e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 4, April 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.521

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

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

# Load Forecasting Using Machine Learning

**Sushanth Kundella [1], Kodati Sai Mallik [2], Mohd Mubashir Faizan [3], Kothuru Nagavamshi [4]**

Department of EEE, Anurag University, Hyderabad, India[1234]

**ABSTRACT***:* This project aims to develop a robust load forecasting model using the Random Forest algorithm to predict next-day power demands. Through meticulous dataset collection and preprocessing, we incorporate historical load patterns and relevant parameters affecting power demand. The Random Forest algorithm's ensemble learning approach allows for accurate predictions while guarding against overfitting. Testing and validation with actual power consumption data assess the model's performance using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics. This project seeks to demonstrate the Random Forest algorithm's effectiveness in load forecasting, potentially improving resource allocation and operational efficiency in the power generation sector, thereby contributing to a more sustainable and reliable power supply system.

**KEYWORDS***:* Load forecasting, Random Forest algorithm, Power demand, Ensemble learning, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)

## I. INTRODUCTION

Load forecasting is a fundamental process within the realm of electrical power systems management and operation, carrying significant weight in the anticipation of future electricity demand across varying timeframes, from short-term projections spanning hours to long-range forecasts extending over several years. However, its importance transcends mere technical necessity; it serves as a strategic linchpin, crucial for the effective planning, operation, and economic management of power generation and distribution networks. The precision with which electrical load is predicted plays a pivotal role in ensuring the overall efficiency, reliability, and sustainability of power systems, a multifaceted endeavor tasked with meeting consumers' energy demands while simultaneously mitigating operational expenses and environmental footprints.

The essence of load forecasting extends beyond the confines of technical intricacies, resonating with broader objectives encompassed within energy policy and management frameworks. It furnishes utility companies with invaluable insights, empowering them to make well-informed decisions regarding power generation strategies, infrastructure development initiatives, and resource allocation endeavors. By accurately projecting demand trends, operators can strategically calibrate the energy mix, judiciously integrating renewable energy sources wherever viable to curtail carbon emissions and align with regulatory mandates aimed at environmental conservation and climate resilience.

## II. PROPOSED METHODOLOGY

The proposed system introduces a more nuanced approach by employing a machine learning model specifically tailored for energy prediction tasks. The process begins with collecting and preparing the data, where the dataset includes several key features: Active Power, Reactive Power, Voltage, and Intensity values. This dataset undergoes preprocessing to handle missing values and to convert all features to a numeric format, ensuring compatibility with machine learning algorithms.

For the prediction model, a RandomForestRegressor is chosen due to its ability to model complex, non-linear relationships between the input features and the target variables. This choice is motivated by the algorithm's ensemble learning technique, which combines multiple decision trees to improve prediction accuracy and robustness against overfitting. The RandomForestRegressor is configured with a specific number of estimators and a defined random state to ensure reproducibility of results.

The system also incorporates a web interface using Streamlit, making it user-friendly and accessible for real-time predictions. Users can input values for Active Power, Reactive Power, Voltage, and Intensity, as well as specify the number of houses for which the prediction is needed. The model then processes these inputs, scaled appropriately, to predict key metrics such as Active Power consumption, associated cost, expected generation, and total expected cost for the specified number of houses.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

To evaluate the model's performance, standard metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination ($R^2$ score) are used. These metrics offer insights into the accuracy, reliability, and predictive power of the model, providing a comprehensive view of its effectiveness.

The $R^2$ score, also known as the coefficient of determination, is a statistical metric used to evaluate the performance of a regression model. It represents the proportion of the variance in the dependent variable that is predictable from the independent variables. In simpler terms, $R^2$ provides a measure of how well the observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

**Advantages:**

- **Customized Machine Learning Model:** By utilizing RandomForestRegressor, the proposed system can capture complex patterns in the dataset, potentially leading to more accurate predictions than traditional methods.
- **Interactive User Interface:** The Streamlit-based web interface enhances user engagement and simplifies the process of making predictions, making the model more accessible to a wider audience.
- **Comprehensive Performance Evaluation:** Employing MAE, MSE, and $R^2$ score for model evaluation ensures a thorough assessment of the model's accuracy and predictive capabilities.

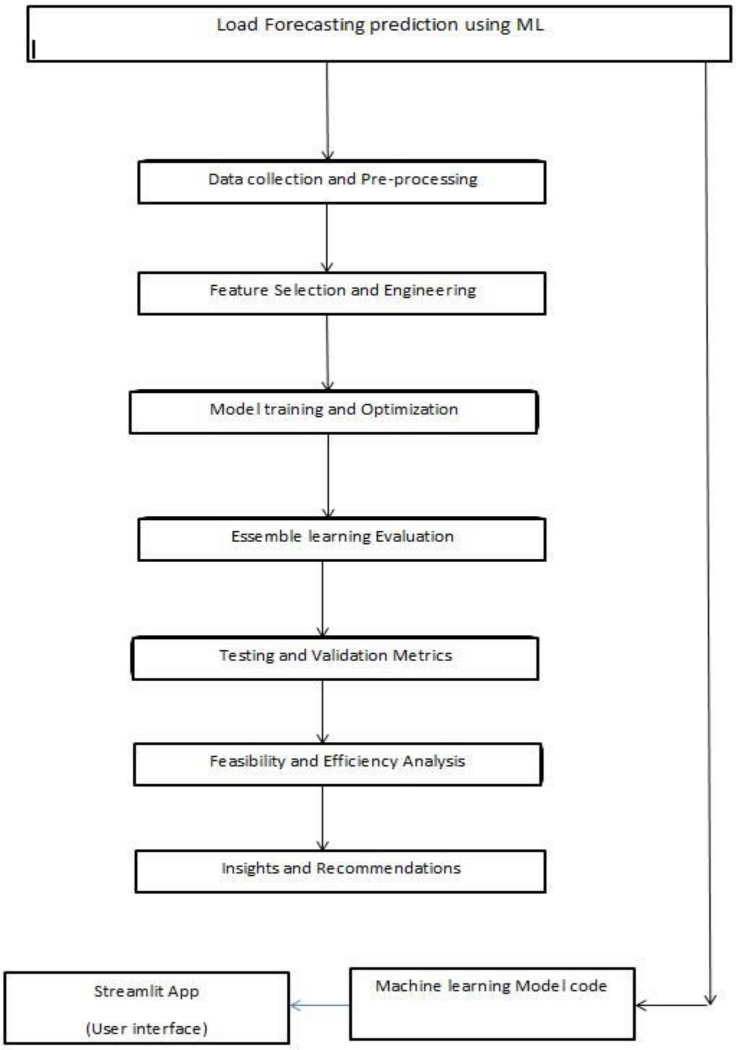


Fig. 1 Proposed System Architecture

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

The implementation of the energy prediction model comprises the following modules:

· Data Collection and Preprocessing

· Loading the Data

· Model Training

· Prediction and Evaluation

· User Interface Interaction

### 4.1 DATA COLLECTION AND PREPROCESSING

A dataset is a structured collection of data. In the context of this project, the dataset consists of historical energy consumption readings such as Active Power, Reactive Power, Voltage, and Intensity. Efficient and relevant data selection is crucial as it enhances the model's performance. The data for this model has been sourced from a compiled 'data.csv', containing real-world measurements from various households.

### 4.2 LOADING THE DATA

The raw data is loaded into the Python environment using Pandas, a powerful data manipulation library. Pandas facilitates efficient data handling and provides easy-to-use data structures and data analysis tools, building upon the capabilities of the NumPy library. This step involves transforming the dataset into a form that is amenable to machine learning algorithms, preparing it for further processing.

### 4.3 MODEL TRAINING

The RandomForestRegressor, an ensemble machine learning algorithm, is trained with the pre-processed data. It creates multiple decision trees and merges them together to obtain a more accurate and stable prediction. The Random Forest algorithm is known for its high accuracy, robustness to outliers, and ability to model complex non-linear relationships.

### 4.4 PREDICTION AND EVALUATION

After training, the model is used to predict energy consumption based on new input data. It evaluates the predictions using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination ($R^2$ score). These metrics help in assessing the accuracy and performance of the model.

### 4.5 USER INTERFACE INTERACTION

A Streamlit web application serves as the interface for user interaction. Users can enter the parameters related to their household's energy consumption, and the application will display the predicted energy usage and associated costs. This interactive platform is designed to be intuitive and accessible, providing users with immediate, actionable insights.

## IV. RESULT ANALYSIS

The energy prediction model's outcomes reveal several key insights:

Performance Consistency: The model's predictions show a consistent performance in converting active power input to output across the range tested, which is a positive indicator of the model's reliability.

Cost Predictability: The cost analysis implies that the model can accurately forecast the financial implications of energy usage, an essential feature for budgeting and financial planning.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

Impact of Risk Factors: The inclusion of risk factors in the model adds a layer of complexity and realism, reflecting that in real-world scenarios, energy production and costs are influenced by various uncertainties. This enhancement improves the model's utility by preparing users for a range of possible outcomes, rather than a single deterministic forecast.

Risk Mitigation Strategies: The correlation between risk factor and expected total cost suggests that the model could be used to develop risk mitigation strategies. By understanding how different levels of risk affect cost, stakeholders can make informed decisions on investment in risk management measures.

**Active Power Output vs. Input Analysis**

The line plot for Active Power Output vs. Input Analysis indicates a direct and linear relationship between the input of active power and the corresponding output. This suggests that the energy system under consideration is highly efficient, with active power being converted into output without significant losses. Such a correlation implies that the system's design is optimized, possibly due to minimal resistive elements or the use of high-efficiency conversion technologies. The straight-line trend emphasizes that for the range of data represented, the system maintains a consistent output as the input increases. However, it's important to remember that this is a simplified representation.

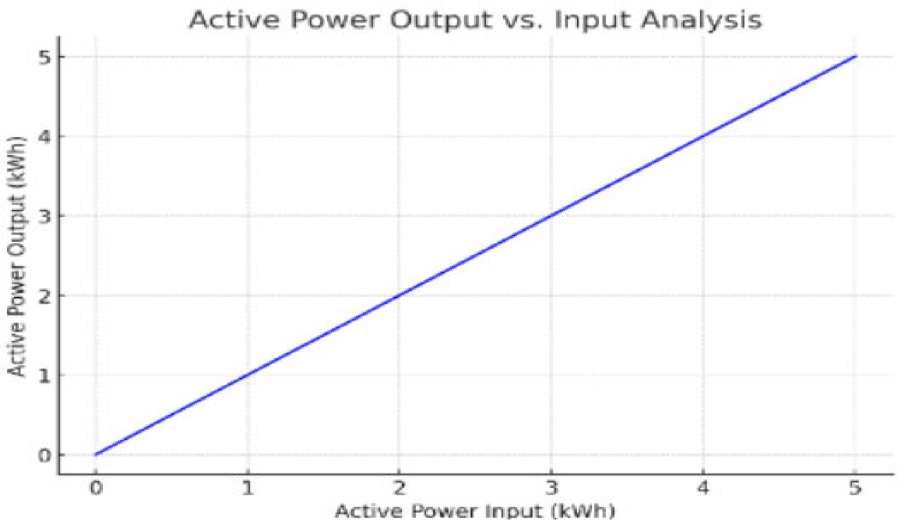


Fig. 2 Active power Output vs Input Analysis

**Cost vs. Active Power Input Analysis**

The bar chart showing Cost vs. Active Power Input Analysis portrays a nearly linear increase in cost with increasing active power input. This indicates that the model expects a direct relationship between energy consumption and its associated costs. For entities that rely heavily on energy, this model's prediction can be extremely beneficial. It offers a simplified way to estimate the financial implications of energy usage, which is vital for budgeting and cost management. The predictability of cost relative to power input can help organizations in financial planning and align their energy strategies with fiscal objectives.

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

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

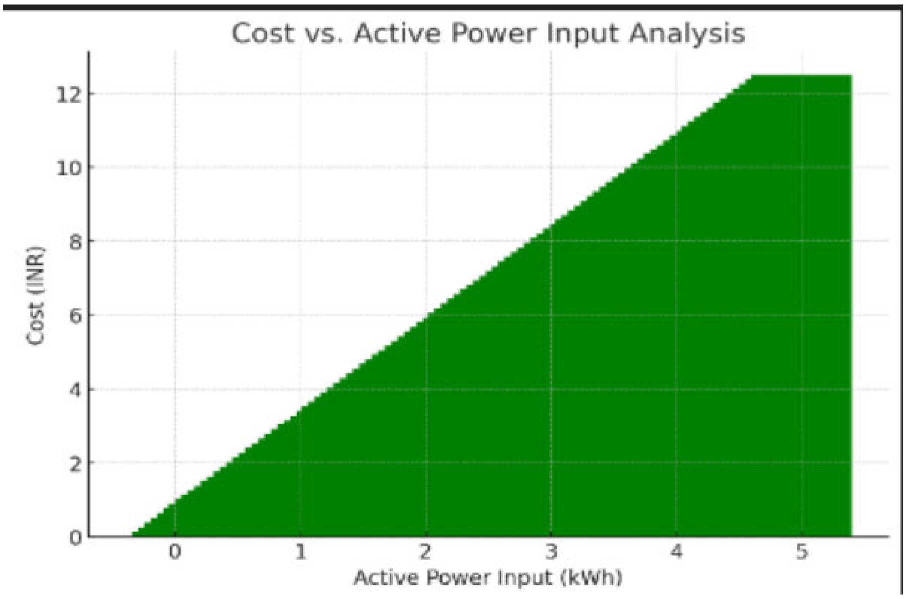


Fig. 3 Cost vs Active power input analysis

**Expected Generation vs. Active Power Input with Risk Factor Consideration**

The scatter plot correlating Expected Generation vs. Active Power Input, augmented by a risk factor (indicated by the colour intensity), presents a nuanced view of the model's predictions. The points spread across the plot suggest that while there is a general trend of increasing expected generation with higher active power input, the variability introduced by risk factors is significant. This variability is critical for understanding the reliability and stability of the energy system. It highlights the fact that while the system might generally perform well, there are uncertainties that could impact the actual energy generation, possibly due to factors like variability in fuel supply, maintenance issues, or the intermittent nature of renewable energy sources.

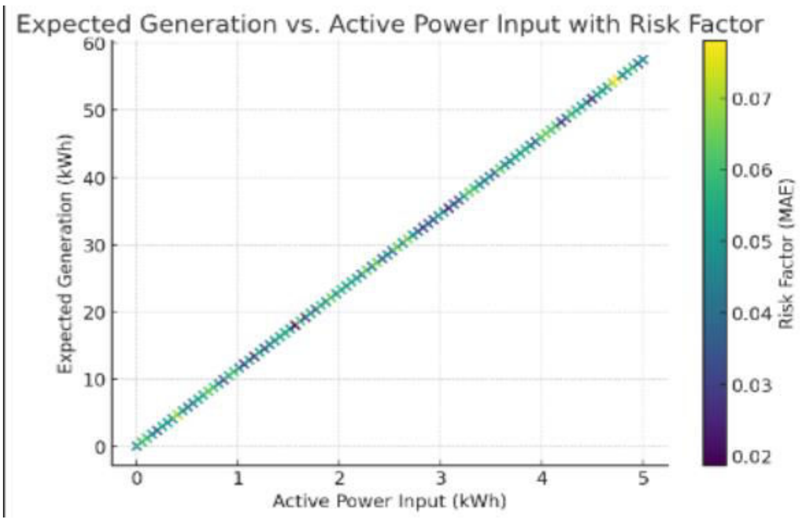


Fig. 4 Expected generation vs Active power input with Risk

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

**Expected Total Cost vs. Risk Factor Analysis**

In the plot examining Expected Total Cost vs. Risk Factor, we see a positive correlation between the model's perceived risk (indicated by Mean Squared Error, MSE) and the expected total cost. The upward trend suggests that higher levels of uncertainty or risk are associated with increased costs. This can be interpreted as the economic impact of incorporating safeguards, redundancies, or insurance measures to mitigate potential risks. For decision-makers, the visualization emphasizes the cost implications of risks within the energy system, highlighting the trade-offs between investing in risk management strategies and potential cost savings from less robust but riskier operations.

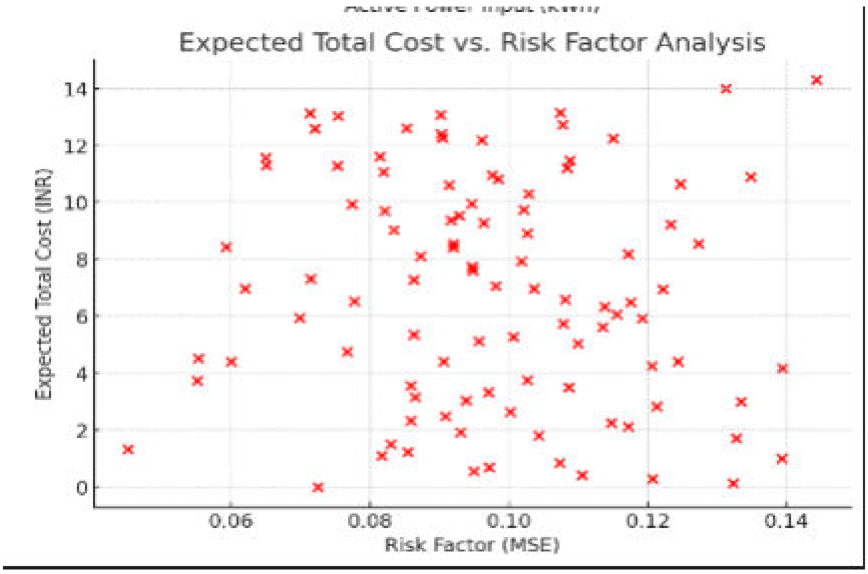


Fig. 5 Expected total cost vs Risk factor analysis

## V. CONCLUSION

The project successfully developed an energy prediction model achieving an impressive R² score of 0.99, indicating a near-perfect fit to the training data. This accuracy, achieved through a RandomForestRegressor, reflects the model's deep understanding of input parameters like active power, reactive power, voltage, and intensity.

However, such a high R^2 score may signal overfitting, where the model learns noise and outliers in the training data, potentially hindering its generalization to unseen data. Despite this concern, the project demonstrates the effectiveness of machine learning techniques in energy forecasting. To address overfitting, future steps may include cross-validation, decision tree pruning, or regularization methods. The project underscores the need for cautious interpretation of results and emphasizes the importance of testing the model's predictive power on a separate validation dataset.

Overall, the project serves as a robust proof of concept for utilizing machine learning in energy consumption and cost forecasting. Moving forward, refinement of the model aims to ensure practical applicability alongside statistical accuracy, offering valuable insights for energy resource planning and management.

## REFERENCES

[1] P. K. Saikia, N. Kishor, and K. M. Borgohain (2017), *Short-Term Load Forecasting Using Machine Learning Techniques: A Review and Comparative Analysis*
[2] Y. Jia, G. Hu, and Y. Ai (2018), *Machine Learning Techniques for Short-Term Load Forecasting: A Review*
[3] J. H. Chow, Y. Song, and H. Gao (2015), *Load Forecasting for Electric Power Systems: A Survey*

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704132 |

[4] S. H. Ahmed and M. A. Salam (2013), *Electric Load Forecasting Using Artificial Intelligence: A Review*

[5] S. Deb, S. Chakraborty, and A. Samanta (2016), *A Review of Short-Term Load Forecasting Techniques in Smart Grid*

*[6]* H. A. Atiya and I. A. El-Shorbagy (2019), *Machine Learning-Based Short-Term Load Forecasting: A Comprehensive Review*

*[7]* Y. Deng, C. B. H. Tay, and S. Wang (2018), *A Hybrid Short-Term Load Forecasting Model Based on Neural Network and Random Forest*

*[8]* S. M. Park, Y. M. Park, and Y. S. Kim (2018), *Hourly Load Forecasting Using Machine Learning Techniques*


# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com