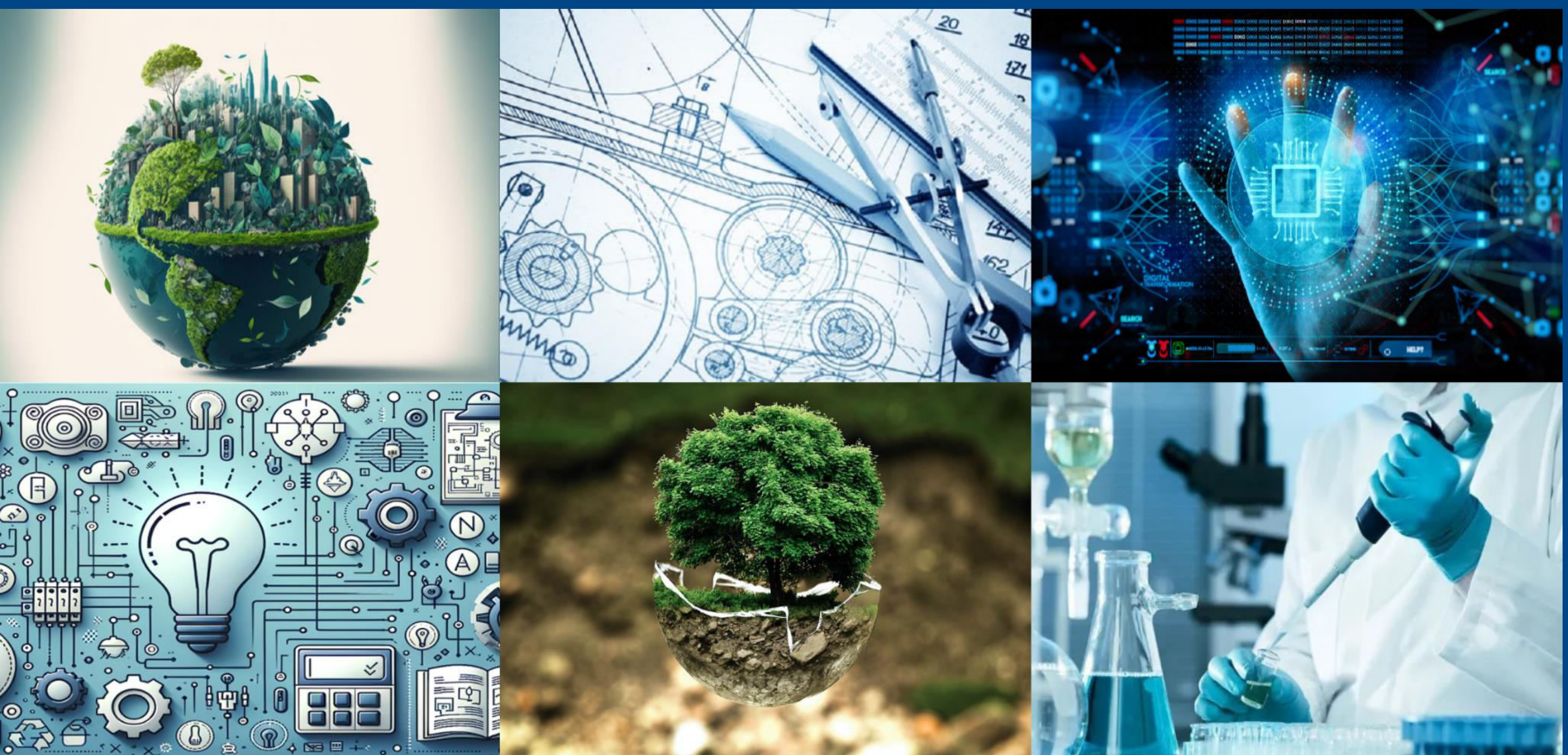
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804540

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Time Series Analysis by XGBoost Model for Future Prediction of Power Consumption

A.Srinivasa Rao

Assistant Professor, Department of AIML, Visvodaya Engineering College, Kavali, Nellore, AP., India

G. Srikanth, Sk.Ashrafsulthan, N. Rajendra, S. Srinivasulu, O. Ashok Kumar

Department of AIML, Visvodaya Engineering College, Kavali, Nellore, AP., India

ABSTRACT: This project presents an integrated system for forecasting energy consumption at both the appliance and group levels using a combination of time series analysis, machine learning, and deep learning techniques. The system is designed to address the growing need for accurate energy forecasting, which is essential for energy planning, cost control, and smart grid management. The project consists of three modules: (1) Appliance-Level Forecasting, using SARIMA and Prophet models to predict the consumption of individual household appliances based on the REFIT Electrical Load Measurements dataset, (2) Area-Level Forecasting and Cost Estimation, leveraging the XGBoost model on the EnergyData_Complete dataset to predict aggregated power consumption, generate synthetic data, and offer a real-time cost estimation tool, and (3) Model Comparison Dashboard, which compares the performance of ARIMA, LSTM, and XGBoost models for forecasting accuracy using metrics such as MAE, RMSE, and MAPE. The system's modular architecture ensures scalability, allowing it to be expanded for smart homes, urban energy monitoring, and energy-saving recommendation engines. By integrating both classical and machine learning models, this system offers a comprehensive solution for forecasting energy usage, reducing operational costs, and promoting sustainability. The results demonstrate that XGBoost outperforms traditional models, making it the preferred choice for energy consumption predictions.

I. INTRODUCTION

In today's rapidly evolving energy landscape, accurate forecasting of electricity consumption is not merely a technical challenge but a vital necessity. With the surge in urbanization, proliferation of smart appliances, and increasing dependence on electricity for everyday life, energy demand has become highly dynamic and unpredictable. At the same time, the integration of renewable energy sources and the advancement of smart grid technologies require a more proactive and intelligent approach to energy management. One of the most effective strategies to address these challenges is through time series-based energy forecasting, which enables utility providers, governments, businesses, and households to optimize energy usage, control costs, and plan future resource allocation efficiently.

Traditionally, energy forecasting has relied on statistical methods such as ARIMA and linear regression models. While these methods are valuable for understanding general trends and seasonal patterns, they often fall short when applied to real-world electricity usage, which is characterized by non-linear behaviors, irregular consumption, and high dependency on human activity and environmental factors. Particularly at the appliance level, energy consumption data can be highly spiky and influenced by individual user behavior, time of day, and lifestyle. These complexities demand more flexible and adaptive forecasting models that can handle both the volatility and multivariate nature of modern energy data.

This project proposes an integrated, multi-level forecasting system that combines classical statistical techniques with advanced machine learning and deep learning models to provide accurate predictions at both the micro (appliance) and macro (group or area-wide) levels. The system is modular and consists of three standalone yet interconnected components: (1) Appliance-Level Forecasting using SARIMA and Prophet models, (2) Area-Level Forecasting and Cost Estimation using XGBoost and synthetic data generation, and (3) a Model Comparison Dashboard evaluating ARIMA, LSTM, and XGBoost based on their performance. Each module is designed to solve a distinct real-world problem while contributing to the overall goal of creating a scalable and user-interactive forecasting platform.

The first module leverages the REFIT Electrical Load Measurements dataset to forecast the power consumption of specific household appliances. This allows individual users to monitor their electricity usage more effectively and make informed decisions to reduce their energy bills. By employing SARIMA and Prophet—both known for handling seasonality and trend components in time series data—this module captures the periodic nature of appliance usage while also addressing the irregularity and abrupt usage spikes commonly found in real-world datasets.

The second module expands the forecasting to an area-wide or aggregated scale using the EnergyData_Complete dataset. This component uses XGBoost, a robust and scalable machine learning algorithm that excels in modeling non-linear relationships and handling large datasets. In addition to forecasting future consumption trends, the module generates synthetic data for future intervals, allowing predictions beyond the historical timeframe. It also introduces a user-driven cost estimation tool, where users can input appliance details such as wattage, duration of use, and quantity to estimate energy costs in real-time. This interactive feature is especially beneficial for property managers, energy auditors, and end users looking to control or monitor power consumption dynamically.

To support data-driven decision-making and model selection, the third module provides a comparative evaluation of multiple forecasting techniques: ARIMA (statistical), LSTM (deep learning), and XGBoost (machine learning). The performance of each model is measured using key error metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These evaluations are visualized in a user-friendly dashboard built with Flask and Plotly, allowing stakeholders to interpret model outputs clearly and choose the best-performing model for specific forecasting needs. This module not only serves as a research tool for comparing model accuracy but also enables developers to understand the trade-offs between different modeling approaches.

The modularity and extensibility of the proposed system make it suitable for broader applications in smart homes, urban energy monitoring, and future development of energy-saving recommendation engines. By integrating both classical models and advanced machine learning techniques, the system strikes a balance between interpretability and predictive power. It also paves the way for incorporating external variables such as weather conditions, occupancy patterns, or time-of-use tariffs to further enhance forecasting accuracy.

This project aims to bridge the gap between theoretical energy forecasting models and practical, real-time applications. It addresses key challenges such as non-linearity, irregular data patterns, and user interactivity while offering a complete framework for energy consumption analysis and prediction. The integration of statistical, machine learning, and deep learning models into a single, user-centric platform provides a comprehensive solution that is both technically robust and practically useful. Through its three-module structure, the system demonstrates the potential of combining traditional and modern forecasting methods to meet the complex demands of today's energy systems.

II. LITERATURE SURVEY

Accurate energy consumption forecasting has been a critical area of research due to its direct impact on energy management, operational efficiency, and sustainable development. Over the past decades, a wide range of forecasting techniques have been proposed, ranging from classical statistical methods to advanced machine learning and deep learning models. This literature survey provides a comprehensive review of the key methodologies used in energy forecasting, along with their application scope, limitations, and potential integration into modern forecasting systems.

One of the earliest approaches to energy forecasting involves the use of Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant SARIMA. ARIMA models are particularly effective for univariate time series data with linear trends and stationary properties. SARIMA extends ARIMA by capturing seasonality, which is common in daily and monthly energy consumption patterns.

- Box & Jenkins (1970) introduced ARIMA as a foundational method for time series modeling.
- Chen et al. (2012) applied SARIMA to residential electricity data and demonstrated reasonable accuracy in capturing periodic usage patterns.

However, the performance of ARIMA/SARIMA deteriorates in the presence of non-linear relationships or sudden fluctuations, which are common in real-world appliance-level consumption. These models also assume stationary data and often struggle to adapt to real-time or dynamic environments.

The Prophet model, developed by Facebook (Taylor & Letham, 2017), is a decomposable time series model designed for scalability and interpretability. It separates time series into trend, seasonality, and holiday components, making it particularly effective for datasets with strong seasonal signals and missing data.

- Taylor and Letham (2017) demonstrated that Prophet performs comparably to ARIMA on many business forecasting tasks and is easier to tune.
- Its applicability in energy forecasting has been explored by Kandananond (2018) for medium-term electricity demand prediction, with satisfactory performance in capturing seasonality.

Prophet's strength lies in its flexibility and minimal pre-processing requirements, making it suitable for appliance-level data where human behavior introduces noise and irregularity.
In recent years, machine learning methods have been widely adopted for energy forecasting due to their ability to model complex, non-linear relationships.

- XGBoost (Extreme Gradient Boosting), proposed by Chen and Guestrin (2016), has gained popularity for its speed, accuracy, and regularization capabilities.
- Studies such as Deng et al. (2019) and Wang et al. (2020) have applied XGBoost to building-level and aggregated energy consumption datasets, outperforming linear regression and ARIMA in terms of RMSE and MAE.
- XGBoost is particularly effective when incorporating multiple features such as temperature, time of day, and historical lagged values, making it suitable for area-level forecasting.

Machine learning models, including XGBoost, are data-driven and do not require prior assumptions about data distribution, making them ideal for real-time and large-scale applications.
Deep learning approaches, particularly Long Short-Term Memory (LSTM) networks, have revolutionized sequence prediction tasks including energy consumption forecasting.

- LSTM, introduced by Hochreiter and Schmidhuber (1997), is capable of learning long-term dependencies and is resilient to the vanishing gradient problem.
- Marino et al. (2016) and Ryu et al. (2017) used LSTM networks to predict residential and commercial electricity usage with superior accuracy compared to traditional models.
- Deep learning models, however, require large datasets, high computational resources, and careful tuning of hyperparameters. They also tend to function as "black boxes," making interpretability a challenge.

Despite their complexity, LSTM models are highly effective for non-linear, multi-feature energy forecasting tasks where temporal dependencies are important.
Hybrid models and benchmarking studies have also gained traction in the literature to balance accuracy, interpretability, and complexity.

- Zhou et al. (2018) proposed a hybrid SARIMA-LSTM model that leveraged the strengths of both statistical and deep learning methods.
- Comparative studies like Amasyali and El-Gohary (2018) systematically evaluated multiple models across different datasets, concluding that no single model performs best in all scenarios.
- Visual dashboards and decision-support systems integrating multiple models are increasingly being used for transparent and comparative forecasting.

These approaches highlight the importance of modular and extensible forecasting systems, which is a core motivation of this project.

III. PROPOSED SYSTEM

The proposed system is a multi-level, modular energy consumption forecasting platform that integrates statistical, machine learning, and deep learning models to predict electricity usage at both the appliance level and the group (area) level. This system addresses the limitations of traditional forecasting methods by incorporating non-linear modeling capabilities, user interactivity, and real-time cost estimation into a unified framework. The design of the system ensures scalability, flexibility, and adaptability, making it suitable for various applications such as smart homes, commercial buildings, and energy auditing.

System Architecture Overview

The system is divided into three independent yet interconnected modules:

1. **Module 1 – Appliance-Level Forecasting using SARIMA and Prophet**
2. **Module 2 – Area-Level Forecasting and Cost Estimation using XGBoost**
3. **Module 3 – Forecasting Model Comparison Dashboard (ARIMA, LSTM, XGBoost)**

Each module is implemented using Python-based libraries and integrated through a Flask web interface, offering a seamless and interactive user experience.

Module 1: Appliance-Level Forecasting

This module focuses on forecasting the energy consumption of individual household appliances. It uses time series models—**SARIMA** and **Prophet**—which are capable of modeling seasonality, trends, and noise in the data. The module is trained on the **REFIT Electrical Load Measurements Dataset (House 1)**, which provides fine-grained appliance-level energy consumption data.

SARIMA is used to capture linear seasonal patterns and dependencies, while Prophet is employed for its robustness in handling missing data and its intuitive parameter tuning. The user can select a specific appliance (e.g., refrigerator, washing machine, or microwave), and the system provides forecasted power usage over a future time window. This helps individuals understand their consumption patterns and make informed decisions to reduce energy usage and optimize appliance operation times.

Module 2: Area-Level Forecasting and Cost Estimation

This module addresses the forecasting of **aggregated energy consumption** for a household, building, or area using the **EnergyData_Complete** dataset. The **XGBoost** machine learning algorithm is chosen for its superior performance in capturing non-linear relationships and its ability to handle multiple input features such as temperature, humidity, and time-based variables.

The module also includes a **synthetic data generation** feature that enables forecasting beyond the historical data window, helping in planning and long-term analysis. A key interactive component of this module is the **real-time cost estimation tool**, where users can input:

- Appliance wattage
- Number of appliances
- Daily usage hours
- Electricity price per unit

Based on these inputs, the system calculates estimated daily, monthly, and annual energy costs, providing valuable insights for users seeking to monitor and reduce their energy expenditures.

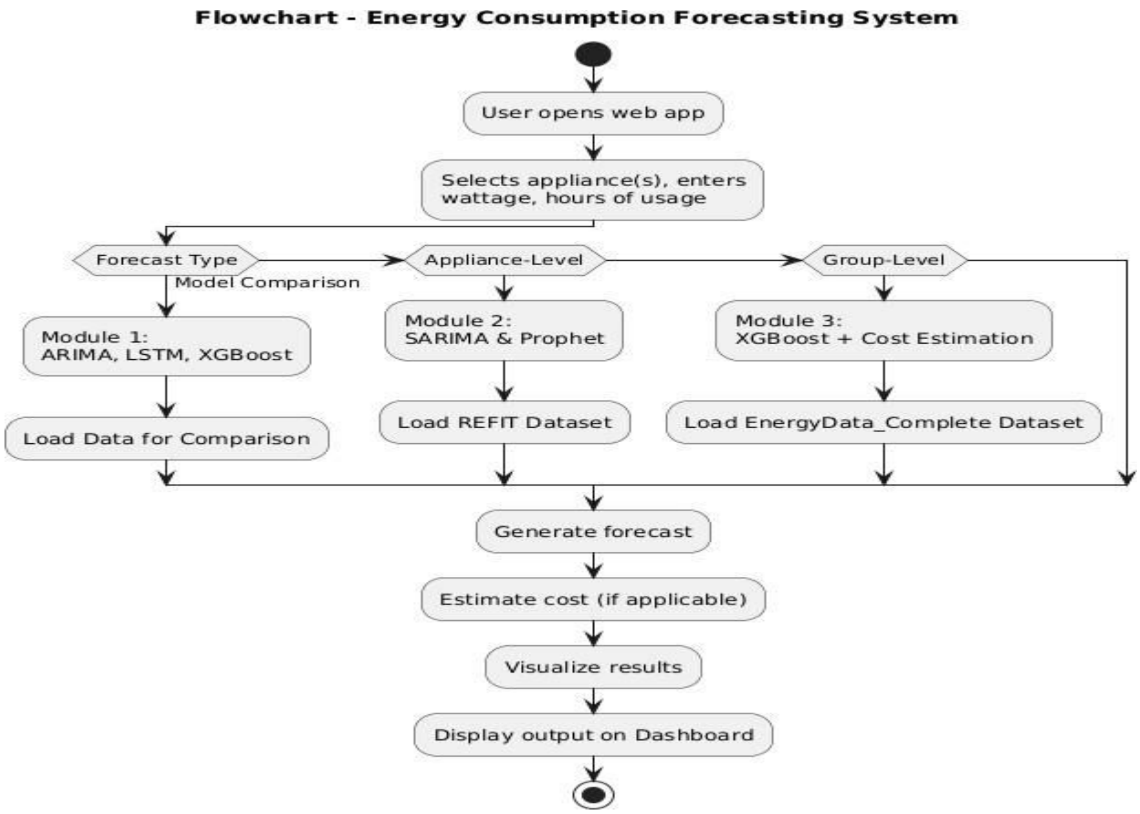
Module 3: Forecasting Model Comparison Dashboard

The third module provides a **benchmarking and visualization dashboard** to compare the performance of various forecasting models including **ARIMA**, **LSTM**, and **XGBoost**. These models are evaluated using common metrics:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Squared Error)**
- **MAPE (Mean Absolute Percentage Error)**

A user-friendly **Flask-based web dashboard** is developed using Plotly and Dash to present the forecasted outputs and model errors visually. Users can upload datasets, choose forecasting intervals, and interactively switch between models to compare their outputs. This helps researchers, developers, and analysts understand which model performs best under specific data conditions and use cases.

Fig 1: Architecture Diagram



IV. WORKING PROCEDURE

The proposed energy forecasting system is composed of three core modules, each handling a specific level of prediction and user interaction. The system is designed to operate as a web-based application using Flask, allowing users to input parameters, run forecasts, and visualize results in real-time. Below is a step-by-step outline of the working procedure:

Step 1: Data Acquisition and Preprocessing

1. Data Sources:
   - REFIT Dataset (House 1) – Appliance-level energy consumption data.
   - EnergyData_Complete Dataset – Group/area-level aggregated energy consumption along with environmental features like temperature and humidity.
2. Preprocessing Tasks:
   - Handling missing or null values.
   - Resampling and normalizing time series data (e.g., from seconds to hourly/daily formats).
   - Feature engineering (e.g., time of day, day of week, lag features).
   - Splitting data into training and testing sets.

Step 2: Appliance-Level Forecasting (Module 1)

1. Model Selection:
   - Users choose between SARIMA and Prophet models.

2. Input Selection:
    - Users select a specific appliance from the REFIT dataset (e.g., fridge, kettle, microwave).
3. Model Training and Forecasting:
    - Selected model is trained on historical data of the chosen appliance.
    - The model forecasts energy consumption for the next user-specified time window (e.g., next 24 hours or 7 days).
4. Output Visualization:
    - Forecast results are plotted using interactive charts (e.g., line graphs).
    - Predictions are compared with actual values (if available) to assess accuracy.

Step 3: Area-Level Forecasting and Cost Estimation (Module 2)

1. Model Execution (XGBoost):
    - The EnergyData_Complete dataset is used to train the XGBoost regression model.
    - Features include environmental data (temperature, humidity), timestamp-related features (hour, day), and previous consumption values.
2. Forecasting:
    - The trained model predicts future energy usage for a building or area.
3. Synthetic Data Generation:
    - If forecast intervals exceed the available data, synthetic data is generated using trends and seasonality patterns.
4. Cost Estimation Tool:
    - Users input:
        - Appliance wattage (e.g., 1500W)
        - Usage duration (e.g., 3 hours/day)
        - Number of appliances (e.g., 2 units)
        - Cost per unit (e.g., ₹7/kWh)
    - The system calculates:
        - Daily energy usage
        - Estimated daily, monthly, and yearly electricity cost
5. Visualization:
    - Predicted usage trends and cost breakdown are shown using graphs and tables.

Step 4: Forecasting Model Comparison (Module 3)

1. Dataset Upload/Selection:
    - Users can upload a dataset or choose a preloaded one.
2. Model Execution:
    - The system runs three models in parallel:
        - ARIMA – For statistical baseline comparison.
        - LSTM – For deep learning-based sequential prediction.
        - XGBoost – For non-linear regression-based forecasting.
3. Performance Evaluation:
    - Predictions are compared against actual values using:
        - MAE (Mean Absolute Error)
        - RMSE (Root Mean Squared Error)
        - MAPE (Mean Absolute Percentage Error)
4. Visualization Dashboard:
    - Forecast results and error metrics are visualized side-by-side using interactive charts.
    - This helps users understand which model performs best for their specific use case.

Step 5: User Interface and Deployment

1. Web Interface (Flask):
    - A user-friendly interface allows users to:
        - Choose forecasting models

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804540

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

  - Enter custom inputs
  - View plots and metrics
  - Download results

2. Technologies Used:
   - Flask – Backend API and routing
   - Plotly/Dash – Interactive chart rendering
   - Pandas/Numpy – Data handling
   - XGBoost, SARIMA (statsmodels), Prophet, LSTM (TensorFlow/Keras) – Model building
   - HTML/CSS/Bootstrap – Frontend styling
3. Deployment Options:
   - Can be deployed locally or on cloud platforms (Heroku, AWS, or PythonAnywhere) for public access.
   -

Step 6: Output and Insights

- Appliance Usage Forecasts – Helps households manage specific appliances.
- Group-Level Forecasts and Cost – Assists energy managers in planning and budgeting.
- Model Benchmarking – Offers insights into model accuracy and applicability for future development.

V. RESULT

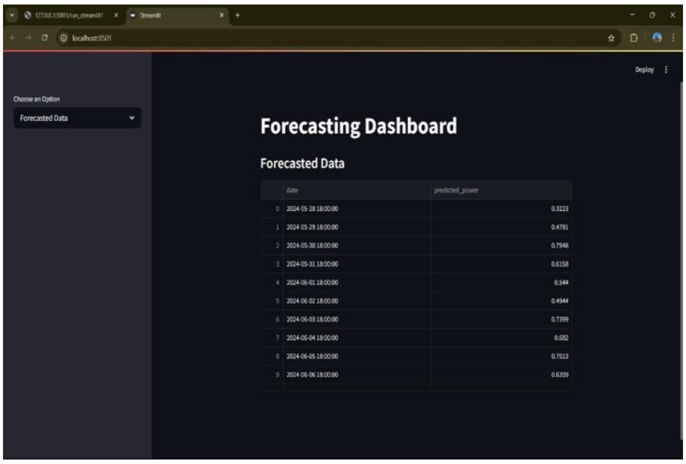


Fig 2: Home Page

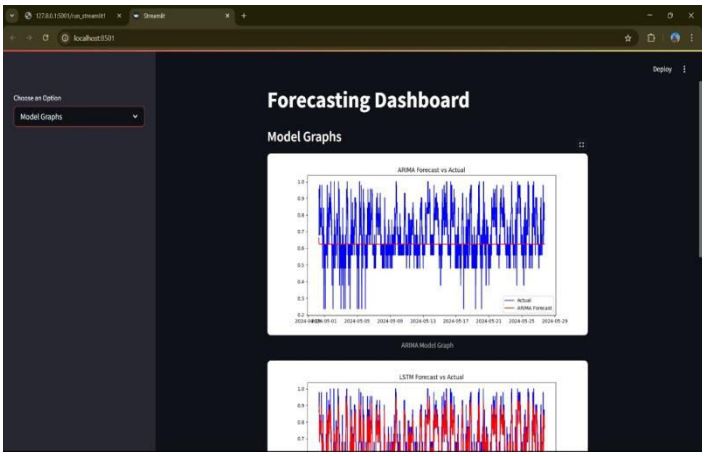


Fig 3: Forecasting

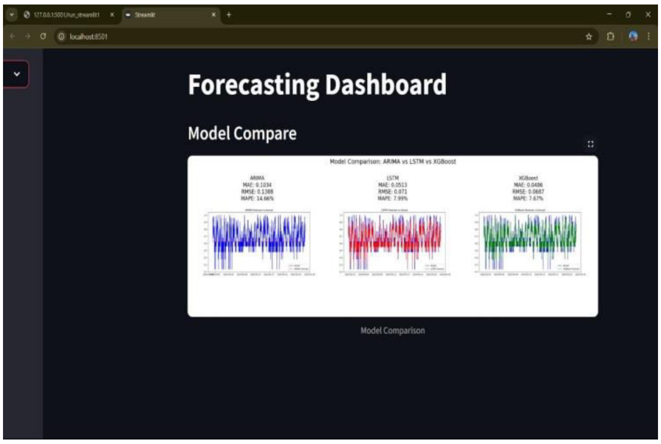


Fig 4: Forecasting

VI. CONCLUSION

This project presents a comprehensive, modular, and intelligent energy forecasting system that integrates statistical, machine learning, and deep learning models to predict electricity consumption at both the appliance and area levels. By leveraging models such as SARIMA, Prophet, XGBoost, ARIMA, and LSTM, the system effectively addresses the challenges of seasonality, non-linearity, irregular appliance behavior, and external influencing factors such as temperature and humidity.

The three-module structure enables flexible and scalable deployment:
Module 1 offers appliance-level forecasts using SARIMA and Prophet, aiding individual users in identifying consumption patterns and managing specific devices.

Module 2 performs area-level forecasting using XGBoost and provides a real-time cost estimation tool, empowering building managers and energy auditors with actionable insights.

Module 3 compares forecasting models in terms of accuracy and performance, helping developers and researchers select the most effective model for a given scenario.

The project also introduces a user-interactive Flask web interface that makes advanced forecasting tools accessible to both technical and non-technical users. This real-time interaction and interpretability bridge the gap between complex models and practical applications.

Overall, the proposed system demonstrates that combining classical time series models with modern machine learning techniques can significantly improve the accuracy and usability of energy forecasting solutions. It lays a strong foundation for future extensions such as real-time IoT integration, adaptive energy-saving recommendations, and deployment in smart home or smart grid environments.

REFERENCES

[1]Semmelmann L., Henni S., Weinhardt C., "Load Forecasting for Energy Communities: A Novel LSTM-XGBoost Hybrid Model," Energy Informatics, vol. 5, no. 1, 2022. doi: 10.1186/s42162-022-00212-9
[2]Li Y., Peng Y., Zhang D., Mai Y., Ruan Z., "XGBoost Energy Consumption Prediction Based on Multi-System Data HVAC," arXiv preprint, 2021. arXiv:2105.09945.
[3]Hewamalage H., Bergmeir C., Bandara K., "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions," arXiv preprint, 2019. arXiv:1909.00590.
[4]YuanY.,ZhangY., Luo Z.,"LoadForecastingBasedon LSTMNeuralNetwork," Journal of Electrical Engineering & Technology, vol. 16, pp. 1621–1628, 2021. doi: 10.1007/s42835-021-00768-8

[5]LiaoW.,Bak-JensenB.,PillaiJ.R.,YangZ.,LiuK.,"Short-TermPowerPrediction Using Hybrid Graph Convolutional Network and LSTM," arXiv preprint, 2021. arXiv:2111.07958.
[6]Facebook Research, "Prophet: Forecasting at Scale," [Online]. Available: https://facebook.github.io/prophet/
[7]Liu N., Hu Y., Ai X., "Power Load Forecasting Based on Random Forest Regression,"IOPConf.Series:EarthandEnvironmentalScience,vol.252,no.3,2019. doi: 10.1088/1755-1315/252/3/032171
[8]MakridakisS.,SpiliotisE.,AssimakopoulosV.,"StatisticalandMachineLearning ForecastingMethods:ConcernsandWaysForward,"PLOSONE,vol.13,no.3,2018. doi: 10.1371/journal.pone.0194889
[9]Lago J., Marcjasz G., De Schutter B., Weron R., "Forecasting Day-Ahead Electricity Prices: A Review," arXiv preprint, 2020. arXiv:2008.08004.
[10]Nabipour M., Nayyeri P., Jabani H., Shamshirband S., Mosavi A., "Predicting StockMarketTrendsUsingMachine LearningandDeep LearningAlgorithms,"IEEE Access, vol. 8, pp. 150199–150212, 2020. doi: 10.1109/ACCESS.2020.3015966

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY