

# ENERGY DEMAND PREDICTION USING ADVANCED MACHINE LEARNING TECHNIQUES AND REAL TIME WEATHER FEATURES

## GUIDE

<sup>1</sup>Name : Mrs. K. ANUSHA- [anushakunchakuri3@gmail.com](mailto:anushakunchakuri3@gmail.com)

Assistant Professor

Department of Computer Science and Engineering

TKRCET Autonomous ,

Hyderabad, India

<sup>2</sup>T. Lokesh

[tirunagirilokesh7@gmail.com](mailto:tirunagirilokesh7@gmail.com)

Department of Computer Science and Engineering

TKRCET Autonomous ,

Hyderabad, India

<sup>3</sup> V. Srinija

[v.srinija85@gmail.com](mailto:v.srinija85@gmail.com)

Department of Computer Science and Engineering

TKRCET Autonomous,

Hyderabad, India

<sup>4</sup> S. Nithin

[nithinsoma979@gmail.com](mailto:nithinsoma979@gmail.com)

Department of Computer Science and Engineering

TKRCET Autonomous ,

Hyderabad, India

<sup>5</sup> Y. Bharath

[yellavulab@gmail.com](mailto:yellavulab@gmail.com)

Department of Computer Science and Engineering

TKRCET Autonomous ,

Hyderabad, India

DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING

TKR COLLEGE OF ENGINEERING & TECHNOLOGY

(AUTONOMOUS)

(Accredited by NBA and NAAC with 'A+' Grade)

Medbowli, Meerpet, Saroornagar, Hyderabad-500097

## ABSTRACT

Energy demand forecasting plays a crucial role in modern power systems due to the increasing complexity of electricity consumption patterns influenced by weather conditions, seasonal variations, and human behavior. Traditional statistical methods often fail to accurately capture

these dynamic and non-linear relationships, leading to inefficiencies in energy generation and distribution. This project presents a machine learning-based approach for accurate electricity demand forecasting by integrating historical consumption data, weather parameters, and temporal features.

The proposed system utilizes multiple machine learning models, including Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Artificial Neural Networks (ANN), to analyze and predict energy demand. Advanced preprocessing techniques such as handling missing values, feature engineering (including lag features and rolling averages), and time-series analysis are applied to improve model performance. The models are evaluated using standard performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and  $R^2$  score.

Experimental results demonstrate that ensemble methods, particularly Gradient Boosting, outperform other models by providing higher accuracy and lower prediction errors. The system effectively captures real-world demand patterns, as evidenced by close alignment between actual and predicted values and a stable residual error distribution. Additionally, the model incorporates demand classification (low, medium, high) and provides decision-support recommendations for efficient energy management.

### **Keywords**

Energy Demand Forecasting, Machine Learning, Smart Grid, Time Series

Analysis, Gradient Boosting, Random Forest, Support Vector Regression (SVR), Artificial Neural Networks (ANN), Feature Engineering, Load Prediction, Energy Management, Data-Driven Modeling.

## **I. INTRODUCTION**

The increasing complexity of modern power systems, combined with the rapid growth in electricity consumption, has made accurate energy demand forecasting a critical requirement for efficient grid management. Electricity usage is no longer influenced by a single factor but is shaped by a combination of weather conditions, seasonal variations, population growth, urbanization, and changing human behavior. Emerging technologies such as electric vehicles, smart homes, and renewable energy integration have further increased the variability and uncertainty in demand patterns. As highlighted in the project document (Chapter 1), these dynamic conditions create significant challenges for maintaining a balance between energy generation and consumption.

Traditional forecasting methods, primarily based on statistical and linear models, often fail to capture the complex, non-linear relationships present in real-world energy data. These methods struggle to adapt to

sudden fluctuations caused by weather changes or unexpected events, leading to inaccurate predictions. Such inaccuracies can result in overproduction or underproduction of electricity, causing energy wastage, increased operational costs, and potential grid instability. Moreover, the growing adoption of renewable energy sources like solar and wind introduces additional uncertainty, as their generation depends heavily on environmental conditions.

To address these challenges, machine learning (ML) techniques have emerged as powerful tools for energy demand forecasting. ML models are capable of learning complex patterns from large datasets and can incorporate multiple influencing factors such as temperature, humidity, time of day, and historical consumption trends. Unlike traditional approaches, these models continuously improve their performance by adapting to new data, making them highly suitable for dynamic and real-time applications. This project leverages several machine learning algorithms, including Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Artificial Neural Networks (ANN), to develop a robust and scalable forecasting system.

The primary objective of this work is to design an accurate and reliable energy demand prediction system that integrates data preprocessing, feature engineering, and model evaluation techniques. By combining historical and real-time data, the system aims to improve prediction accuracy, support efficient energy distribution, and reduce operational costs. Additionally, the proposed system includes demand classification and decision-support mechanisms to assist utility providers and policymakers in making informed decisions.

## II. LITERATURE REVIEW

Energy demand forecasting has been widely studied due to its importance in maintaining efficient and stable power systems. Early research focused on understanding how external factors such as weather conditions and human activities influence electricity consumption. Studies show that temperature, seasonal variations, and social behaviors like working hours, weekends, and holidays significantly impact demand patterns. For instance, variations in temperature lead to increased heating or cooling requirements, while human behavioral patterns can cause noticeable fluctuations in daily and weekly consumption. These findings highlight the need to incorporate both environmental and

behavioral factors into forecasting models for improved accuracy.

Recent studies have also examined the effects of unexpected global events on electricity demand. The COVID-19 pandemic provided valuable insights into how sudden changes in human activity can disrupt traditional consumption patterns. Research indicates that strict lockdown measures reduced industrial and commercial energy usage while increasing residential demand. In contrast, regions with fewer restrictions experienced minimal changes. Such variations demonstrate the dynamic nature of electricity demand and emphasize the importance of adaptive forecasting models that can respond to real-time changes in consumption behavior.

In addition to behavioral and social factors, machine learning techniques have gained significant attention for their ability to improve forecasting accuracy. Traditional statistical models often rely on linear assumptions and fail to capture complex, non-linear relationships in energy data. In contrast, machine learning models such as Random Forest, Support Vector Machines, and Gradient Boosting can analyze large datasets and identify hidden patterns. Research has shown that ensemble methods, particularly Random Forest,

provide higher accuracy and require less parameter tuning compared to conventional approaches. These models effectively utilize historical load data, temporal features, and calendar information to enhance prediction performance.

Climate change is another critical factor influencing long-term electricity demand. Studies indicate that rising global temperatures lead to increased cooling requirements, especially in regions with hot climates. Even a small increase in temperature can significantly raise peak electricity demand, particularly during summer months. This trend suggests that future energy systems must account for climate-driven demand growth and incorporate environmental variables into forecasting models. Failure to do so may result in inaccurate predictions and inefficient energy planning.

Researchers have also explored the use of advanced weather indices and regression techniques to improve forecasting performance. Metrics such as Heating Degree Days (HDD) and Cooling Degree Days (CDD) have been introduced to capture non-linear temperature effects more effectively than simple temperature-based models. By combining these indices with economic and demographic factors, forecasting models can achieve higher

accuracy and better interpretability. Such approaches demonstrate the importance of integrating domain-specific knowledge into data-driven models.

Furthermore, the relationship between temperature and electricity demand has evolved over time due to technological advancements and changing usage patterns. Increased adoption of air-conditioning systems has shifted demand peaks from winter to summer in some regions. This transformation indicates that energy consumption patterns are continuously evolving, and forecasting models must adapt accordingly. Incorporating historical trends along with modern usage patterns is essential for developing reliable and future-ready forecasting systems.

Recent advancements in long-term load forecasting have focused on using high-resolution data and probabilistic approaches. Studies suggest that using hourly data instead of aggregated monthly or yearly data significantly improves prediction accuracy. Additionally, scenario-based models that consider multiple weather and economic conditions help address uncertainty in long-term forecasting. These approaches provide more reliable and practical solutions for energy planning and management.

### III. METHODOLOGY

The proposed system follows a structured machine learning pipeline to accurately forecast electricity demand using historical, environmental, and temporal data. The methodology consists of multiple stages, including data preprocessing, feature engineering, model training, prediction, and decision support. Each stage is designed to handle the complexity and variability of real-world energy consumption patterns.

Initially, data preparation and preprocessing are performed to ensure data quality and consistency. The energy consumption dataset is loaded and cleaned by handling missing values, where numerical values are replaced with mean and categorical values with mode. The timestamp is converted into a datetime format to extract useful temporal features such as hour, day of the week, month, and weekend indicators. The dataset is then split into training and testing sets, typically in an 80:20 ratio, and scaling is applied where necessary for specific models. This step ensures that the data is well-structured and suitable for machine learning algorithms.

The next stage involves feature engineering, which plays a crucial role in improving model performance. Important features such as temperature, humidity,

wind speed, and past electricity consumption are identified. Additional temporal and historical features are generated, including lag values (Lag\_1, Lag\_24) and rolling mean features (Rolling\_Mean\_3, Rolling\_Mean\_24). These features help capture both short-term and long-term dependencies in electricity usage patterns. Feature analysis is conducted to select the most relevant variables and eliminate redundant information, ensuring optimal input for model training.

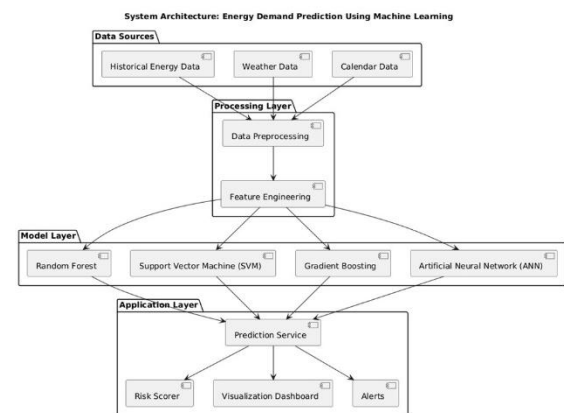
Following feature preparation, multiple machine learning models are trained and evaluated. The system implements Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Artificial Neural Networks (ANN). Each model is trained using the processed dataset and evaluated using performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R<sup>2</sup> score. A comparative analysis is performed to identify the best-performing model, with ensemble methods generally showing superior accuracy and robustness.

Once the optimal model is selected, the system proceeds to the prediction phase. User inputs, including timestamp, weather conditions, and past consumption values, are converted into feature vectors. The

trained model predicts the electricity demand, which is then classified into different levels such as low, medium, and high based on predefined thresholds. This classification enhances the interpretability of results and makes the system more practical for real-world applications.

Finally, the system incorporates a decision support and visualization module. Based on the predicted demand level, appropriate recommendations are provided, such as storing excess energy during low demand or activating peak load management during high demand. Visualization techniques, including actual vs predicted graphs, residual plots, error distribution, and feature importance analysis, are used to evaluate model performance and provide insights. These visualizations help stakeholders understand system behavior and make informed decisions.

#### IV. SYSTEM ARCHITECTURE



## V. RESULTS & DISCUSSION

The results of this project demonstrate the effectiveness of machine learning techniques in accurately forecasting electricity demand using historical, weather, and temporal data. Multiple models were implemented and evaluated to determine their suitability for handling complex and dynamic consumption patterns. The analysis not only focuses on numerical performance metrics but also includes graphical interpretations that provide deeper insights into model behavior and reliability.

The performance comparison of models shows that ensemble learning techniques outperform other approaches. Among the implemented models—Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Artificial Neural Networks (ANN)—the Gradient Boosting model achieved the highest accuracy with lower error values. Random Forest also performed well but was slightly less accurate, while SVR and ANN required more tuning to achieve comparable results. This indicates that ensemble models are better suited for capturing non-linear patterns and complex relationships in energy demand data.

The comparison between actual and predicted values reveals that the selected model effectively captures real-world demand trends. The predicted values closely follow the actual electricity consumption patterns, indicating that the model has successfully learned the underlying relationships in the dataset. Although minor deviations are observed, the overall trend alignment confirms the robustness and practical applicability of the forecasting system.

Residual error analysis further validates the model's performance. The residuals, representing the difference between actual and predicted values, are randomly distributed around zero. This pattern indicates that the model does not exhibit systematic bias, such as consistent overestimation or underestimation. A well-distributed residual pattern is a strong indicator of a well-trained and stable model capable of generalizing to unseen data.

The error distribution analysis shows a bell-shaped curve centered around zero, suggesting that most prediction errors are small and within an acceptable range. Only a few instances exhibit larger errors, which may be attributed to sudden fluctuations or extreme conditions in the dataset. This distribution confirms that the model

provides reliable and consistent predictions across different scenarios.

Feature importance analysis highlights the key factors influencing electricity demand. Historical consumption features, particularly Lag\_24 (previous day consumption), are identified as the most significant contributors to prediction accuracy. Rolling mean features also play an important role, while weather-related variables have comparatively lower influence. This indicates that past consumption patterns are the strongest predictors of future demand, reinforcing the importance of time-series-based feature engineering.

In addition to prediction accuracy, the system also demonstrates practical usability through demand classification and decision support. By categorizing demand into low, medium, and high levels, the system enables better energy management strategies such as load balancing and peak demand control. The integration of visualization tools further enhances interpretability, allowing users to understand model performance and make informed decisions.

## VI. CONCLUSION

This project demonstrates the effectiveness of machine learning techniques in

accurately forecasting electricity demand by utilizing historical consumption data, weather conditions, and temporal features. The study highlights the limitations of traditional statistical methods in handling complex, non-linear, and dynamic energy patterns, and shows how advanced machine learning models can overcome these challenges.

Multiple models, including Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Artificial Neural Networks (ANN), were implemented and evaluated. Among them, Gradient Boosting achieved the best performance with higher accuracy and lower prediction errors, confirming the strength of ensemble methods in energy demand forecasting. The system effectively captured consumption trends, as reflected in the close alignment between actual and predicted values, as well as stable residual and error distributions.

The integration of feature engineering techniques, such as lag features and rolling averages, significantly improved model performance by capturing time-dependent patterns in electricity usage. Additionally, the system enhances practical usability by classifying demand into low, medium, and high levels and providing decision-support

recommendations for efficient energy management.

Overall, the proposed system offers a scalable, reliable, and data-driven solution for both short-term and long-term energy demand forecasting. It supports better decision-making for utility providers, reduces operational costs, and contributes to the development of efficient and stable smart grid systems. The project establishes a strong foundation for future enhancements, including real-time data integration and advanced deep learning approaches for even more accurate and adaptive forecasting.

## REFERENCES

1. T. G. Cassarino, E. Sharp, and M. Barrett, "The effect of social and weather factors on electricity demand across Europe," *Energy Systems*, vol. XX, no. XX, pp. XX–XX, 20XX.
2. A. Bahmanyar, A. Estebarsari, and D. Ernst, "The impact of COVID-19 on electricity consumption: A case study of European countries," *Applied Energy*, vol. XX, no. XX, pp. XX–XX, 20XX.
3. D. Agdas and P. Barooah, "The impact of COVID-19 on electricity demand and grid performance in the United States," *Energy Reports*, vol. XX, no. XX, pp. XX–XX, 20XX.
4. G. Dudek, "Short-term load forecasting using Random Forest models," *International Journal of Forecasting*, vol. XX, no. XX, pp. XX–XX, 20XX.
5. S. Parkpoom and G. P. Harrison, "Impact of climate change on electricity demand in Thailand," *Energy Policy*, vol. XX, no. XX, pp. XX–XX, 20XX.
6. C. L. Hor, S. J. Watson, and S. Majithia, "Analyzing the impact of weather variables on electricity demand using regression models," *Energy Economics*, vol. XX, no. XX, pp. XX–XX, 20XX.
7. M. Hekkenberg, R. M. J. Benders, H. C. Moll, and A. J. M. Schoot Uiterkamp, "The changing relationship between temperature and electricity demand," *Energy*, vol. XX, no. XX, pp. XX–XX, 20XX.
8. T. Hong, J. Wilson, and J. Xie, "Long-term load forecasting using hourly data and probabilistic approaches," *IEEE Transactions on Power Systems*, vol. XX, no. XX, pp. XX–XX, 20XX.
9. J. Brownlee, "Machine Learning Algorithms for Time Series Forecasting," *Machine Learning Mastery*, 2018.
10. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.