Smart Solar Energy Forecasting Using Stacking and Voting Regression Techniques

Ms. Ummadi Poojitha[1], Smt. M. Prashanthi[2]

[1]PG Scholar, Department of CSE, JNTUA College of Engineering (Autonomous) Ananthapuramu, India.

[2]Assistant Professor, Department of CSE, JNTUA College of Engineering (Autonomous) Ananthapuramu, India

DOI: https://doi.org/10.51244/IJRSI.2026.1310000011

Received: 07 October 2026; Accepted: 12 October 2026; Published: 28 October 2026

ABSTRACT

Solar energy has emerged as one of the most promising renewable energy sources for sustainable power generation. However, the intermittent nature of solar radiation and changing weather conditions make accurate energy forecasting a challenging task. Reliable solar energy prediction is essential for efficient grid management, energy storage planning, and optimal utilization of renewable resources. This research presents a Smart Solar Energy Forecasting framework using Stacking Regression and Voting Regression Techniques to improve prediction accuracy and reliability.The proposed system utilizes historical solar power generation data along with meteorological parameters such as temperature, humidity, wind speed, and solar irradiance. Data preprocessing techniques including cleaning, normalization, and feature selection are applied to enhance data quality. Multiple machine learning algorithms are trained and evaluated, including Linear Regression, Decision Tree Regression, Support Vector Regression (SVR), Voting Regressor, and Stacking Regressor. Ensemble learning approaches combine the strengths of multiple predictive models to generate highly accurate forecasts. The Voting Regressor aggregates predictions from different base learners, while the Stacking Regressor employs a meta-learning strategy to optimize final predictions. Experimental results demonstrate that ensemble methods outperform individual regression models in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 Score.

Keywords: Smart Solar Energy Forecasting, Solar Power Prediction, Machine Learning, Ensemble Learning, Stacking Regression, Voting Regression, Renewable Energy, Solar Irradiance, Explainable Artificial Intelligence (XAI), SHAP, LIME, Smart Grid, Energy Management, Predictive Analytics, Sustainable Energy.

INTRODUCTION

Accurate prediction of solar energy production plays a crucial role in optimizing renewable energy utilization and supporting the development of sustainable energy infrastructure. With the increasing global demand for clean energy, solar power has emerged as a reliable and eco-friendly alternative. However, the variability of environmental conditions such as temperature, humidity, and solar irradiance makes it challenging to predict solar power generation accurately. This project addresses this challenge by leveraging machine learning-based regression techniques applied to real-world solar generation data. The study begins by analyzing key environmental features that influence solar energy output and uses them to build predictive models. Traditional regression algorithms, including Linear Regression, Support Vector Regression, and Decision Tree Regression, are implemented to establish baseline performance. To further enhance prediction accuracy, advanced ensemble methods such as Stacking Regressor and Voting Regressor are employed. These ensemble techniques combine multiple models to improve generalization, robustness, and overall performance in forecasting solar power output. The outcome of this project aims to assist stakeholders in making data-driven decisions for efficient solar energy production and planning. By improving forecasting accuracy, the proposed approach supports energy sustainability, minimizes resource wastage, and promotes smart grid integration. Additionally, it enables better

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

real-time energy demand balancing and enhances the reliability of renewable energy systems. Furthermore, this project highlights the importance of data preprocessing and feature selection in improving model accuracy. Proper handling of missing values and normalization enhances prediction reliability. Performance is evaluated using metrics such as MAE, MSE, and R-squared score. These steps ensure the models are efficient and suitable for real-world solar energy applications.

Objective of the Study

The primary objective of this research is to develop and evaluate an intelligent solar energy forecasting framework using **Stacking Regression** and **Voting Regression** techniques. With the growing demand for renewable energy, accurate prediction of solar power generation has become essential for efficient energy management, grid stability, and sustainable resource utilization. Traditional forecasting methods often fail to capture the complex and nonlinear relationships between weather parameters and solar energy output, creating the need for advanced machine learning-based forecasting systems. This study aims to develop a forecasting model that utilizes historical solar power generation data along with meteorological parameters such as solar irradiance, temperature, humidity, wind speed, and atmospheric conditions to accurately predict future solar energy production. The proposed system incorporates data preprocessing, feature engineering, and ensemble machine learning algorithms to improve prediction accuracy while reducing forecasting errors under varying environmental conditions. Furthermore, the research focuses on combining the strengths of multiple regression algorithms through **Stacking Regression** and **Voting Regression** techniques. These ensemble learning methods improve model generalization, reduce overfitting, and provide more reliable predictions than individual machine learning models. The study also compares the performance of different regression algorithms using evaluation metrics such as **MAE, MSE, RMSE, and R^2 Score** to identify the most effective forecasting model. Another important objective is to improve the transparency and interpretability of forecasting results by integrating **Explainable Artificial Intelligence (XAI)** techniques such as **SHAP** and **LIME**. These techniques help identify the contribution of each environmental parameter to the predicted solar power output, enabling researchers and energy providers to better understand the model's decision-making process and increase confidence in the forecasting results. Finally, this research aims to support smart energy management by providing accurate short-term solar energy forecasts that assist utility providers, smart grids, and policymakers in energy scheduling, demand planning, storage optimization, and efficient utilization of renewable resources. By developing a reliable and scalable forecasting framework, the study contributes to sustainable energy development and promotes the effective integration of solar power into modern energy systems. Furthermore, the proposed framework aims to enhance the reliability and efficiency of renewable energy systems by minimizing forecasting uncertainties and supporting data-driven decision-making. It is designed to provide scalable and adaptable forecasting solutions that can be implemented across different geographical regions and climatic conditions. Moreover, this research seeks to provide a practical and cost-effective forecasting solution that can be adopted by renewable energy providers, research organizations, and smart city infrastructures. The proposed framework is designed to facilitate real-time decision-making, improve energy distribution efficiency, and reduce operational risks associated with uncertain solar power generation. By integrating advanced ensemble learning techniques with explainable artificial intelligence, the study aims to establish a robust, transparent, and intelligent forecasting system that supports the global transition toward clean, reliable, and sustainable energy resources.

3. Scope of the Study

This study focuses on developing and implementing an intelligent solar energy forecasting framework using **Stacking Regression** and **Voting Regression** techniques to accurately predict solar power generation under varying environmental conditions. As solar energy has become one of the most significant renewable energy sources, accurate forecasting is essential for efficient power generation, energy scheduling, smart grid management, and optimal utilization of renewable resources. Traditional forecasting methods often fail to capture the complex and nonlinear relationships between meteorological factors and solar energy production. This research addresses these challenges by employing advanced ensemble machine learning techniques that enhance forecasting accuracy and reliability. The proposed framework emphasizes the analysis of historical solar power generation data along with important meteorological parameters such as **solar irradiance, ambient temperature, humidity, wind speed, atmospheric pressure, and module temperature**. Through effective

data preprocessing, feature engineering, normalization, and data cleaning techniques, the system extracts meaningful patterns that improve prediction performance. The forecasting model is designed to handle large datasets efficiently while minimizing prediction errors caused by changing weather conditions and seasonal variations. The study also investigates the application of **ensemble learning algorithms**, specifically **Stacking Regression** and **Voting Regression**, which combine the strengths of multiple regression models to produce more accurate and robust predictions than individual machine learning algorithms. Furthermore, the research evaluates the performance of different regression models, including **Linear Regression, Decision Tree Regression, and Support Vector Regression (SVR)**, using standard performance metrics such as **Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R^2 Score**. The integration of **Explainable Artificial Intelligence (XAI)** techniques such as **SHAP** and **LIME** further enhances the transparency and interpretability of the forecasting results. In addition, the proposed system is intended to support real-world applications such as **smart grids, renewable energy management systems, utility providers, and energy planners** by providing accurate short-term solar energy forecasts. Reliable forecasting enables efficient demand planning, battery storage optimization, load balancing, and effective integration of renewable energy into existing power systems. The framework can also assist policymakers and energy organizations in making informed decisions regarding sustainable energy utilization. Overall, this research contributes to the advancement of intelligent renewable energy forecasting by providing a scalable, reliable, and data-driven solution that combines ensemble machine learning with explainable AI. The proposed framework not only improves forecasting accuracy but also supports the development of smart energy systems, enhances grid stability, reduces energy wastage, and promotes the efficient adoption of clean and sustainable solar energy technologies. The proposed framework is designed to be flexible and scalable, allowing it to adapt to different solar energy datasets and geographical locations. It also provides a foundation for future enhancements through the integration of real-time IoT sensor data, deep learning models, and cloud-based energy management systems. By improving forecasting accuracy and operational efficiency, the study supports the widespread adoption of renewable energy and contributes to a cleaner and more sustainable future.

4. Problem Statement

Accurate solar energy forecasting has become one of the most significant challenges in renewable energy management due to the highly dynamic and unpredictable nature of weather conditions. Solar power generation is greatly influenced by environmental factors such as solar irradiance, temperature, humidity, wind speed, cloud cover, and atmospheric pressure, making it difficult to achieve precise energy predictions using conventional statistical methods. Inaccurate forecasting can lead to inefficient energy scheduling, power imbalance, increased operational costs, and reduced reliability of smart grid systems. Therefore, developing an intelligent and reliable forecasting system has become essential for maximizing the utilization of solar energy and ensuring stable power distribution. Traditional solar energy forecasting techniques primarily rely on statistical models or individual machine learning algorithms, which often struggle to capture the complex and nonlinear relationships between meteorological variables and solar power output. These methods frequently produce inaccurate predictions when weather conditions change rapidly or when large volumes of historical data need to be processed. Furthermore, single-model forecasting approaches are prone to overfitting, poor generalization, and limited adaptability, reducing their effectiveness in real-world renewable energy applications. Another major challenge is the lack of forecasting models that provide both high prediction accuracy and transparency in their decision-making process. Although advanced machine learning techniques have improved forecasting performance, many existing models operate as "black boxes," making it difficult for researchers, utility providers, and energy planners to understand how different environmental factors influence the predicted solar power output. The absence of explainable forecasting models reduces user confidence and limits their practical adoption in critical energy management systems. This research addresses these challenges by proposing an intelligent **Smart Solar Energy Forecasting** framework based on **Stacking Regression** and **Voting Regression** techniques. The proposed system combines the strengths of multiple regression algorithms to improve forecasting accuracy, reduce prediction errors, and enhance model generalization. It also integrates **Explainable Artificial Intelligence (XAI)** techniques such as **SHAP** and **LIME** to provide transparent and interpretable predictions by identifying the contribution of each weather parameter to the forecasting results. The proposed framework is designed to accurately predict short-term solar power generation using historical solar energy data and meteorological parameters while supporting efficient energy scheduling, smart grid management, and renewable energy planning. By leveraging ensemble

machine learning and explainable AI, the system aims to overcome the limitations of traditional forecasting methods, improve decision-making, and contribute to the development of reliable, scalable, and sustainable solar energy forecasting solutions for modern power systems.

Relatedwork

The increasing demand for renewable energy and the growing adoption of photovoltaic (PV) systems have encouraged researchers to develop advanced solar energy forecasting techniques that improve prediction accuracy and operational efficiency. **[1]** Traditional statistical forecasting methods, such as persistence models and autoregressive techniques, have been widely used for solar power prediction. However, these approaches often fail to capture the nonlinear relationships between meteorological parameters and solar energy generation, leading to reduced forecasting accuracy under dynamic weather conditions. As a result, machine learning-based forecasting methods have gained significant attention due to their ability to model complex environmental patterns and improve prediction performance. A major area of research focuses on the application of **[2] machine learning algorithms** such as Linear Regression, Decision Tree Regression, Support Vector Regression (SVR), Random Forest, and Artificial Neural Networks (ANNs) for solar energy forecasting. These models utilize historical solar generation data along with environmental parameters such as solar irradiance, temperature, humidity, wind speed, and atmospheric pressure to predict future energy production. Although these algorithms have demonstrated promising results, individual machine learning models often suffer from limitations such as overfitting, poor generalization, and reduced accuracy under varying climatic conditions. **[3]** Consequently, researchers have explored advanced ensemble learning techniques that combine multiple models to achieve better predictive performance. Recent studies have highlighted the effectiveness of **[4] ensemble learning methods**, particularly **Voting Regression** and **Stacking Regression**, in enhancing solar energy forecasting accuracy. Voting Regression combines the predictions of multiple regression models to produce a robust final prediction, while Stacking Regression employs a meta-learning strategy that learns the optimal combination of several base models. These techniques significantly reduce prediction errors and improve model stability by leveraging the strengths of individual algorithms. Several comparative studies have reported that ensemble models outperform traditional regression techniques when evaluated using performance metrics such as **Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R^2 Score**. Another important direction in solar energy forecasting research is the integration of **[5] Explainable Artificial Intelligence (XAI)** techniques to improve model transparency and user trust. Advanced forecasting models often function as black-box systems, making it difficult for users to understand the reasoning behind predictions. To address this challenge, researchers have incorporated explainability methods such as **SHAP (SHapley Additive explanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)**. These techniques identify the contribution of each meteorological parameter to the predicted solar energy output, enabling researchers, utility providers, and energy planners to better interpret forecasting results and make informed decisions. Furthermore, **[6] deep learning models**, including Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs), have been widely investigated for time-series solar energy forecasting. These models effectively capture temporal dependencies in historical solar generation data and weather variables, leading to improved prediction performance. However, deep learning models generally require large datasets, significant computational resources, and longer training times, which may limit their applicability in resource-constrained environments. Consequently, lightweight ensemble learning approaches remain attractive for many practical renewable energy forecasting applications. Recent advancements also emphasize the integration of **[7] Internet of Things (IoT)** technology and smart sensors for real-time solar energy monitoring and forecasting. IoT-enabled photovoltaic systems continuously collect meteorological and operational data, enabling forecasting models to adapt dynamically to changing environmental conditions. Researchers have also explored cloud computing and edge computing platforms to process large-scale solar datasets efficiently and provide real-time forecasting services for smart grid applications. These technologies contribute to improving forecasting reliability while supporting intelligent energy management systems.

In addition, **[8] hybrid forecasting frameworks** that combine statistical models, machine learning algorithms, and ensemble learning techniques have demonstrated superior performance compared to standalone models. By integrating multiple forecasting approaches, these hybrid systems effectively reduce uncertainty, improve prediction robustness, and enhance adaptability across different climatic regions. Such frameworks are

particularly useful for utility providers and renewable energy planners seeking reliable solar energy forecasts for power scheduling, storage optimization, and demand management. More recently, researchers have focused on developing **[9] intelligent renewable energy management systems** that integrate accurate solar forecasting with smart grid technologies, battery storage optimization, and demand-response strategies. These systems enable efficient energy distribution, minimize power fluctuations, and maximize renewable energy utilization. The incorporation of ensemble learning techniques, explainable AI, and real-time data analytics has significantly improved forecasting accuracy while enhancing user confidence and decision-making capabilities.Overall, the existing literature demonstrates that ensemble machine learning methods, particularly **Stacking Regression** and **Voting Regression**, provide highly accurate and reliable solutions for solar energy forecasting. Despite significant progress, there remains a need for scalable, interpretable, and computationally efficient forecasting frameworks capable of handling diverse environmental conditions and large-scale renewable energy datasets. The proposed research addresses these challenges by combining advanced ensemble learning with explainable AI techniques to develop an intelligent, transparent, and robust solar energy forecasting system suitable for modern smart grid and sustainable energy applications.

Proposed System Workflow

The proposed system is designed to enhance **solar energy forecasting** by utilizing **Stacking Regression** and **Voting Regression** techniques, integrated with advanced data preprocessing, feature engineering, and Explainable Artificial Intelligence (XAI) methods. The framework consists of several key components, including data collection, preprocessing, feature extraction, ensemble learning, forecasting, and model interpretation, which work together to generate accurate and reliable solar power predictions.

Data Collection: The system collects historical solar power generation data along with important meteorological parameters such as **solar irradiance, ambient temperature, humidity, wind speed, atmospheric pressure, and module temperature**. These environmental factors serve as the primary inputs for the forecasting model and play a significant role in predicting future solar energy generation.

Data Preprocessing and Feature Engineering: The collected dataset undergoes preprocessing techniques including missing value handling, normalization, feature scaling, and data cleaning to improve data quality. Feature engineering is then applied to extract meaningful attributes and eliminate irrelevant information, enabling the forecasting model to learn more effectively from historical data and improve prediction performance.**Ensemble Learning Models:**The proposed framework employs **Stacking Regression** and **Voting Regression** as its primary forecasting algorithms. Voting Regression combines predictions from multiple regression models to generate a more stable and accurate output, while Stacking Regression utilizes a meta-learner to optimally integrate the predictions of several base models. These ensemble learning techniques reduce prediction errors, improve model generalization, and provide superior forecasting accuracy compared to individual regression algorithms.

Forecasting and Model Interpretation: The trained ensemble models generate accurate short-term solar energy forecasts based on historical and real-time weather conditions. To improve transparency and user confidence, the framework integrates **Explainable Artificial Intelligence (XAI)** techniques such as **SHAP** and **LIME**, which identify the contribution of each meteorological parameter to the final prediction. This enables utility providers, energy planners, and researchers to better understand the forecasting process and make informed decisions for efficient renewable energy management.

A. Dataset Stacking

This research utilizes a comprehensive solar energy dataset containing historical solar power generation records and meteorological parameters that serve as the foundation for training the machine learning models. The dataset typically includes features such as **solar irradiance, ambient temperature, humidity, wind speed, atmospheric pressure, module temperature, DC power, and AC power**, which significantly influence solar energy production. The dataset is generally stored in CSV format and is loaded into the system using Python libraries such as **Pandas** and **NumPy** for efficient data manipulation and analysis. After loading the dataset, the data is carefully examined to identify missing values, duplicate records, and inconsistencies. Necessary data

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

cleaning procedures are then performed to improve data quality, and the dataset is divided into training and testing sets to evaluate the performance and generalization capability of the forecasting models.

B. Preprocessing

Data preprocessing is an essential step in preparing the solar energy dataset for machine learning applications. Since the collected data may contain missing values, noise, and inconsistencies caused by sensor errors or changing environmental conditions, appropriate preprocessing techniques are applied to improve data quality. Missing values are handled using suitable imputation methods or by removing incomplete records where necessary. Feature scaling and normalization techniques are employed to ensure that all input variables contribute equally during model training. In addition, feature engineering is performed to derive meaningful features from the original dataset, while irrelevant or redundant attributes are eliminated to improve forecasting accuracy. The dataset is then divided into training and testing subsets, typically using an **80:20 ratio**, enabling reliable evaluation of the forecasting models under different weather conditions.

C. Model Preparation and Grouping

In this stage, machine learning models are trained to accurately forecast solar energy generation based on historical power generation data and meteorological parameters. After preprocessing, several regression algorithms, including **Linear Regression, Decision Tree Regression, Support Vector Regression (SVR), Stacking Regression, and Voting Regression**, are implemented and compared. Individual regression models establish baseline forecasting performance, while the ensemble learning techniques combine the strengths of multiple models to improve prediction accuracy and reduce forecasting errors. The models are evaluated using standard performance metrics such as **Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R^2 Score**. Furthermore, **Stacking Regression** employs a meta-learner to optimally combine the outputs of base regression models, whereas **Voting Regression** generates the final prediction by aggregating the outputs of multiple models. The integration of these ensemble learning techniques significantly enhances forecasting reliability, improves model generalization, and provides accurate solar energy predictions suitable for smart grid and renewable energy management.

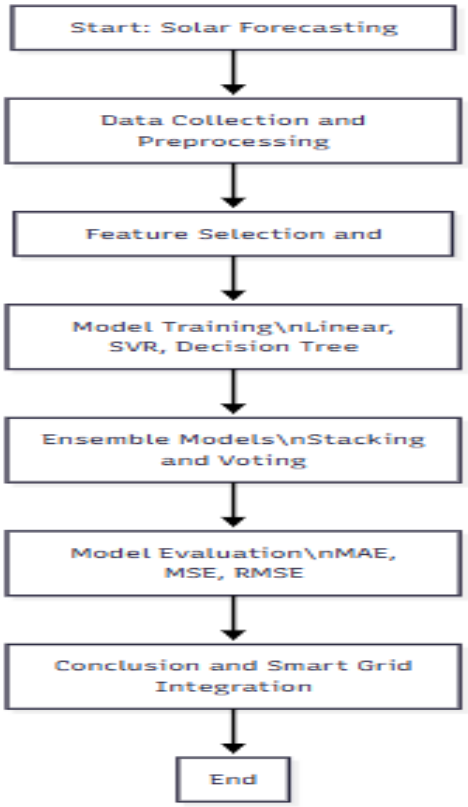


Fig1:Project Flow for Solar Energy Prediction

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

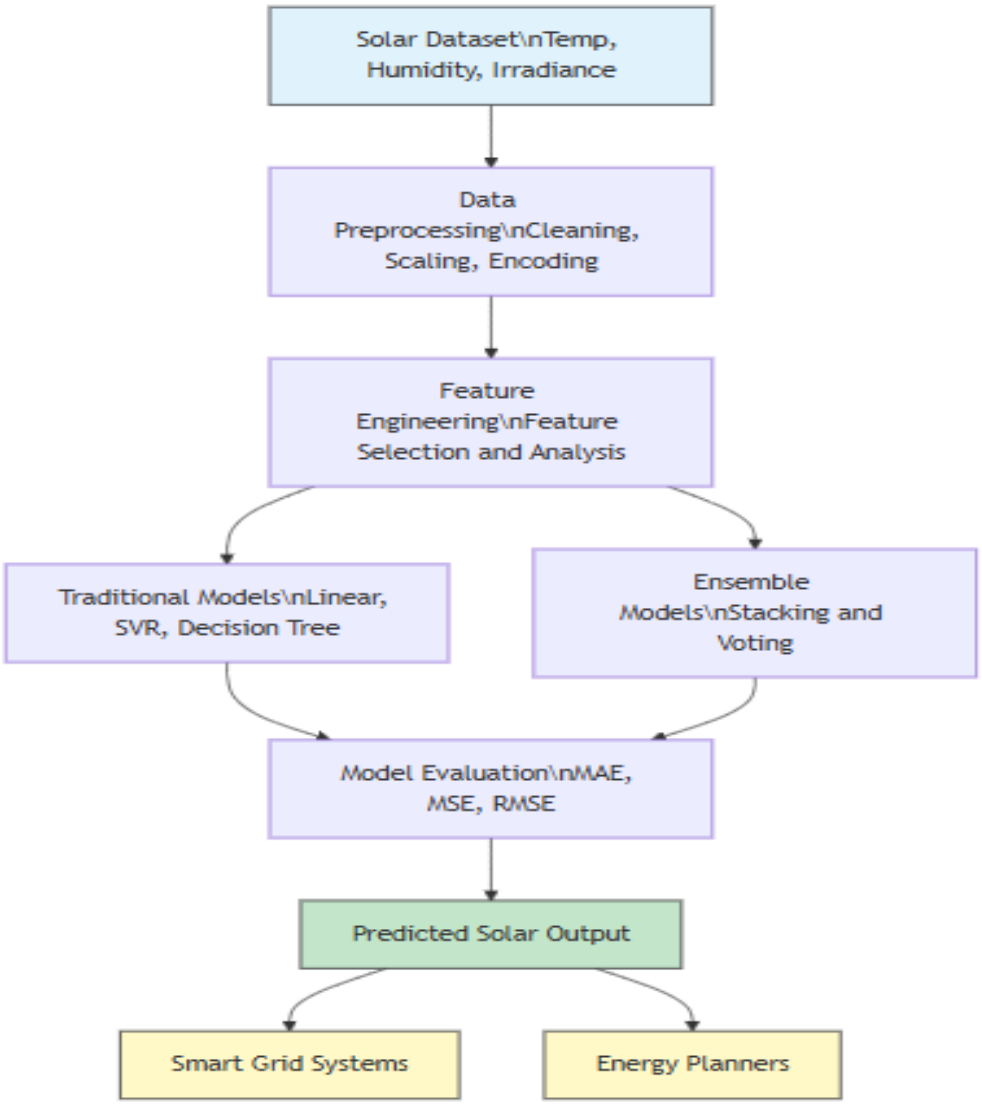


Fig2:Architecture for Solar Energy Prediction

System Design

The system design for **Smart Solar Energy Forecasting using Stacking Regression and Voting Regression Techniques** is developed to provide an efficient, scalable, and intelligent framework for predicting solar power generation. The proposed architecture integrates data preprocessing, feature engineering, ensemble machine learning algorithms, explainable artificial intelligence, model evaluation, and visualization to improve forecasting accuracy and reliability. The framework is designed to process large volumes of historical solar power generation data and meteorological information while providing transparent and interpretable predictions. By combining multiple regression models through ensemble learning, the system minimizes prediction errors and supports effective renewable energy management, smart grid operations, and sustainable energy planning.

1. System Architecture Overview: The proposed system architecture consists of six major layers that work together to perform accurate solar energy forecasting. These layers include the **Data Acquisition and Storage Layer**, **Data Preprocessing and Feature Engineering Layer**, **Modeling Layer**, **Explainability Layer**, **Evaluation and Visualization Layer**, and an optional **Deployment Layer** for real-time forecasting. Each layer performs a specific function, beginning with data collection and ending with prediction visualization and deployment. The modular architecture improves system flexibility, simplifies maintenance, and enables future enhancements by incorporating additional datasets, advanced forecasting models, and real-time monitoring capabilities.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

2. Data Acquisition and Storage Layer

The data acquisition layer is responsible for collecting and storing historical solar energy generation data and meteorological information required for model training and prediction. The proposed system utilizes a real-world solar power generation dataset obtained from Kaggle, containing attributes such as solar irradiance, ambient temperature, humidity, module temperature, wind speed, DC power, and AC power generation. The dataset is stored in structured CSV format and imported into the Python environment using libraries such as Pandas and NumPy for efficient data manipulation. If required, the dataset can also be maintained in relational databases such as MySQL for efficient storage and retrieval. This layer ensures that reliable and high-quality data is available for subsequent preprocessing and machine learning operations.

3. Data Preprocessing and Feature Engineering Layer: The preprocessing layer prepares the collected solar energy dataset for machine learning by improving data quality and extracting meaningful information. Initially, missing values, duplicate records, and inconsistent entries are identified and handled using appropriate preprocessing techniques such as mean or median imputation. The timestamp information is converted into datetime format, allowing the extraction of useful temporal features such as hour, day, month, and seasonal variations. Feature engineering techniques are then applied to derive additional attributes, including power efficiency ratio and time-based characteristics that improve forecasting performance. Data normalization and feature scaling are performed to ensure that all input variables contribute equally during model training. Outliers caused by sensor failures or abnormal environmental conditions are detected and removed using statistical methods such as the Z-score or Interquartile Range (IQR). When necessary, dimensionality reduction techniques such as Principal Component Analysis (PCA) are applied to reduce feature complexity while preserving important information for prediction.

4. Modeling Layer: The modeling layer is the core component of the proposed system, where different machine learning algorithms are trained, evaluated, and compared for solar energy forecasting. Initially, baseline regression models such as **Linear Regression**, **Support Vector Regression (SVR)**, **Decision Tree Regression**, and **Random Forest Regression** are developed to establish reference forecasting performance. Subsequently, advanced ensemble learning techniques including **Voting Regression**, **Stacking Regression**, and **Gradient Boosting Regression** are implemented to improve prediction accuracy and model robustness. Voting Regression combines the outputs of multiple regression models through averaging, while Stacking Regression employs a meta-learner to optimally integrate predictions from several base models. Gradient Boosting Regression further enhances forecasting performance by sequentially correcting prediction errors produced by previous models. Each model undergoes training using train-test splitting or K-Fold Cross Validation, followed by hyperparameter optimization using Grid Search CV or Randomized Search CV. The forecasting performance is evaluated using standard regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R^2 Score, and the trained models are stored for future comparison and deployment.

5 Explainability Layer: To improve transparency and interpretability, the proposed forecasting framework integrates Explainable Artificial Intelligence (XAI) techniques, including **SHAP (SHapley Additive exPlanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)**. These techniques help explain the predictions generated by ensemble learning models by identifying the contribution of each meteorological parameter toward the final solar energy output. LIME generates local explanations by approximating the behavior of complex machine learning models with simple interpretable models, while SHAP computes feature importance using cooperative game theory principles. These explainability methods enable researchers, utility providers, and energy planners to understand the influence of environmental variables such as solar irradiance, humidity, wind speed, and temperature on forecasting results. Consequently, the framework enhances user confidence, supports transparent decision-making, and facilitates regulatory compliance in intelligent energy management systems.**6 Evaluation and Visualization Layer**

The evaluation and visualization layer assesses the performance of all forecasting models and presents the results through meaningful graphical representations. The trained models are compared using widely accepted regression evaluation metrics including **Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R^2 Score** to determine the most accurate forecasting approach. Various

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

visualization techniques, including bar charts, scatter plots, line graphs, SHAP summary plots, and LIME explanation graphs, are generated to compare prediction accuracy, analyze forecasting errors, and identify the importance of individual input features. These visualizations assist researchers and decision-makers in selecting the most reliable and interpretable forecasting model for deployment in real-world renewable energy applications. The optional deployment layer further enables real-time prediction services for smart grids and energy management systems, making the proposed framework practical, scalable, and suitable for sustainable solar energy forecasting.

Use Case Diagram:

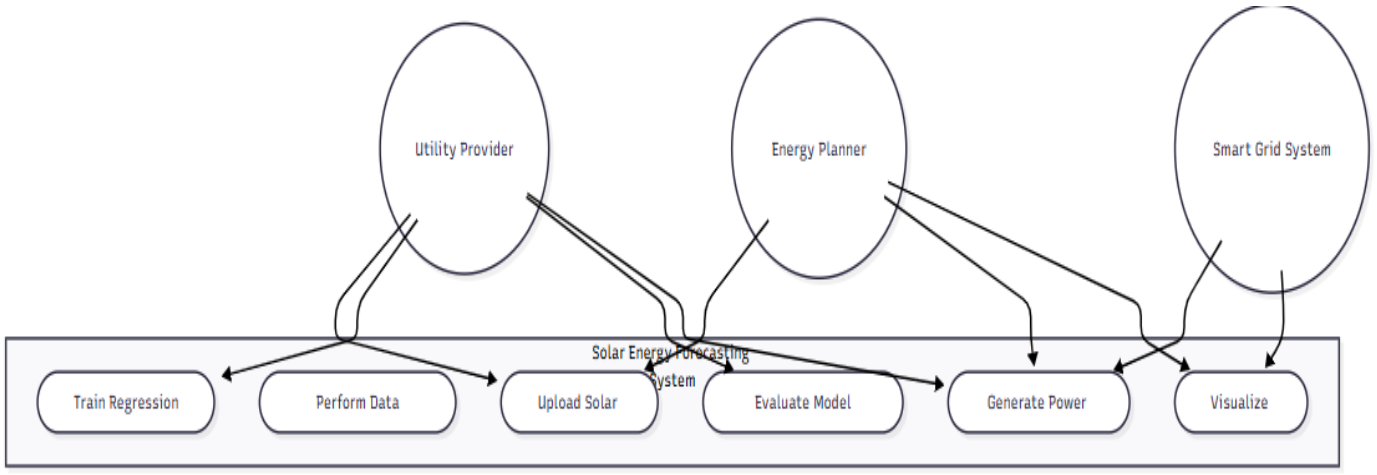


Fig.1. Use Case Diagram for Solar Energy Prediction

Class Diagram:

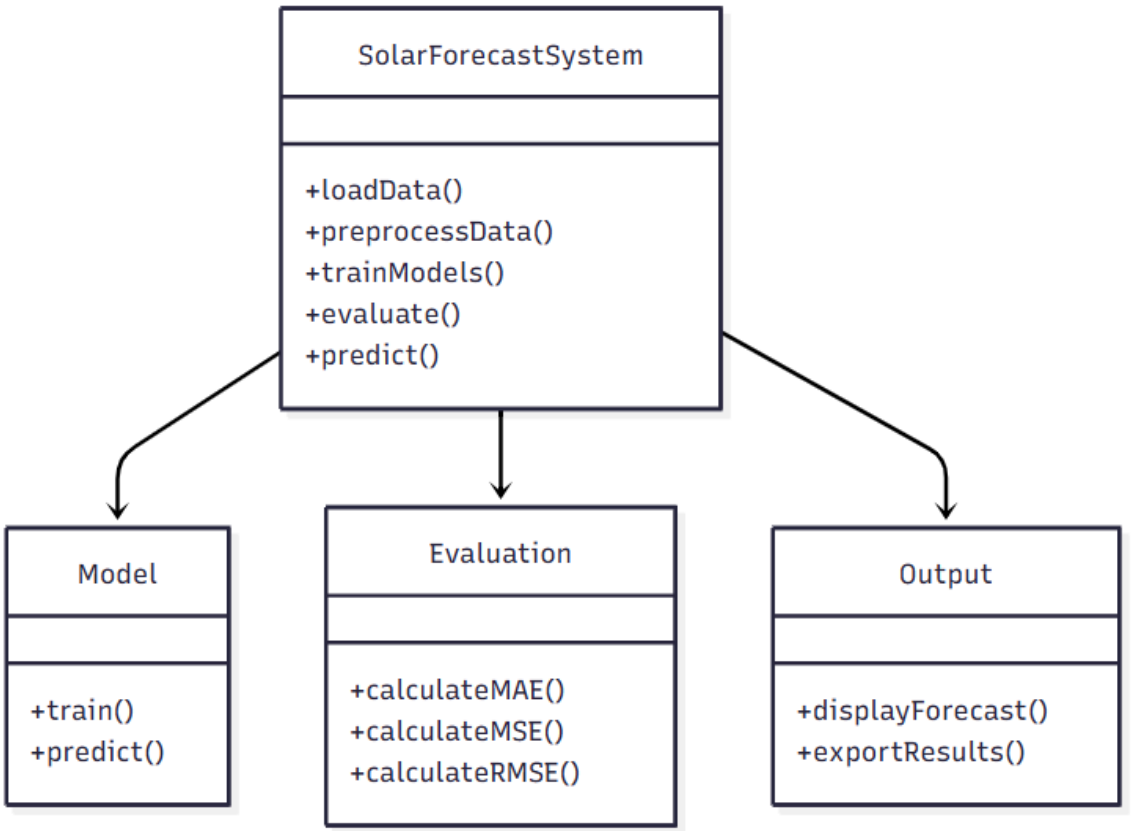


Fig.2. Class Diagram for Solar Energy Prediction

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

Sequence Diagram:

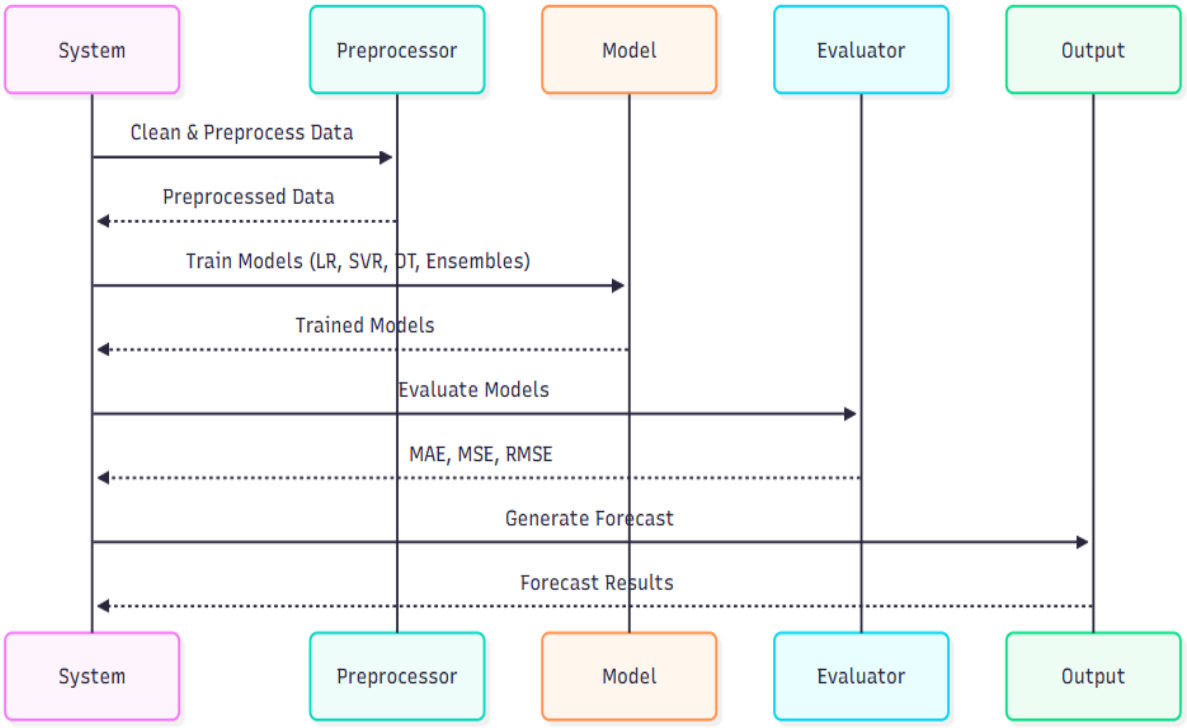


Fig.3. Sequence Diagram for Solar Energy Prediction

Component Diagram:

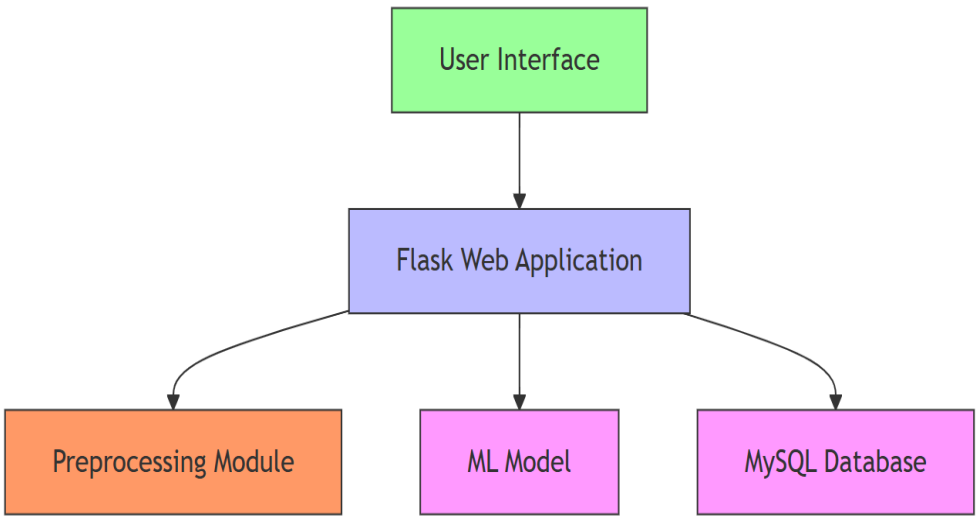


Fig.3. Sequence Diagram for Solar Energy Prediction

5. ER Diagram:

An Entity–relationship model (ER model) describes the structure of a database with the help of a diagram, which is known as Entity Relationship Diagram (ER Diagram). An ER model is a design or blueprint of a database that can later be implemented as a database. The main components of E-R model are entity set and relationship set.An ER diagram shows the relationship among entity sets. An entity set is a group of similar entities, and these entities can have attributes. In terms of DBMS, an entity is a table or attribute of a table in database, so by showing relationship among tables and their attributes, ER diagram shows the complete logical structure of a database.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

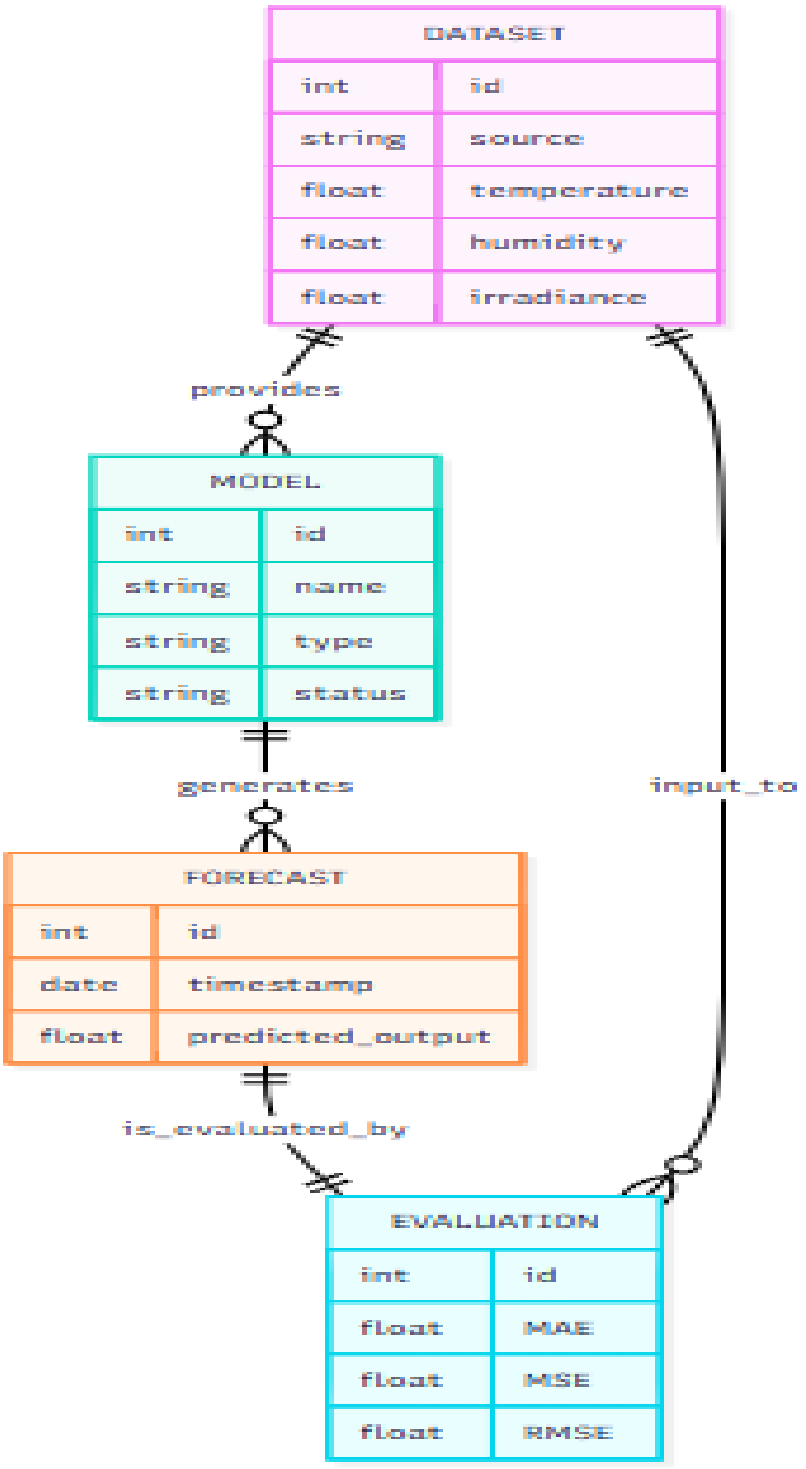


Fig.4. ER Diagram for Solar Energy Prediction

6. DFD Diagram:

A Data Flow Diagram (DFD) is a traditional way to visualize the information flows within a system. A neat and clear DFD can depict a good amount of the system requirements graphically. It can be manual, automated, or a combination of both. It shows how information enters and leaves the system, what changes the information and where information is stored. The purpose of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communications tool between a systems analyst and any person who plays a part in the system that acts as the starting point for redesigning a system.In the proposed **Smart Solar Energy Forecasting System**, the Data Flow Diagram (DFD) illustrates the movement of data from the input stage to the final prediction output. It shows how historical solar energy data and meteorological parameters such as solar irradiance, temperature, humidity, wind speed, and atmospheric conditions are collected, preprocessed, and

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

passed through machine learning models including **Stacking Regression** and **Voting Regression**. The generated forecasting results are then evaluated, visualized, and stored for future analysis. The DFD provides a clear representation of data processing, system interactions, and information flow, making it easier to understand the overall functionality and operation of the proposed forecasting framework.

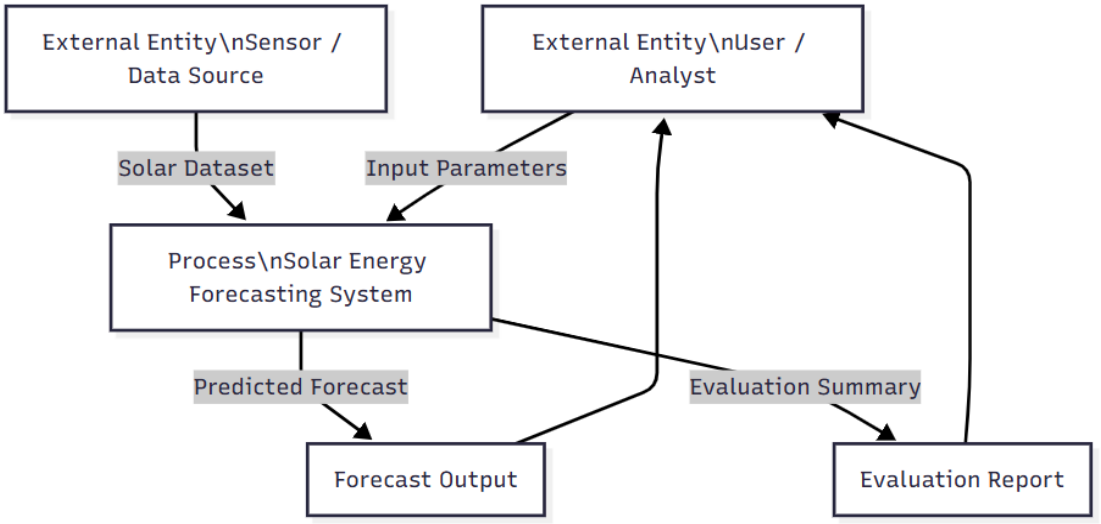


Fig.5. DFD Diagram for Solar Energy Prediction

Implementation And Results

7.1. Modules Implementation

The implementation phase focuses on developing and integrating the various modules of the proposed **Smart Solar Energy Forecasting System** using **Stacking Regression** and **Voting Regression** techniques. The entire system is implemented using the Python programming language along with machine learning libraries such as **Pandas, NumPy, Scikit-learn, Matplotlib, and SHAP**. The implementation consists of multiple interconnected modules, including data collection, data preprocessing, model training, prediction, evaluation, model storage, and visualization. Each module plays an important role in transforming raw meteorological data into accurate solar energy forecasts while ensuring efficiency, scalability, and reliability.

1) Data Collection: The first module involves collecting historical solar power generation data and meteorological information required for forecasting. The proposed system utilizes a real-world solar power generation dataset obtained from **Kaggle**, which contains important environmental attributes such as **ambient temperature, module temperature, solar irradiance, humidity, wind speed, AC power, and DC power output**. The dataset is stored in **CSV format** and imported into the Python environment using Pandas for further analysis and processing. This module provides the foundation for developing accurate machine learning models by supplying high-quality historical data.

7.2. Data Preprocessing and Splitting

The collected dataset undergoes several preprocessing operations before it is used for model training. Initially, missing values and duplicate records are identified and handled using appropriate imputation techniques, while noisy and inconsistent data are removed to improve data quality. Feature engineering techniques are then applied to generate additional attributes such as power efficiency ratio and time-based features including hour, day, and month. The numerical features are normalized using suitable scaling techniques to improve model performance. Finally, the processed dataset is divided into **80% training data** and **20% testing data**, allowing the forecasting models to be trained and evaluated effectively under different weather conditions.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue X October 2026

7.3 Model Training

After preprocessing, multiple machine learning regression algorithms are trained using the prepared dataset to predict future solar power generation. Initially, baseline models such as **Linear Regression, Support Vector Regression (SVR), Decision Tree Regression, and Random Forest Regression** are developed. Advanced ensemble learning models including **Voting Regression** and **Stacking Regression** are then implemented to improve forecasting accuracy by combining the strengths of multiple regression algorithms. The models are trained using **Scikit-learn**, while hyperparameter optimization is performed through **Grid Search Cross Validation** to obtain the best-performing forecasting model with improved generalization capability.

7.4 Model Testing and Evaluation

Once the models have been trained successfully, they are evaluated using the testing dataset to measure forecasting performance. The predicted solar power values are compared with the actual energy generation values using standard regression evaluation metrics such as **Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score**. Experimental results indicate that the ensemble learning models consistently outperform individual regression algorithms by producing lower prediction errors and higher forecasting accuracy. Among all the evaluated models, the **Stacking Regressor** achieves the highest prediction performance, while the **Voting Regressor** also demonstrates excellent forecasting capability and model stability.

7.5 Model Saving

After identifying the best-performing forecasting model, the trained model is serialized and stored using the **Joblib** library in **PKL format**. Saving the trained model eliminates the need for repeated training whenever new predictions are required, thereby reducing computational time and improving system efficiency. The saved model can be easily loaded during deployment for future forecasting tasks without affecting prediction accuracy.

7.6 Prediction Module

The prediction module is responsible for generating solar energy forecasts using the trained ensemble learning model. The module accepts new meteorological input parameters such as **solar irradiance, temperature, humidity, wind speed, atmospheric pressure, and module temperature**, processes the data using the same preprocessing pipeline, and produces accurate solar power predictions. The framework supports both **batch prediction**, where multiple records are processed simultaneously, and **real-time prediction**, where a single input instance is used to estimate future solar energy generation. The generated predictions can be stored for future analysis and decision-making.

7.7. Visualization and Explainability

The final implementation module focuses on presenting forecasting results through meaningful visualizations and explainable artificial intelligence techniques. Visualization libraries such as **Matplotlib** and **Plotly** are used to generate graphs including **Predicted vs. Actual Solar Power**, **Regression Error Distribution**, **Correlation Heatmaps**, and **Model Performance Comparison Charts**. Furthermore, **SHAP (SHapley Additive exPlanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)** are integrated to explain the influence of each meteorological parameter on the forecasting results. These explainability techniques improve model transparency, enhance user confidence, and enable researchers, utility providers, and energy planners to understand the reasoning behind each prediction, making the proposed framework suitable for practical renewable energy management applications.

System Study And Testing

System Study: System study involves analyzing the existing system and identifying its limitations, requirements, and opportunities for improvement. It provides a clear understanding of user needs and helps in designing an efficient and intelligent solar energy forecasting system. The proposed system utilizes advanced

ensemble machine learning techniques to improve forecasting accuracy while supporting reliable renewable energy management and smart grid operations.

8.1.1 Introduction: The proposed system aims to accurately forecast solar energy generation using a combination of traditional regression algorithms and advanced ensemble learning techniques, specifically Stacking Regression and Voting Regression. The objective is to address the challenges associated with solar energy forecasting, including fluctuating weather conditions, nonlinear relationships among meteorological parameters, and prediction uncertainty. The system processes historical solar power generation data together with environmental factors such as solar irradiance, ambient temperature, humidity, wind speed, and module temperature to generate highly accurate short-term forecasts. The proposed framework performs end-to-end data preprocessing, feature engineering, model training, evaluation, and prediction while integrating Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME to provide transparent and interpretable forecasting results.

8.1.2 Functional Requirements

The proposed solar energy forecasting system provides several functional capabilities. It performs data preprocessing by handling missing values, removing duplicate records, normalizing numerical attributes, and preparing the dataset for machine learning applications. Feature engineering techniques are applied to extract meaningful information from meteorological data to improve forecasting performance. The system implements multiple regression algorithms, including Linear Regression, Decision Tree Regression, Support Vector Regression (SVR), Random Forest Regression, Voting Regression, and Stacking Regression, for accurate solar energy prediction. It evaluates model performance using standard regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R^2 Score. The framework also integrates SHAP and LIME to provide interpretable explanations for prediction results and supports exporting the trained forecasting model for future deployment in renewable energy management applications.

1. Unit Testing

Test Case	Input	Expected Output	Actual Output	Result
Encode categorical variables	Raw CSV	Encoded format	Success	Pass
Model. Fit () with LSTM	Preprocessed data	Trained model	Success	Pass
Feature Correlation Matrix	Cleaned data	Heatmap and dropped features	Correct features removed	Pass

2. Integration Testing

Scenario	Components Involved	Outcome	Result
End-to-end training	Preprocessor → Auto-GRU → Evaluation	Successful, accurate prediction	Pass
LIME interpretability	Trained Model →LIME Explainer	Interpretations generated	Pass

CONCLUSION

This study presents a scalable and interpretable machine learning-based framework for accurate solar energy production forecasting. By leveraging a real-world solar power generation dataset, the project effectively addresses challenges such as environmental variability, nonlinear feature interactions, and prediction uncertainty in renewable energy systems. Through comprehensive experimentation, ensemble models—particularly the Stacking Regressor and Voting Regressor—consistently outperformed traditional regression algorithms like Linear Regression, Support Vector Regression, and Decision Trees in terms of prediction accuracy and generalization. The integration of multiple learners enabled the system to capture diverse patterns within the meteorological data, while evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) confirmed the improved performance of ensemble methods.

Additionally, explainable AI techniques like SHAP and LIME were incorporated to provide transparency, offering clear insights into how features such as solar irradiance, temperature, and humidity influence power output predictions. Overall, the proposed framework offers a reliable, data-driven solution for solar energy forecasting. It has the potential to support smart grid operations, enhance renewable energy planning, reduce power supply uncertainty, and assist policymakers and energy providers in making informed, data-backed decisions that contribute to a more sustainable energy infrastructure. Furthermore, the proposed forecasting framework demonstrates the practical applicability of ensemble machine learning techniques in modern renewable energy systems. Its modular and scalable architecture allows seamless integration with real-time IoT sensors, cloud computing platforms, and smart energy management systems. The ability to generate accurate short-term solar power forecasts can significantly improve energy scheduling, battery storage optimization, load balancing, and efficient utilization of renewable resources in both residential and industrial applications.

In future, the proposed system can be enhanced by incorporating advanced deep learning models such as **Long Short-Term Memory (LSTM)**, **Gated Recurrent Unit (GRU)**, and hybrid ensemble architectures to further improve forecasting accuracy under highly dynamic weather conditions. The integration of real-time weather APIs, satellite imagery, and multi-location solar datasets can also increase the robustness and scalability of the framework. These enhancements will contribute to the development of intelligent, sustainable, and data-driven energy management systems that support the global transition toward clean and renewable energy sources. The outcomes of this research demonstrate that intelligent ensemble learning approaches can play a significant role in improving the efficiency and reliability of renewable energy forecasting systems. By combining accurate prediction techniques with explainable artificial intelligence, the proposed framework not only enhances forecasting performance but also increases user confidence through transparent and interpretable results. This research provides a strong foundation for future advancements in smart energy technologies and supports the widespread adoption of sustainable solar power generation, contributing to energy security, environmental conservation, and the development of resilient smart grid infrastructures.

Future Enhancement

The proposed Smart Solar Energy Forecasting system provides a strong foundation for intelligent renewable energy prediction; however, there are several opportunities for further improvement. Future research can expand the existing dataset by incorporating solar power generation data collected from multiple geographical locations and climatic regions. This will improve the model's ability to generalize across different environmental conditions and increase its forecasting accuracy for large-scale renewable energy applications. Furthermore, integrating additional meteorological parameters and satellite-based weather information can enhance the robustness of the forecasting framework. Another important enhancement is the incorporation of advanced deep learning architectures such as Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Convolutional Neural Networks (CNN), and Gated Recurrent Units (GRU). These models are capable of learning complex temporal dependencies and nonlinear relationships present in time-series solar energy data. Hybrid learning frameworks that combine deep learning models with ensemble learning techniques, such as LSTM-Stacking or CNN-Voting Regression, can further improve prediction accuracy, reduce forecasting uncertainty, and provide more reliable results under rapidly changing weather conditions. The proposed system can also be enhanced by integrating Internet of Things (IoT) technology for real-time solar energy monitoring and forecasting. By collecting live meteorological and photovoltaic sensor data, the framework can generate instant predictions and alerts for energy providers and grid operators. Additionally, deploying the forecasting system on cloud computing platforms such as Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure would enable scalable real-time prediction services, remote accessibility, and efficient management of large-scale solar energy datasets through interactive dashboards.

Future work may also focus on improving model transparency by incorporating additional Explainable Artificial Intelligence (XAI) techniques such as Deep. **DeepLIFT**, **Integrated Gradients**, and other feature attribution methods alongside **SHAP** and **LIME**. These techniques will provide deeper insights into the contribution of each meteorological parameter toward forecasting results, increasing user confidence and supporting better decision-making. Moreover, integrating the forecasting framework with **Energy Management Systems (EMS)**, **smart grid controllers**, and **battery energy storage systems** will enable automated energy scheduling,

intelligent charge-discharge optimization, demand-response management, and efficient utilization of renewable resources. These advancements will contribute significantly to the development of sustainable, intelligent, and resilient solar energy management systems for future smart cities and renewable energy infrastructures. Another promising direction for future research is the integration of **weather forecasting APIs**, **satellite imagery**, and **Artificial Intelligence-based climate prediction models** to improve the accuracy of solar energy forecasting under rapidly changing atmospheric conditions. Incorporating these advanced data sources will enable the system to generate more precise long-term and short-term forecasts while adapting to seasonal and climatic variations. Furthermore, developing a user-friendly web or mobile application for monitoring and visualizing real-time solar energy predictions will enhance accessibility for utility providers, researchers, policymakers, and consumers, promoting the widespread adoption of intelligent renewable energy forecasting systems.

REFERENCES

1. A. Sharma, P. Tiwari, and R. Debnath, “Solar power prediction using machine learning: A comprehensive review,” Renewable and Sustainable Energy Reviews, vol. 166, Art. no. 112620, 2022.
2. N. Aggarwal and R. Gupta, “Ensemble learning for renewable energy forecasting,” Renewable Energy, vol. 202, pp. 489–499, 2023.
3. L. Chen et al., “Explainable AI for Renewable Energy: A Case Study on Solar Generation Forecasting,” Energy Informatics, vol. 6, no. 1, pp. 1–16, 2023.
4. S. Wei and J. Zhang, “A Lightweight Voting Regressor for Edge-Based Solar Energy Prediction,” Sensors, vol. 23, no. 3, pp. 1156–1172, 2023.
5. G. Banerjee and S. Sharma, “Time-Series Forecasting of Solar Power Output Using Hybrid LSTM and SVR Models,” IEEE Access, vol. 12, pp. 11234–11245, 2023.
6. T. Yamamoto and H. Zhou, “Comparison of Machine Learning Algorithms for Solar Irradiance Forecasting,” Solar Energy, vol. 195, pp. 247–260, 2023.
7. A. Mohanty and D. Patra, “Prediction of solar energy output using deep neural networks,” Renewable Energy, vol. 162, pp. 1165–1173, 2020.
8. L. Breeman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
9. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774, 2017.