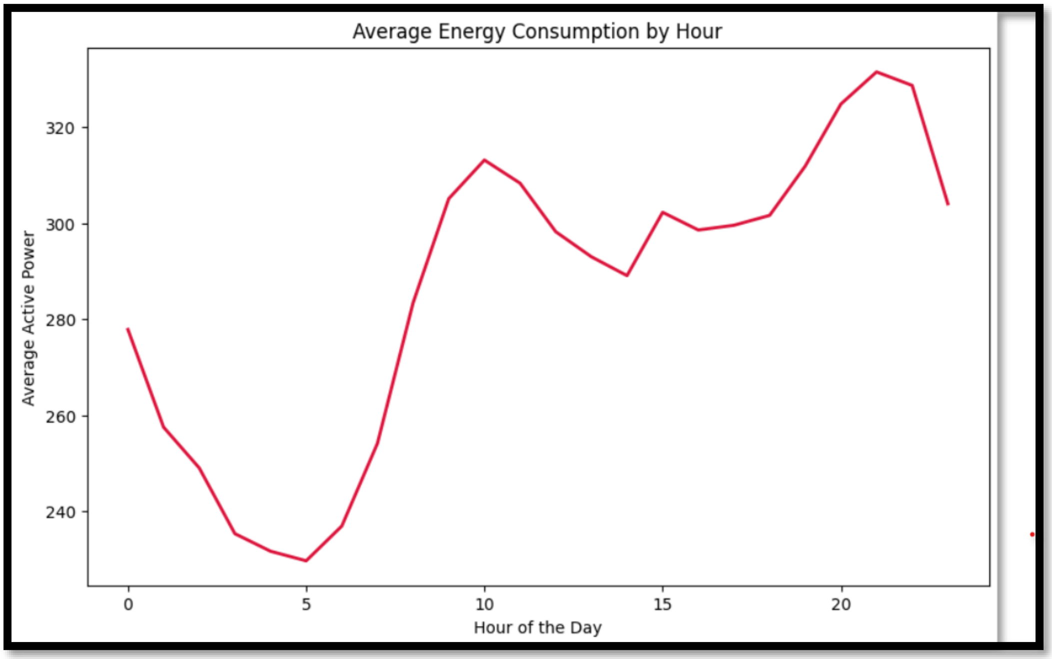

Fig. A2.5. Average Energy Consumption By Hour

The Line Plot as shown in the **Fig. A2.6**, shows variability in energy usage over the past 24 hours, which can help detect unusual spikes or drops.

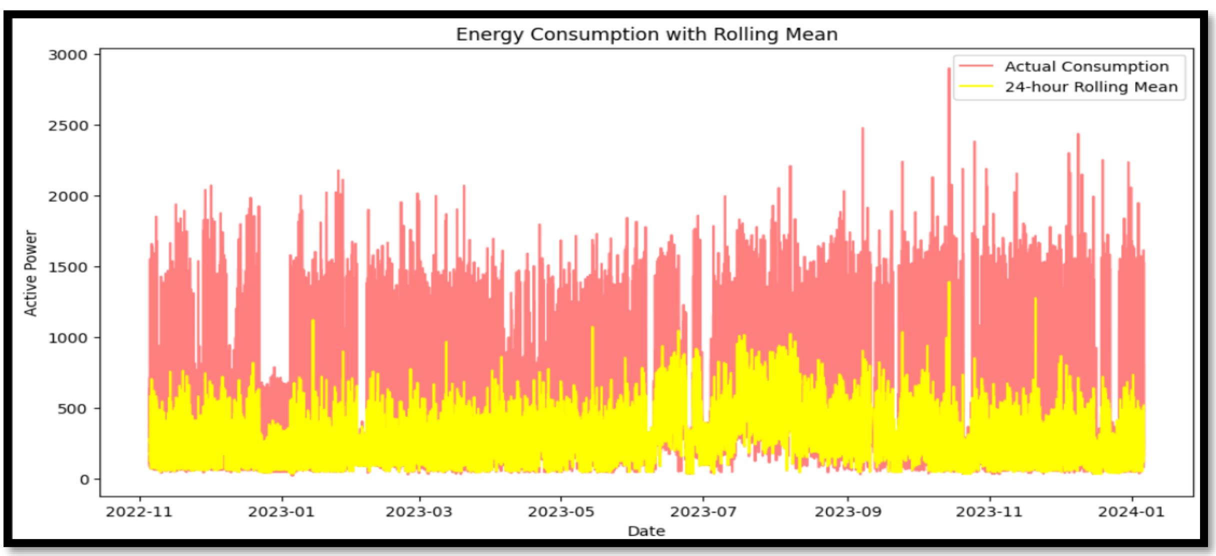

Fig. A2.6. Energy Consumption with Rolling Mean Time Series Plot

The Line Plot as shown in the **Fig. A2.7**, shows the energy usage from the same hour the previous day. This is helpful for capturing daily usage patterns.

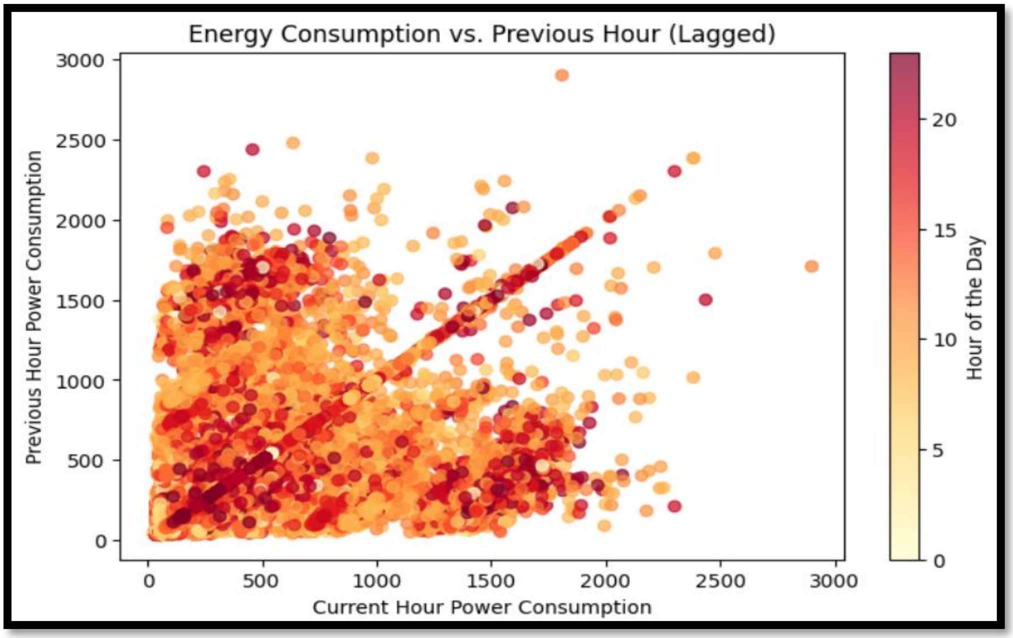

Fig. A2.7. Energy Consumption over the previous hour Scatter Plot

The correlation Heat Map shown in **Fig. A2.8**, shows the energy consumption often changes with weather. This feature multiplies temperature by humidity to show the combined impact**.**

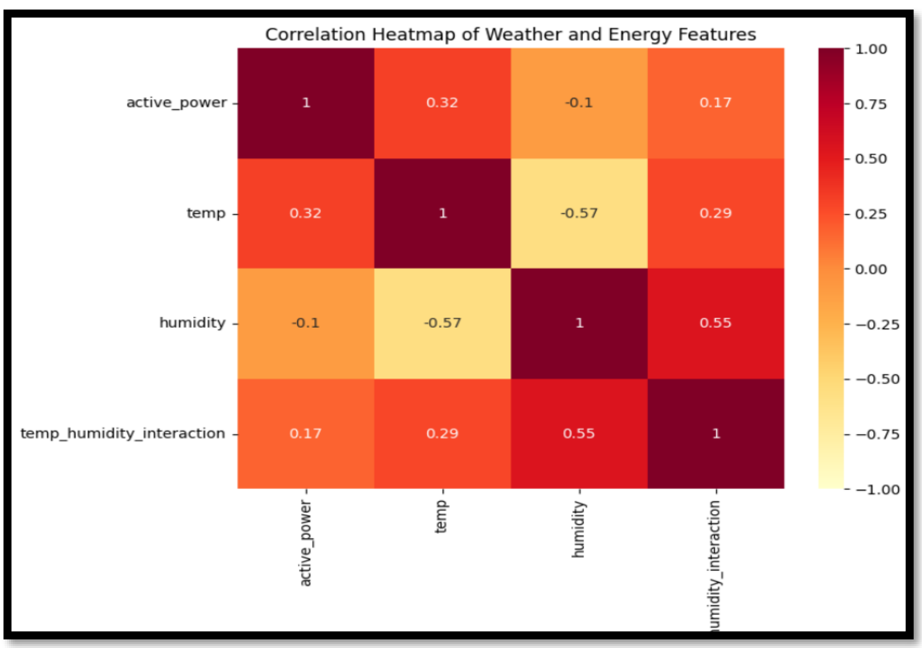

Fig. A2.8. Correlation Heatmap of Weather and Energy Features

A2.3. Model Training and Prediction:

The Fig. A2.9 shows the initialization parameters of the XGBRegressor class in the XGBoost library, used for training gradient-boosted regression models with customizable hyperparameters

```
XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=None, device=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric=None, feature_types=None,
             gamma=None, grow_policy=None, importance_type=None,
             interaction_constraints=None, learning_rate=0.1, max_bin=None,
             max_cat_threshold=None, max_cat_to_onehot=None,
             max_delta_step=None, max_depth=6, max_leaves=None,
             min_child_weight=None, missing=nan, monotone_constraints=None,
             multi_strategy=None, n_estimators=100, n_jobs=None,
             num_parallel_tree=None, random_state=42, ...)
```

Fig. A2.9. XGBRegressor Model Overview

A2.4. Model Evaluation and Interpretation:

A scatter plot of actual vs. predicted values as shown in Fig. A2.10, helps visualize how well the model captures the data's trends. The closer the points are to the line y = x, the better the model.

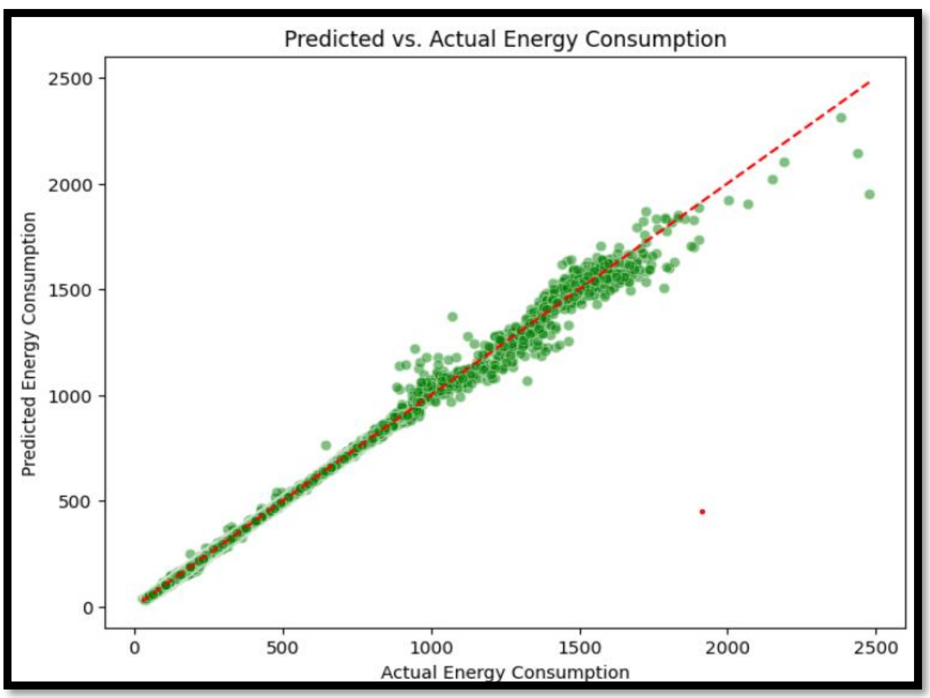

Fig. A2.10. Predicted vs. Actual Energy Consumption Regression Line

A histogram of residuals as shown in **Fig. A2.11,** can indicate whether there's any systematic error in the model's predictions.

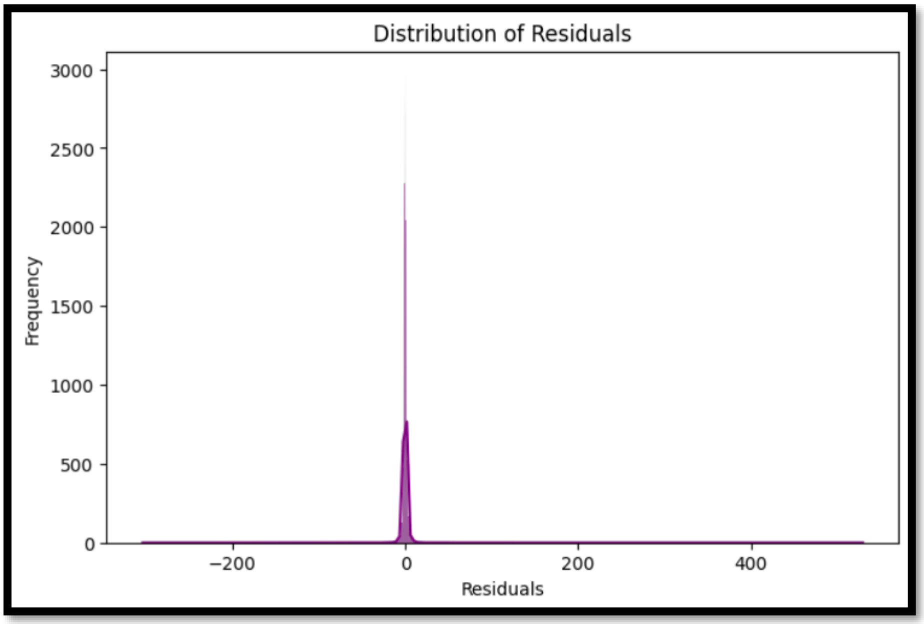

Fig. A2.11. Distribution of Residual – Histplot

A2.5. Generating User Insights:

A time-series chart of forecasted consumption for the next day/week as shown in the **Fig. A2.12**, helping users to visually grasp consumption trends.

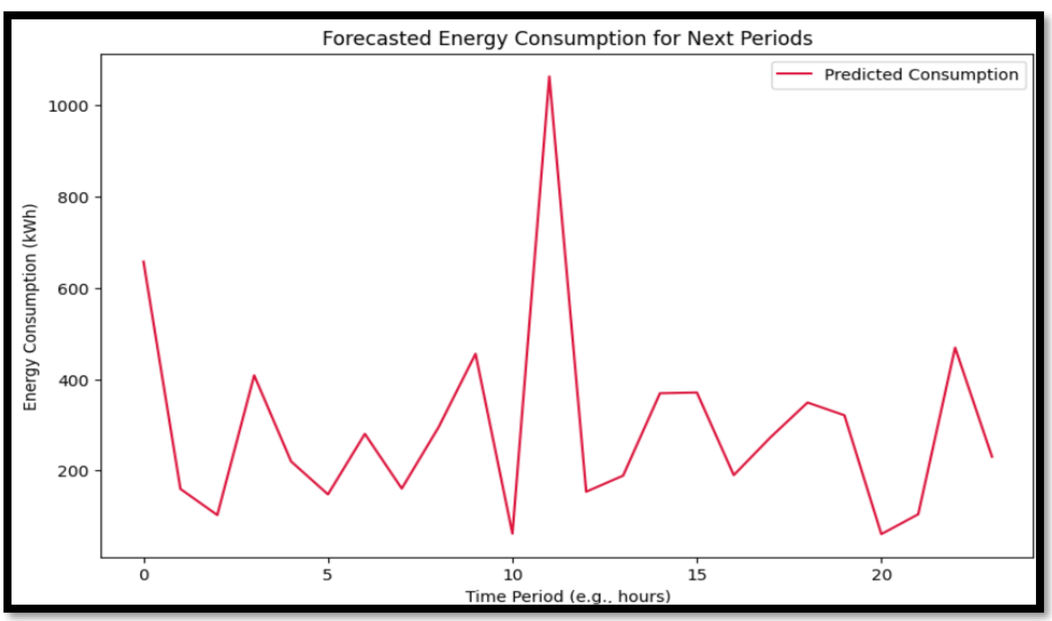

Fig. A2.12. Forecasted Energy Consumption for the next hour Line plot

REFERENCES

[1] L. H. M. Truong, K. H. K. Chow, R. Luevisadpaibul, G. S. Thirunavukkarasu, M. Seyedmahmoudian, B. Horan, S. Mekhilef, and A.Stojcevski, "Accurate Prediction of Hourly Energy Consumption in a Residential Building Based on the Occupancy Rate Using Machine Learning Approaches," *Applied Sciences*, vol. 11, no. 5, pp. 2229, Mar. 2021. doi: 10.3390/app11052229.

[2] Z. Wang, T. Hong, H. Li, et al., "Predicting city-scale daily electricity consumption using data-driven models," *Adaptive Energy*, vol. 16, no.1, pp. 100025, May 2021. doi: 10.1016/j.adapen.2021.100025.

[3] J. L. Harris and L.-M. Liu, "Dynamic structural analysis and forecasting of residential electricity consumption," *Economics and Forecasting*,P.O. Box 15.51 (Mail Code 5A5), Carolina Power and Light Company, Raleigh, NC 27602, USA, and Department of Information and Decision Sciences

[4] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Journal Name*, vol. XX, no.YY, pp. ZZ-ZZ, Month Year.

[5] Y.-S. Kim and J. Srebric, "Impact of occupancy rates on the building electricity consumption in commercial buildings," *Energy and Buildings*, vol. XX, pp. 7235, Dec. 2016. doi: 10.1016/j.enbuild.2016.12.056.

[6] D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," *Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia*, Email: marinodl@vcu.edu, amarasinghek@vcu.edu.

[7] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya; Electricity load forecasting: A systematic review; Journal of Electrical Systems and Information Technology, vol. 7, no. 1, pp. 1–19, 2020. DOI: 10.1186/s43067-020-00021-8.

[8] C. Sheng and H. Yu, “An optimized prediction algorithm based on XGBoost” in Proc. 2022 Int. Conf. Networking and Network Applications (NaNA), Xinjiang, China, Nov. 2022, pp. 442–447. [Online]. Available: IEEE Xplore. DOI: 10.1109/NaNA56854.2022.00082.