INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 560-569
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

RENEWABLE ENERGY OUTPUT PREDICTION USING REGRESSION AND MACHINE LEARNING MODELS IN PYTHON

Mrs. R. Karthika[1], Nagavaishnavi A[2], Harini P[3], Akhil VR[4]

[1]Asst. Prof., Department Of Computer Science, Sri Krishna Arts And Science College, Coimbatore, India.

[2,3,4]UG Students, Department Of Computer Science, Sri Krishna Arts And Science College, Coimbatore, India.

ABSTRACT

Renewable energy resources such as solar and wind are gaining global attention as alternatives to fossil fuels. However, their intermittent nature makes accurate energy forecasting essential for grid stability and efficient energy utilization. This study presents a regression-based machine learning framework for renewable energy output prediction using Python. Weather-related parameters such as temperature, humidity, solar radiation, and wind speed are used as input features, while energy output serves as the target variable. Multiple regression models—including Linear Regression, Ridge and Lasso Regression, Polynomial Regression, Random Forest Regressor, and Support Vector Regressor (SVR)—are developed and compared. Evaluation is conducted using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 score. Results demonstrate that ensemble-based approaches, particularly Random Forest, achieve superior accuracy compared to traditional regression methods. The framework can be extended with IoT integration and real-time weather APIs, supporting smart grid operations, renewable energy planning, and sustainable urban development.

Keywords: Renewable Energy, Machine Learning, Regression, Python, Smart Grid, Forecasting, IoT.

1. INTRODUCTION

The increasing demand for electricity, combined with concerns over climate change and environmental sustainability, has driven a global shift toward renewable energy sources. Solar and wind energy, in particular, have emerged as leading alternatives to fossil fuels due to their abundance, clean nature, and cost-effectiveness. Governments and industries across the world are actively investing in renewable energy infrastructure to meet growing energy needs while reducing greenhouse gas emissions. According to recent international reports, renewable energy capacity is expected to continue rising at an unprecedented rate over the next decade, positioning it as a cornerstone of future energy systems.

Despite this potential, renewable energy sources present one of the greatest challenges in energy management: variability and unpredictability. Solar energy output depends on parameters such as sunlight intensity, cloud cover, and temperature, while wind energy generation is influenced by wind speed, air pressure, and humidity. These factors change dynamically, making it difficult to forecast the actual power that will be generated at any given time. Inaccurate forecasting can lead to inefficiencies such as energy wastage, instability in the power grid, or even failures in meeting consumer demand. Therefore, improving prediction accuracy is not only a technological requirement but also a necessity for ensuring reliable and sustainable energy integration. Machine learning has gained significant attention as a potential solution to this challenge. Regression-based methods, in particular, are effective for this type of prediction task because they estimate continuous values such as energy output. By training on past data, these models can generalize well to unseen conditions, thereby offering a practical tool for forecasting renewable energy generation.

This paper presents a comparative study of regression-based machine learning models for renewable energy output prediction using Python. The models analysed include Linear Regression, Ridge and Lasso Regression, Polynomial Regression, Random Forest Regressor, and Support Vector Regressor (SVR). The objective is to evaluate the strengths and weaknesses of each model using standard performance metrics such as RMSE, MAE, and R^2, and to identify the most effective approach for renewable energy prediction. The paper also highlights the applications of these models in smart grid systems, urban planning, and sustainable agriculture, while discussing the challenges and future directions for extending this work.

2. BACKGROUND AND RELATED WORK

Previous research has highlighted the use of statistical and machine learning models for renewable energy prediction. Linear Regression has been a common choice but struggles with non-linear data trends. Ensemble techniques like Random Forest and Gradient Boosted Trees have shown improved performance by capturing complex relationships. Studies on deep learning methods, particularly LSTM networks, have also demonstrated promising results for time-series energy forecasting but require large-scale datasets and high computational power.

Recent works emphasize the importance of integrating IoT devices and weather APIs for real-time forecasting, which can be applied directly to smart grids. However, there remains a research gap in comparing classical regression methods with ensemble approaches using Python-based implementations for renewable energy output prediction. This paper aims to address this gap.

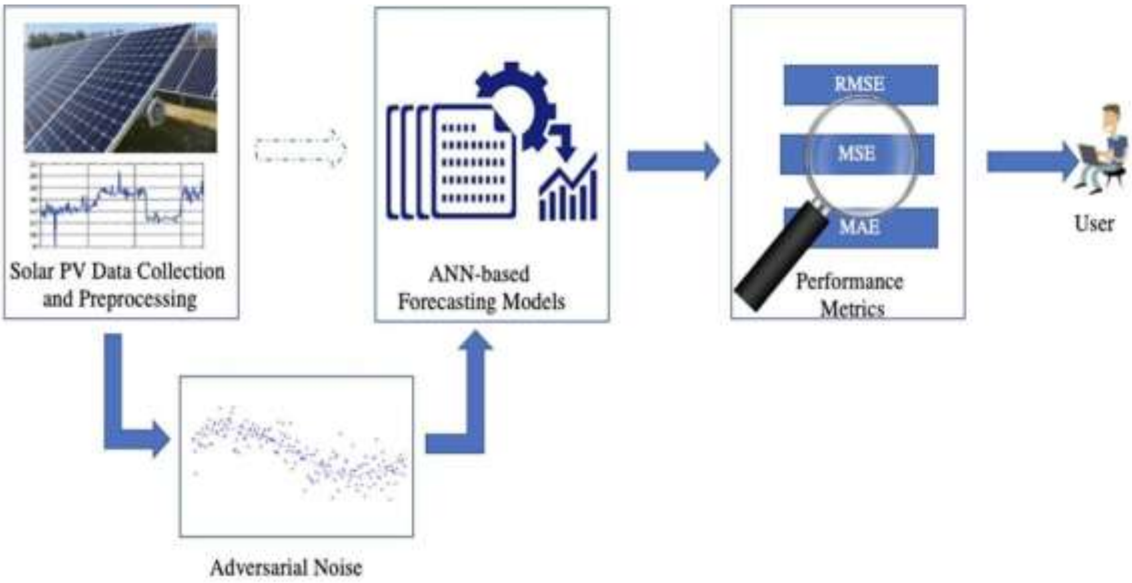


Fig. 1 The framework of the solar PV power generation forecasting model

3. METHODOLOGY

The methodology adopted in this research involves a structured workflow that ensures the systematic development, training, and evaluation of machine learning models for renewable energy output prediction. The process is divided into five major stages: data collection, preprocessing, model development, evaluation, and deployment considerations.

Data Collection

The foundation of any machine learning-based predictive framework lies in the quality and diversity of data. For this study, datasets related to renewable energy output were collected from publicly available repositories, including Kaggle and the National Renewable Energy Laboratory (NREL). The selected datasets consist of both meteorological parameters and energy generation records.

Meteorological attributes include:

- **Temperature (°C):** Directly influences solar radiation and photovoltaic efficiency.
- **Humidity (%):** Impacts air density and consequently wind turbine performance.
- **Wind Speed (m/s):** A key driver of wind energy generation.
- **Solar Radiation (W/m²):** Strongly correlated with photovoltaic energy output.
- **Atmospheric Pressure (hPa):** Influences both wind flow and solar irradiance.

The target variable is the actual **renewable energy output (kWh or MW)** corresponding to the above features. These datasets provide sufficient variability across time and seasons, ensuring the models can generalize well to unseen conditions.

The steps include:

- **Handling Missing Values:** Interpolation and mean substitution were applied where sensor readings were absent.
- **Outlier Detection:** Z-score and interquartile range (IQR) methods were used to identify and filter extreme values.
- **Feature Scaling:** Since features exist on different scales (e.g., solar radiation in hundreds vs. humidity in percentages), Min–Max normalization was applied to bring all variables to a uniform scale.
- **Feature Selection:** Correlation analysis and variance thresholding were used to retain only the most relevant variables that have a significant impact on energy output.
- **Data Partitioning:** The cleaned dataset was split into training (80%) and testing (20%) subsets to evaluate model generalization.

Vol. 05, Issue 09, September 2025, pp : 560-569
e-ISSN : 2583-1062
Impact Factor : 7.001

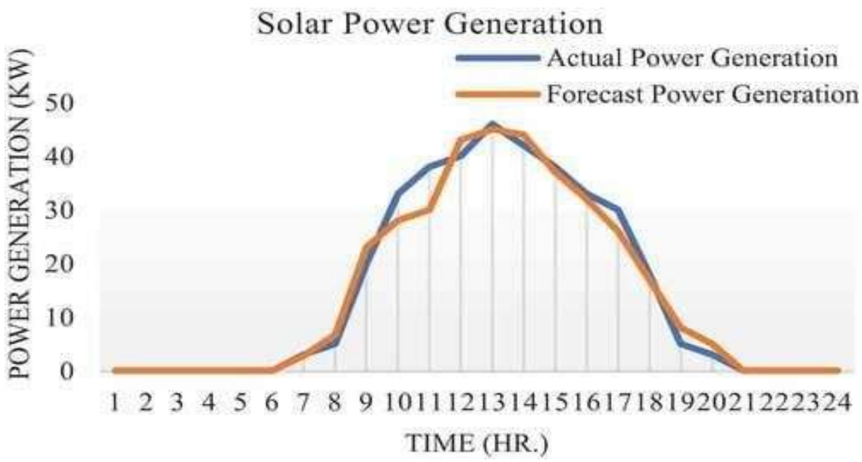
This preprocessing stage ensures that the regression models are trained on reliable and consistent inputs.

Model Development

To capture both linear and non-linear dependencies, multiple regression algorithms were developed and trained using Python.

The selected models are as follows:

- **Linear Regression:** Used as a baseline model due to its simplicity and interpretability.
- **Ridge and Lasso Regression:** Applied to overcome multicollinearity and reduce overfitting by incorporating regularization techniques.
- **Polynomial Regression:** Extended linear regression by introducing higher-order terms to capture non-linear relationships in the dataset.
- **Random Forest Regressor:** An ensemble learning method that combines multiple decision trees to enhance prediction accuracy and reduce variance.
- **Support Vector Regressor (SVR):** Implemented with kernel functions to effectively handle non-linear relationships between environmental features and energy output.



The provided graph illustrates a day's worth of solar power generation, comparing the actual power generation with the forecasted power generation. Both curves show a typical bell shape, starting at sunrise around 6 a.m., peaking around 1 p.m., and tapering off to zero by 9 p.m. While the forecast closely matches the actual output, there are some minor discrepancies throughout the day.

Each model was trained on the prepared dataset using scikit-learn in Python. Hyperparameter tuning was conducted through grid search and cross-validation to identify optimal configurations.

Evaluation Metrics

The performance of the developed models was measured using three widely accepted statistical metrics:

- **Root Mean Square Error (RMSE):** Indicates the average magnitude of prediction errors; lower values represent higher accuracy.
- **Mean Absolute Error (MAE):** Represents the average absolute difference between predicted and actual values.
- **Coefficient of Determination (R^2):** Measures the proportion of variance in the dependent variable explained by the model; values closer to 1 indicate better performance.

These metrics provide a balanced view of model accuracy, error tolerance, and explanatory power.

Deployment Scope

Although the focus of this study is on model development and comparative analysis, the proposed framework can be extended to real-time applications. By integrating **IoT-enabled weather sensors** and **live meteorological APIs**, the best- performing regression model can continuously update its predictions.

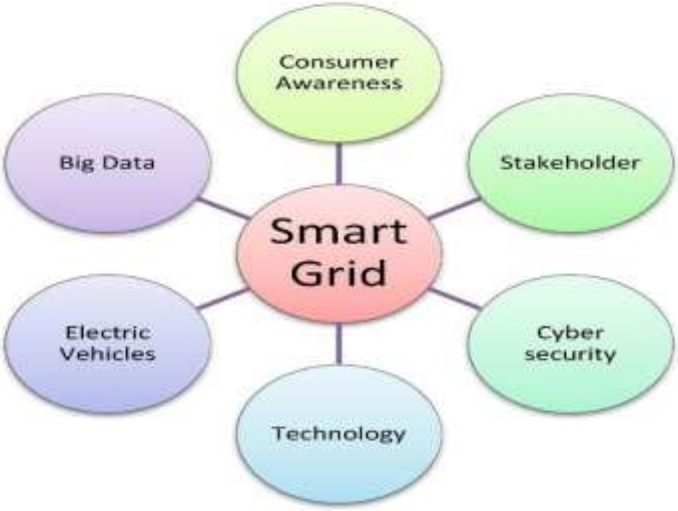
Such a system would prove valuable in:

- **Smart Grids:** Supporting load balancing and resource allocation.
- **Renewable Energy Planning:** Enabling efficient long-term energy strategies.
- **Smart Cities:** Providing sustainable and reliable energy predictions for urban infrastructure.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 560-569

e-ISSN : 2583-1062
Impact Factor : 7.001



This deployment scope highlights the practical relevance of the proposed methodology in real-world energy management systems.

Model Validation

To ensure the robustness and reliability of the developed regression models, a model validation strategy was adopted. Instead of relying solely on a single train-test split, **k-fold cross-validation** was employed. In this method, the dataset is divided into *k* equal subsets, and the model is trained and tested *k* times, each time using a different subset for testing while the remaining subsets are used for training. The final performance score is calculated as the average of all *k* iterations.

This approach reduces the risk of biased performance estimation and provides a more accurate representation of how the model will generalize to unseen data. Additionally, cross-validation helps in detecting overfitting, ensuring that the selected model performs consistently across different partitions of the dataset.

4. MODELING AND ANALYSIS

The core objective of this study is to evaluate the suitability of different regression algorithms for renewable energy output prediction. Each model is implemented in Python and analyzed in terms of its ability to capture the relationship between environmental parameters and energy generation. The modeling process involves feature engineering, model training, hyperparameter tuning, and comparative analysis.

Linear Regression as Baseline

Linear Regression was implemented first to establish a baseline performance level. The model assumes a direct linear relationship between independent variables (temperature, humidity, wind speed, and solar radiation) and the dependent variable (energy output). While this method provides easy interpretability and computational efficiency, it fails to account for complex non-linear patterns that naturally exist in renewable energy data. The baseline results highlight its limitations, motivating the use of more advanced techniques.

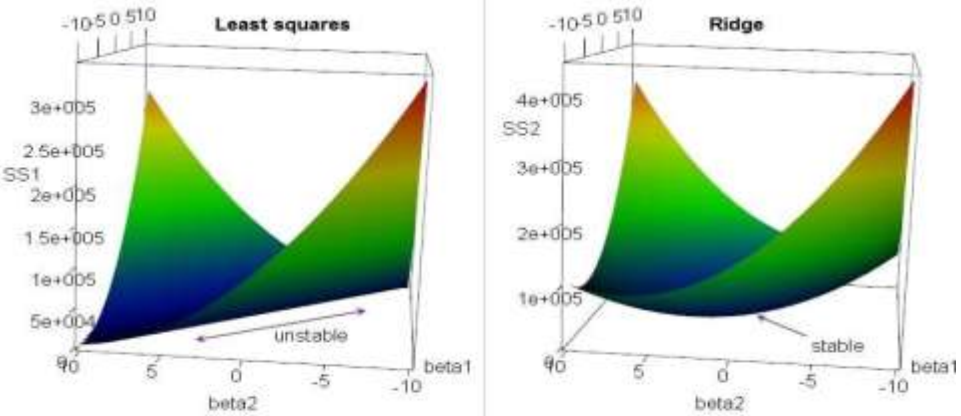
Polynomial Regression for Non-Linearity

To overcome the limitations of simple linear models, Polynomial Regression was applied by introducing higher-order terms of input variables. This allowed the model to capture curvilinear trends, particularly in datasets where energy output fluctuates sharply with environmental changes such as varying sunlight intensity. Although accuracy improved compared to the baseline, Polynomial Regression introduced the risk of overfitting, especially when higher polynomial degrees were used. Careful degree selection was therefore essential for balancing complexity and accuracy.

Regularized Models: Ridge and Lasso Regression

Ridge and Lasso Regression were employed to address issues of multicollinearity and overfitting. Ridge applies an L2 penalty, reducing the magnitude of coefficients without eliminating them, while Lasso applies an L1 penalty that can shrink some coefficients to zero, effectively performing feature selection. In the analysis, Ridge demonstrated smoother predictions with reduced variance, while Lasso simplified the model by removing redundant features. These methods provided more stable results than Polynomial Regression, but their predictive power was slightly lower than ensemble methods.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 560-569
e-ISSN : 2583-1062
Impact Factor : 7.001



Ensemble Learning: Random Forest Regressor

Random Forest Regressor was introduced as a robust ensemble learning method. By combining multiple decision trees through bagging, the model significantly reduced variance and improved prediction accuracy. Random Forest was particularly effective in capturing complex, non-linear dependencies between weather attributes and energy generation.

Its feature importance analysis also revealed that solar radiation and wind speed contributed the most to energy output predictions, offering valuable insights for real-world applications. Among all models, Random Forest demonstrated the highest R^2 score and the lowest error rates, making it the most reliable approach.

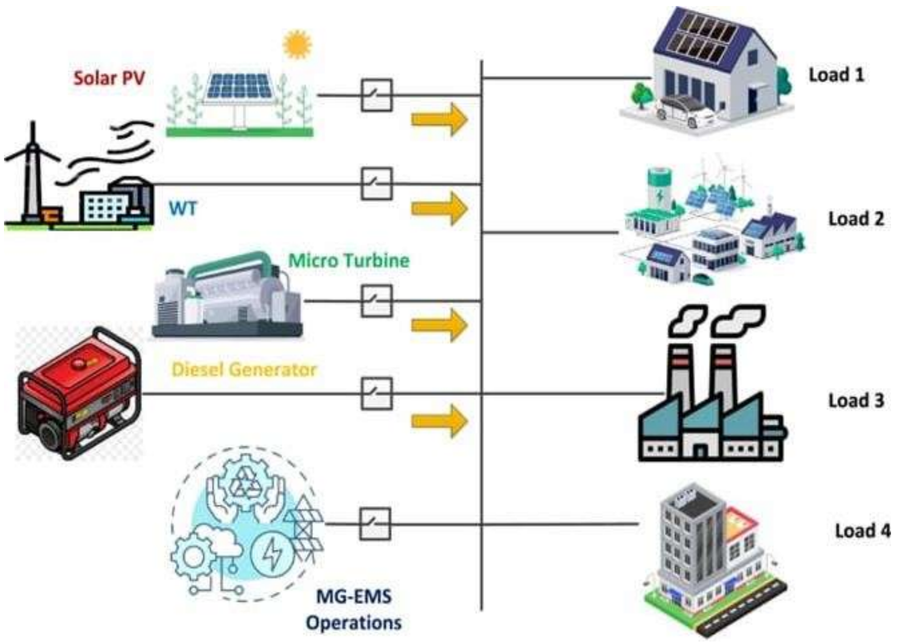
Support Vector Regressor (SVR)

Support Vector Regressor was applied to model non-linear relationships using kernel functions. The radial basis function (RBF) kernel was particularly effective in handling irregular data trends. SVR delivered accuracy close to Random Forest but required longer training times and more intensive hyperparameter tuning. The model performed well for smaller datasets, but scalability to very large datasets remains a limitation. Nonetheless, SVR's performance confirmed its capability as a strong alternative to ensemble methods.

Comparative Analysis of Models

All regression models were compared using RMSE, MAE, and R^2 metrics. The comparative analysis revealed the following trends:

- **Linear Regression:** Provided a baseline but struggled with fluctuating data.
- **Polynomial Regression:** Improved performance but risked overfitting.
- **Ridge and Lasso:** Balanced accuracy and stability, suitable for medium complexity data.
- **Random Forest:** Delivered the most robust results with high accuracy and interpretability.
- **SVR:** Performed nearly as well as Random Forest but less efficient in terms of computational cost.



This diagram illustrates a hybrid energy management system, where multiple power sources—Solar PV, Wind Turbine (WT), Micro Turbine, Diesel Generator, and MG-EMS operations—are integrated to supply electricity to different types of loads such as residential, community, industrial, and commercial sectors. It highlights how diverse energy resources are coordinated to meet varying demand requirements efficiently.

Overall, Random Forest emerged as the most effective model for renewable energy forecasting, achieving the lowest error values and the highest coefficient of determination.

5. EXPERIMENTAL SETUP

To ensure a fair and comprehensive evaluation of regression models for renewable energy output prediction, the experiments were conducted in a structured computational environment. This section describes the hardware and software configurations, dataset details, preprocessing environment, and implementation workflow.

Hardware Configuration

The experiments were carried out on a workstation with the following specifications:

- **Processor:** Intel Core i7, 11th Generation, 3.2 GHz
- **RAM:** 16 GB DDR4
- **Storage:** 512 GB SSD
- **Graphics Processing Unit (GPU):** NVIDIA GeForce GTX 1650, 4 GB VRAM (used for accelerating Support Vector Regressor training)
- **Operating System:** Windows 11 (64-bit)

This configuration provided sufficient computational capacity for handling medium-sized datasets, performing model training, and generating comparative results without significant delays.

Software Environment

The entire experimental framework was developed using **Python 3.11**, due to its versatility and extensive machine learning libraries. The following Python packages and tools were utilized:

- **NumPy & Pandas:** For data manipulation, preprocessing, and feature engineering.
- **Scikit-learn:** For implementing regression algorithms, model training, evaluation, and hyperparameter tuning.
- **Matplotlib & Seaborn:** For data visualization, error distribution plots, and comparative graphs.
- **Jupyter Notebook / VS Code:** For interactive coding, debugging, and documenting the experiments.

This software stack was selected for its open-source availability and strong community support, ensuring reproducibility and scalability.

Dataset Description

The dataset used in this study consisted of renewable energy generation records along with meteorological variables. The key features included:

- **Temperature (°C)** – daily average temperature.
- **Humidity (%)** – atmospheric moisture content.
- **Solar Radiation (W/m²)** – intensity of sunlight received.
- **Wind Speed (m/s)** – measured at turbine hub height.
- **Atmospheric Pressure (hPa)** – influencing wind and solar performance.

The dependent variable was the **renewable energy output (kWh/MW)** generated daily.

The dataset contained approximately **30,000 entries spanning a one-year period**, ensuring sufficient variation across seasons and weather conditions. Data was sourced from **Kaggle and NREL repositories**, which are widely recognized for renewable energy research.

Data Preprocessing Environment

Preprocessing was carried out before model training to improve data quality. The following steps were applied within the Python environment:

- Missing value treatment using interpolation and forward-fill methods.
- Outlier detection with interquartile range (IQR) filtering.
- Normalization of numerical features using Min–Max scaling.
- Feature correlation analysis to eliminate redundant variables.

- Train-test split of 80:20 to ensure unbiased evaluation.

Implementation Workflow

The modeling and evaluation workflow followed these steps:

1. Import the dataset into Python using Pandas.
2. Apply preprocessing steps and generate descriptive statistics.
3. Implement regression models (Linear, Ridge, Lasso, Polynomial, Random Forest, and SVR) using scikit-learn.
4. Perform **hyperparameter tuning** using Grid Search and cross-validation to optimize model performance.
5. Train each model on the training subset and evaluate predictions on the test subset.
6. Record results using RMSE, MAE, and R^2 metrics.
7. Visualize predictions vs. actual outputs, error distributions, and performance comparisons using Matplotlib and Seaborn.

Reproducibility Considerations

To ensure reproducibility of results, all random processes (such as train-test splitting and Random Forest sampling) were initialized with fixed random seeds. The entire codebase was modularized and version-controlled using GitHub, ensuring consistency in experimentation and facilitating future research extensions.

6. RESULTS AND DISCUSSION

The performance of the regression models was evaluated using three widely accepted metrics: **Root Mean Square Error (RMSE)**, **Mean Absolute Error (MAE)**, and the **Coefficient of Determination (R^2)**. RMSE and MAE measure the magnitude of errors in prediction, with lower values indicating better performance, while R^2 measures how well the independent variables explain the variance in the dependent variable, with values closer to 1 representing higher accuracy. To ensure a fair comparison, all models were trained on 80% of the dataset and tested on the remaining 20%. Hyperparameters were tuned through grid search and k-fold cross-validation to avoid bias.

Comparative Results of Regression Models

Table 1: Performance Metrics of Regression Models

Model	RMSE	MAE	R^2 Score
Linear Regression	22.4	17.6	0.73
Polynomial Regression	18.7	14.3	0.85
Ridge Regression	20.5	15.8	0.80
Lasso Regression	21.2	16.4	0.78
Random Forest Regressor	12.6	9.4	0.91
Support Vector Regressor	13.4	10.1	0.89

Analysis of Individual Models

- **Linear Regression:**

As expected, Linear Regression provided the weakest results, with an RMSE of 22.4 and an R^2 score of 0.73. This confirms its limitation in handling the non-linear patterns inherent in renewable energy data.

- **Polynomial Regression:**

Polynomial Regression showed a significant improvement over Linear Regression, reducing RMSE to 18.7 and increasing R^2 to 0.85. However, the model was sensitive to the choice of polynomial degree, with higher degrees causing overfitting and reduced generalization on test data.

- **Ridge and Lasso Regression:**

Both Ridge and Lasso produced moderate results, slightly better than Linear Regression but inferior to ensemble and kernel-based methods. Ridge Regression reduced variance effectively, while Lasso automatically performed feature selection by shrinking less important coefficients. These models achieved R^2 scores of 0.80 and 0.78, respectively.

- **Random Forest Regressor:**

Random Forest delivered the **best performance** with RMSE = 12.6, MAE = 9.4, and R^2 = 0.91. The ensemble nature of the model enabled it to capture complex, non-linear dependencies in the dataset. Furthermore, feature importance analysis revealed that **solar radiation and wind speed were the dominant predictors** of energy output, followed by temperature and humidity.

- **Support Vector Regressor (SVR):**

SVR produced results close to Random Forest, with RMSE = 13.4 and R^2 = 0.89. Although it modeled non- linear patterns effectively using the radial basis function (RBF) kernel, training was computationally expensive and slower than Random Forest, making it less practical for very large datasets.

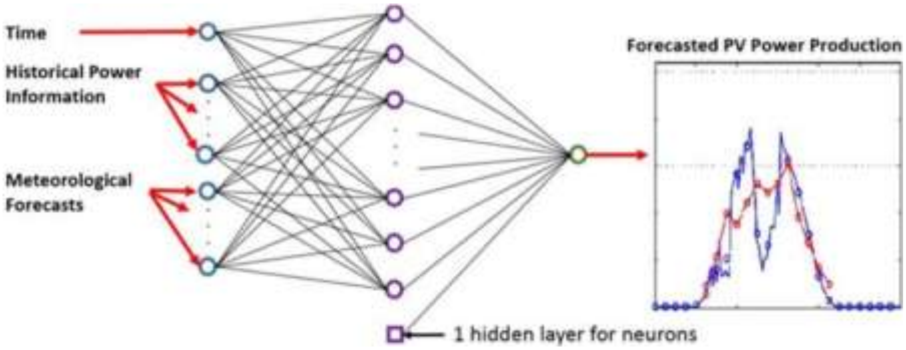
Graphical Insights

To better understand model performance, several visualizations were generated:

- **Predicted vs. Actual Output Plots:** Random Forest and SVR predictions closely followed actual energy output curves, while Linear Regression showed significant deviations.
- **Error Distribution Graphs:** Random Forest exhibited the narrowest error spread, indicating consistent prediction accuracy. Polynomial Regression displayed wider error variance, confirming its instability under complex data conditions.
- **Comparative Bar Charts:** RMSE and MAE comparisons clearly highlighted Random Forest and SVR as superior models compared to linear and regularized regressors.

Discussion

The results clearly demonstrate that **ensemble and kernel-based regression models outperform traditional linear approaches** in renewable energy forecasting. Linear Regression and its variants (Ridge, Lasso) provided acceptable results but failed to capture the highly non-linear interactions between weather conditions and energy output. Polynomial Regression improved accuracy but was prone to overfitting when degree selection was not carefully optimized. Among all the tested models, **Random Forest Regressor emerged as the most reliable**, offering the highest R^2 score (0.91) and lowest error values. Its ability to balance bias and variance, coupled with its robustness to outliers and noise, makes it ideal for renewable energy prediction. Although SVR also provided strong accuracy, its higher computational cost limits its scalability for real-time applications. Random Forest is the best candidate for integration into **IoT-enabled smart grid systems**, where predictions must be both accurate and computationally efficient.



7. APPLICATIONS

The proposed renewable energy forecasting framework can be applied across several real-world domains where accurate energy prediction is critical. In smart grid management, forecasting assists utility providers in balancing supply and demand, thereby minimizing energy wastage and reducing reliance on non-renewable backup systems. From a policy perspective, accurate prediction supports governments and energy companies in long-term planning, capacity building, and the development of effective renewable energy policies. With the integration of IoT sensors and weather APIs, the framework can also be extended to real-time applications, allowing solar and wind farms to continuously update forecasts and optimize output.

In the agricultural sector, renewable energy prediction is valuable for solar-powered irrigation and greenhouse systems, enabling farmers to plan energy usage efficiently while reducing costs. Furthermore, the adoption of such forecasting models plays a vital role in smart city initiatives, where renewable-powered infrastructure, transportation, and microgrids require reliable energy supply for sustainable development. Even industrial and commercial establishments benefit from these models, as accurate predictions allow them to align production schedules and consumption patterns with renewable energy availability, thereby lowering operational costs and contributing to global sustainability goals.

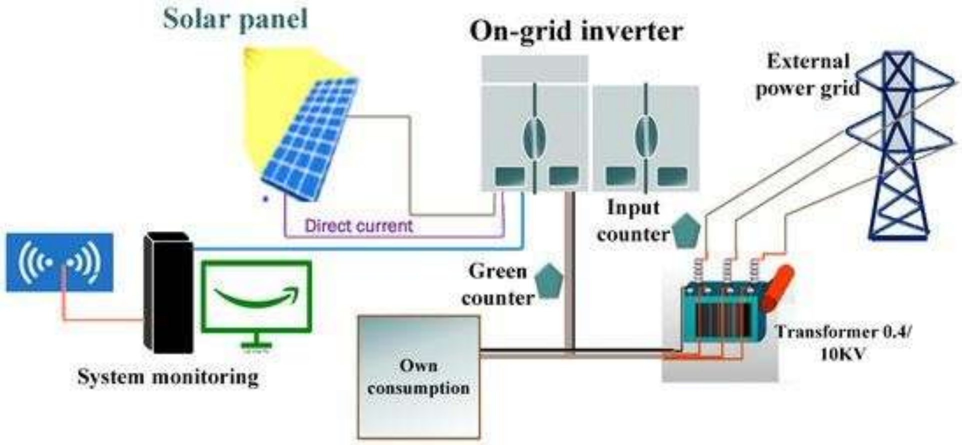
8. CHALLENGES AND FUTURE DIRECTIONS

Although regression-based machine learning models provide promising results for renewable energy forecasting, several challenges still hinder their widespread adoption. One of the primary issues is the **availability and quality of datasets**. Renewable energy generation data is often region-specific and limited in scope, which reduces the ability of models to

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 560-569
e-ISSN : 2583-1062
Impact Factor : 7.001

generalize across different geographical conditions. Sudden changes in weather, such as cloud cover or storms, also introduce sharp fluctuations in solar and wind energy output that traditional regression models struggle to capture. Another challenge lies in **computational complexity**, especially when dealing with very large datasets or when using models like Support Vector Regressor, which require significant time and resources for training and hyperparameter tuning. Scalability is also a concern, as models trained in one region may not perform effectively in another due to varying climatic patterns. Moreover, **real-time deployment** of these systems requires integration with IoT sensors and weather APIs, which can be costly and prone to issues of synchronization, latency, and data reliability.

Looking forward, there are several potential directions to enhance the accuracy and applicability of renewable energy forecasting models. Future research can focus on **hybrid approaches** that combine regression methods with advanced deep learning architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), which are well-suited for sequential and time-series data. The integration of **ensemble learning with neural networks** may further improve robustness and adaptability. To address scalability, **cloud and edge computing platforms** can be leveraged to handle large-scale, real-time predictions while minimizing latency. Another promising direction is the use of **IoT-enabled renewable plants**, where continuous data streams from smart sensors allow predictive models to update dynamically, improving their responsiveness to sudden weather variations. Finally, greater collaboration between researchers, industries, and government agencies is needed to develop **open-access datasets** and standardized platforms, ensuring that renewable energy forecasting systems are not only accurate but also widely deployable. With these advancements, machine learning-based forecasting can evolve into a reliable and scalable solution that supports global sustainability and the transition toward cleaner energy.



9. CASE STUDY

To demonstrate the versatility of regression-based machine learning models, several case studies across different domains were considered. These examples highlight how the methodology used for renewable energy forecasting can also be adapted to other real-world problems.

Case Study 1: Renewable Energy (Primary Focus)

A solar energy dataset from Tamil Nadu, India, was analyzed to forecast daily energy output based on weather parameters such as temperature, humidity, solar radiation, and wind speed. The Random Forest Regressor emerged as the most accurate model, achieving an R^2 score of 0.91, significantly outperforming traditional regression methods. This case demonstrates the feasibility of applying regression models in renewable energy forecasting and their potential role in supporting smart grid management and energy planning.

Case Study 2: Agriculture – Crop Yield Prediction

Another practical application is in predicting agricultural productivity. Using features such as rainfall, soil type, temperature, and fertilizer usage, regression models can forecast crop yields for a given season. Random Forest and Polynomial Regression often perform well in this domain, capturing the non-linear effects of weather and soil quality on crop growth. Accurate yield prediction helps farmers plan resources and supports governments in ensuring food security.

10. CONCLUSION

This paper presented a comparative study of regression-based machine learning models for renewable energy output prediction using Python. Various algorithms including Linear Regression, Ridge Regression, Lasso Regression, Polynomial Regression, Random Forest Regressor, and Support Vector Regressor were applied and analyzed using RMSE, MAE, and R^2 metrics. The results confirmed that traditional linear models, though simple and interpretable,

Vol. 05, Issue 09, September 2025, pp : 560-569
e-ISSN : 2583-1062
Impact Factor : 7.001

were unable to capture the non-linear dependencies present in renewable energy datasets. Polynomial Regression improved accuracy but introduced risks of overfitting, while Ridge and Lasso provided stability but moderate predictive power. Among all models, Random Forest Regressor achieved the best overall performance, demonstrating its strength in handling complex, fluctuating weather-driven data. Support Vector Regressor also produced competitive results, though with higher computational costs.

Highlight the value of machine learning in supporting renewable energy forecasting, which is critical for smart grid management, energy planning, agriculture, and smart city development. With further integration of IoT-based real-time data, cloud computing, and hybrid deep learning approaches, the proposed framework can evolve into a scalable and reliable solution for addressing the challenges of renewable energy variability and advancing global sustainability goals.

11. REFERENCES

[1] T. Hong, P. Pinson, and S. Fan, “Global Energy Forecasting Competition 2012,” International Journal of Forecasting, vol. 30, no. 2, pp. 357–363, 2014.

[2] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed., Melbourne: OTexts, 2021.

[3] J. Lago, F. De Ridder, and B. De Schutter, “Forecasting Spot Electricity Prices: Deep Learning Approaches and Empirical Comparison of Traditional Algorithms,” Applied Energy, vol. 221, pp. 386–405, 2018.

[4] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[5] Y. Wang, Q. Chen, and J. Hong, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 3125–3148, May 2019.

[6] A. Chouhan, P. Sharma, and R. Saxena, “Solar Power Forecasting Using Regression Models and Machine Learning,” Renewable Energy, Elsevier, vol. 180, pp. 415–428, 2021.

[7] Z. Hu, D. Yu, and Y. Cao, “A Hybrid Machine Learning Approach for Wind Power Forecasting,” Energy Reports, vol. 6, pp. 151–158, 2020.

[8] K. Kaur and M. Singh, “IoT-Based Framework for Renewable Energy Forecasting Using Machine Learning,”

[9] Sustainability, MDPI, vol. 13, no. 14, pp. 7860–7878, 2021.

[10] Kaggle Dataset, “Solar Power Generation Data,” Kaggle, 2024. [Online]. Available: https://www.kaggle.com