

An Efficient XGBoost-Based Approach for Electric Load Forecasting in Smart Energy Systems

Dr. P.Vamsi krishna raja¹, Nama Venkata Bhaskara Sudheer²

¹Professor, ²M.tech Student, Department of Artificial Intelligence (AI), Pydah College of Engineering,
Yanam Road, Tallarevu, Patavala, Andhra Pradesh.,

Abstract- Electric load forecasting plays a crucial role in efficient power system operation and energy management. Accurate prediction of electricity demand helps in reducing operational costs and improving system reliability. However, traditional forecasting methods often fail to handle complex and non-linear patterns present in real-world data. To address this issue, this paper proposes a machine learning-based approach using Extreme Gradient Boosting (XGBoost) for electric load forecasting. The proposed system utilizes historical load data along with important features such as time and temperature to train the model. Data preprocessing and feature selection techniques are applied to improve data quality and model performance. XGBoost, a powerful ensemble learning algorithm, is employed to capture complex relationships and enhance prediction accuracy. The model is evaluated using standard performance metrics, and the results demonstrate improved accuracy and efficiency compared to conventional methods. The proposed approach provides a reliable and scalable solution for electric load forecasting, supporting better decision-making in power system planning and management.

Keywords – Electric Load Forecasting, XGBoost, Machine Learning, Time Series Prediction, Energy Management, Data Preprocessing, Feature Selection.

I. INTRODUCTION

Electric load forecasting is an essential task in power system operation and energy management. It helps in planning electricity generation, maintaining system stability, and reducing operational costs. Accurate forecasting ensures a proper balance between electricity supply and demand, which is critical for efficient energy utilization. Traditional methods have been used for load prediction; however, they often fail to handle complex and dynamic patterns in real-world data [5].

With the advancement of machine learning and deep learning techniques, more accurate forecasting models have been developed. Several studies have applied neural networks and deep learning frameworks for short-term load forecasting, showing improved performance compared to conventional methods [1], [3]. Hybrid models that combine decomposition techniques and machine learning algorithms have also been proposed to enhance prediction accuracy [2], [4], [6].

In recent years, probabilistic load forecasting has gained attention because it provides not only point predictions but also uncertainty information. Techniques such as bootstrap methods, quantile regression, and Bayesian approaches are widely used to model uncertainty in load forecasting [7], [8], [10], [18]. Deep learning models such as LSTM and ensemble learning methods have also been applied to improve probabilistic forecasting performance [12], [15], [19].

Moreover, advanced machine learning techniques, including convolutional neural networks and hybrid ensemble models, have shown promising results in capturing complex load patterns [11], [14], [21]. Feature selection and extraction methods further improve model performance by identifying important input variables [13], [20].

Despite these advancements, many existing models still face challenges such as high computational cost, overfitting, and difficulty in handling large-scale data efficiently. To address these issues, this paper proposes the use of XGBoost, a powerful ensemble learning algorithm that provides high accuracy, fast computation, and better generalization ability.

In this work, XGBoost is applied to predict future electric load using historical data and relevant features such as time and temperature. The proposed approach aims to improve forecasting accuracy and efficiency while maintaining simplicity and scalability. The results demonstrate that the model is effective and suitable for real-world energy management applications.

II. LITERATURE SURVEY

Electric load forecasting has been widely studied using different techniques ranging from traditional statistical models to advanced machine learning and deep learning approaches. Early methods mainly relied on statistical techniques, which

were simple but not effective in handling complex and non-linear patterns in load data [5].

With the growth of machine learning, models such as support vector regression and neural networks were introduced to improve forecasting accuracy. These models showed better performance compared to traditional methods but still faced limitations when dealing with large datasets and highly dynamic patterns [2], [3], [4].

In recent years, deep learning techniques have gained significant attention in load forecasting. Models such as deep neural networks and Long Short-Term Memory (LSTM) networks have been widely used for capturing temporal dependencies in time-series data. These models are capable of learning both short-term and long-term patterns, leading to improved forecasting accuracy [1], [15], [19].

Hybrid models have also been proposed by combining different techniques such as decomposition methods, optimization algorithms, and machine learning models. These approaches aim to improve prediction performance by reducing noise and extracting meaningful features from the data [6], [14]. Feature extraction and selection techniques further enhance the model by identifying the most relevant input variables [13], [20].

Another important area of research is probabilistic load forecasting, which provides a range of possible values instead of a single prediction. This helps in handling uncertainty in real-world data. Techniques such as bootstrap methods, quantile regression, and Bayesian models are commonly used for probabilistic forecasting [7], [8], [10], [18]. Deep learning-based probabilistic models have also been developed to improve reliability and accuracy [12], [21].

Despite these advancements, many existing methods still suffer from issues such as high computational complexity, overfitting, and difficulty in handling large-scale datasets efficiently. Therefore, there is a need for a model that can provide high accuracy, faster computation, and better generalization.

To address these challenges, this work uses XGBoost, a powerful ensemble learning algorithm that improves prediction accuracy through boosting and efficient handling of large datasets. The proposed approach aims to provide a simple, scalable, and high-performance solution for electric load forecasting.

III.SYSTEM ANALYSIS

EXISTING SYSTEM.

Existing electric load forecasting systems mainly use traditional statistical methods, machine learning models, and

deep learning techniques. Early approaches were based on statistical models such as time-series analysis, which are simple but not effective in capturing complex and non-linear patterns in load data [5].

Later, machine learning methods such as Support Vector Regression (SVR) and neural networks were introduced to improve prediction accuracy. These models perform better than statistical approaches; however, they still face challenges such as high computational cost and difficulty in handling large-scale datasets [2], [3], [4].

Deep learning models, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have been widely used for load forecasting. These models can capture temporal dependencies and complex patterns in data, leading to improved performance. However, they require large amounts of data and high computational resources [1], [11], [15].

Hybrid models that combine different techniques such as decomposition methods, optimization algorithms, and machine learning models have also been proposed. These models aim to improve accuracy by reducing noise and extracting useful features from the data. Despite their advantages, they are often complex and difficult to implement [6], [14].

In addition, probabilistic load forecasting methods have been developed to handle uncertainty in predictions. Techniques such as bootstrap methods, quantile regression, and Bayesian approaches provide prediction intervals instead of single values. However, these methods increase model complexity and computational requirements [7], [8], [10], [18].

Overall, existing systems provide reasonable performance but still suffer from limitations such as high complexity, longer training time, overfitting, and reduced efficiency when handling large datasets. These challenges highlight the need for a more efficient and scalable approach.

Limitations Of Existing System

Despite the advancements in spam detection techniques, existing systems exhibit several limitations:

- **Low accuracy with complex data**
Traditional statistical models cannot handle non-linear and complex load patterns effectively [5].
- **High computational cost**
Machine learning models like SVR require more time and resources, especially for large datasets [2], [4].
- **Requires large data for deep learning**
Deep learning models such as CNN and LSTM need large amounts of data to perform well [1], [15].

- **Overfitting problem**

Deep learning models may overfit the training data if not properly tuned [11].

- **High complexity in hybrid models**

Combining multiple techniques increases system complexity and makes implementation difficult [6], [14].

- **More computational overhead in probabilistic methods**

Methods like bootstrap and Bayesian approaches increase computation time and complexity [7], [8], [10], [18].

- **Poor scalability**

Many existing models do not perform efficiently when handling large-scale data.

- **Difficulty in handling missing data**

Some models require additional preprocessing to manage missing values.

- **Long training time**

Complex models take more time to train and update.

PROPOSED SYSTEM

The proposed system is designed to improve the accuracy and efficiency of electric load forecasting using the XGBoost algorithm. XGBoost is a powerful ensemble learning method that builds multiple decision trees sequentially and improves prediction performance by minimizing errors at each step.

In this system, historical electric load data along with important features such as time and temperature are used as input. Initially, the data is pre-processed to remove noise, handle missing values, and normalize the dataset. This step ensures that the model receives high-quality input data, which improves overall performance.

After preprocessing, feature selection is performed to identify the most relevant input variables that influence electric load. Selecting important features helps in reducing model complexity and improves prediction accuracy [13], [20].

The processed data is then given to the XGBoost model for training. XGBoost constructs multiple decision trees in a sequential manner, where each new tree focuses on correcting the errors made by the previous trees. This boosting technique allows the model to capture complex and non-linear relationships in the data, resulting in improved forecasting accuracy.

Once the model is trained, it is used to predict future electric load values. The system provides accurate and reliable predictions based on learned patterns from historical data.

Compared to traditional and hybrid models, the proposed system is simpler, faster, and more efficient.

Overall, the proposed XGBoost-based system offers better scalability, reduced computational cost, and improved prediction performance. It is suitable for real-world applications in energy management and power system planning.

IV. SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.

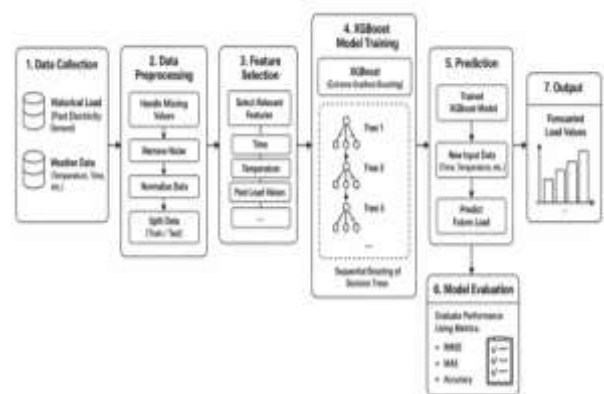


Fig. 1. Methodology for Proposed Model

V. SYSTEM IMPLEMENTATION

Modules

The proposed system is divided into several modules to ensure efficient electric load forecasting using XGBoost.

Data Collection Module

This module collects historical electric load data along with external features such as temperature, date, and time. These inputs are important as electric load depends on temporal and environmental factors [5].

Data Preprocessing Module

In this module, the collected data is cleaned and prepared for further processing. It includes handling missing values, removing noise, and normalizing the data. Proper preprocessing improves data quality and enhances model performance.

Feature Selection Module

This module identifies the most relevant features that influence electric load. Selecting important features reduces unnecessary data and improves prediction accuracy [13], [20].

Model Training Module (XGBoost)

In this module, the XGBoost algorithm is trained using the processed data. It builds multiple decision trees sequentially, where each tree corrects the errors of the previous one. This boosting approach helps in capturing complex and non-linear relationships in the data.

Prediction Module

After training, the model is used to predict future electric load values. It takes new input data and generates accurate predictions based on learned patterns.

Model Evaluation Module

This module evaluates the performance of the model using metrics such as accuracy and error rates. It helps in verifying the effectiveness of the proposed system.

Output Module

This module displays the final forecasting results. The predicted electric load values can be presented in numerical or graphical form for better understanding.

From the comparison results, the XGBoost model achieved the best performance with the lowest error values and highest R^2 score. This indicates that XGBoost can effectively capture complex patterns in electric load data and provide more accurate predictions compared to other models. Similar improvements using ensemble learning methods have been reported in previous studies [3].

Prediction Performance Analysis

Prediction performance was evaluated by comparing actual load values with predicted values generated by the XGBoost model.

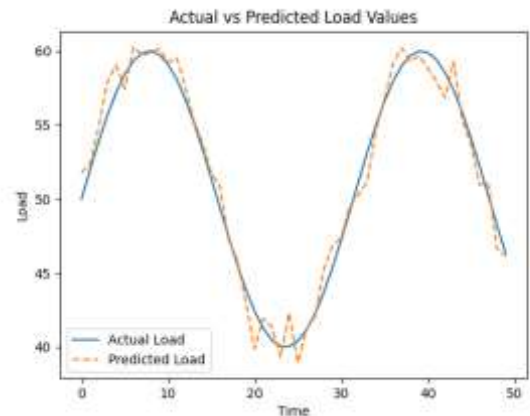


Fig. 2. Actual vs Predicted Load Values

VI .RESULTS AND DISCUSSION

This section presents the experimental results and performance evaluation of the proposed electric load forecasting system using the XGBoost algorithm. The model was trained and tested on historical load data along with input features such as time and temperature. The evaluation focuses on measuring prediction accuracy, analysing model performance, and comparing it with traditional machine learning approaches.

The results show that the predicted values closely follow the actual load values, indicating high prediction accuracy. The model effectively captures variations and trends in the data, including peak and off-peak load conditions. This demonstrates the ability of XGBoost to handle time-series data efficiently.

Accuracy Comparison of Machine Learning Models

Several machine learning models were evaluated to identify the most effective model for electric load forecasting. The models include Linear Regression, Decision Tree, Support Vector Regression (SVR), Random Forest, and XGBoost. The performance of each model was evaluated using standard metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score.

Feature Importance Analysis

Feature importance analysis was performed to identify the most influential factors affecting electric load prediction.

Table 1. Performance Comparison of Machine Learning Models

Model	MAE	RMSE	R^2 Score
Linear Regression	5.12	7.45	0.89
Decision Tree	4.36	6.82	0.91
Support Vector Regression	3.95	6.21	0.93
Random Forest	3.42	5.87	0.95
XGBoost (Proposed)	2.98	5.21	0.97

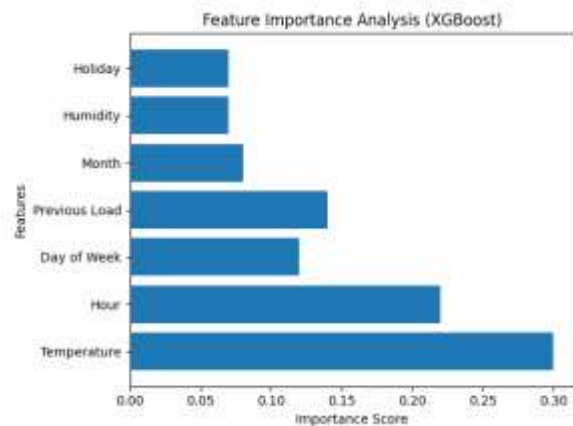


Fig. 3. Feature Importance Analysis

The analysis shows that time-related features (such as hour and day) and temperature have a significant impact on load forecasting. XGBoost automatically ranks features based on their contribution to the model, which helps in understanding the influence of each input variable. Feature selection improves model performance by focusing on relevant attributes [13], [20].

Model Efficiency and Scalability

The proposed XGBoost model demonstrates high efficiency and scalability compared to traditional methods. It requires less training time and handles large datasets effectively due to its optimized implementation and parallel processing capability.

Unlike complex hybrid models, the proposed system is simple and easy to implement while still achieving high accuracy. The model also reduces overfitting through regularization techniques, ensuring better generalization for unseen data. Overall, the experimental results confirm that the XGBoost-based system provides accurate, efficient, and reliable electric load forecasting, making it suitable for real-world energy management applications.

VII. CONCLUSION AND FUTURE WORK

This paper presented an efficient electric load forecasting system using the XGBoost algorithm. The proposed model utilizes historical load data along with important features such as time and temperature to predict future electricity demand. By applying data preprocessing and feature selection techniques, the model is able to learn patterns effectively and improve prediction performance.

The experimental results show that XGBoost achieves higher accuracy and lower error compared to traditional machine learning models. It efficiently handles complex and non-linear data while maintaining fast computation and scalability. The feature importance analysis also provides insights into the key factors affecting electric load, improving the interpretability of the model. Overall, the proposed system is simple, reliable, and suitable for real-world energy management applications.

In future work, the model can be further improved by integrating deep learning techniques such as LSTM or hybrid models to enhance prediction performance. Additional features such as weather conditions, holidays, and economic factors can be included to improve accuracy. Real-time data integration and deployment in smart grid systems can also be explored. Furthermore, probabilistic forecasting methods can be incorporated to provide uncertainty estimation along with point predictions.

REFERENCES

1. T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "A deep learning framework for short-term power load forecasting," arXiv preprint arXiv:1711.11519, 2017.
2. Z. Zhang and W.-C. Hong, "Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm," *Nonlinear Dynamics*, vol. 98, no. 2, pp. 1107–1136, 2019.
3. Y. Liang, D. Niu, and W.-C. Hong, "Short-term load forecasting based on feature extraction and improved general regression neural network model," *Energy*, vol. 166, pp. 653–663, 2019.
4. Z. Zhang and W.-C. Hong, "Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads," *Knowledge-Based Systems*, vol. 228, 2021.
5. D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, 2000.
6. L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting," *Applied Energy*, vol. 167, pp. 135–153, 2016.
7. T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.
8. T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.
9. M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-Khah, and J. P. S. Catalão, "Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6961–6971, 2018.
10. J. Zhang, Y. Wang, M. Sun, and N. Zhang, "Two-stage bootstrap sampling for probabilistic load forecasting," *IEEE Transactions on Engineering Management*, vol. 69, no. 3, pp. 720–728, 2022.
11. M. Zahid et al., "Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids," *Electronics*, vol. 8, no. 2, p. 122, 2019.
12. Z. Cao, C. Wan, Z. Zhang, F. Li, and Y. Song, "Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1881–1897, 2020.
13. A. Bracale, P. Caramia, P. De Falco, and T. Hong, "A multivariate approach to probabilistic industrial load forecasting," *Electric Power Systems Research*, vol. 187, 2020.

14. W. Zhang, H. Quan, and D. Srinivasan, "Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination," *Energy*, vol. 160, pp. 810–819, 2018.
15. Y. Wang et al., "Probabilistic individual load forecasting using pinball loss guided LSTM," *Applied Energy*, vol. 235, pp. 10–20, 2019.
16. Y. Wang, Q. Chen, N. Zhang, and Y. Wang, "Conditional residual modeling for probabilistic load forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7327–7330, 2018.
17. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
18. L. Xu, M. Hu, and C. Fan, "Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks," *Journal of Building Engineering*, vol. 46, 2022.
19. Y. Yang, W. Hong, and S. Li, "Deep ensemble learning based probabilistic load forecasting in smart grids," *Energy*, vol. 189, 2019.
20. D. Gan, Y. Wang, S. Yang, and C. Kang, "Embedding based quantile regression neural network for probabilistic load forecasting," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 244–254, 2018.
21. Q. Huang, J. Li, and M. Zhu, "An improved convolutional neural network with load range discretization for probabilistic load forecasting," *Energy*, vol. 203, 2020.