CHAPTER 1

INTRODUCTION

1.1 GENERAL

Renewable energy forecasting has received a lot of attention recently, as the use of renewable energy from solar and wind grows. Other kinds of renewable energy may require some forecasting duties, but solar and wind have received the most attention during the previous decade due to their low predictability. Many scholars are working to construct models for predicting long-term daily or monthly average solar radiation and wind speed. Solar radiation and wind velocity models for any region are particularly useful in determining the power generation of a certain solar and wind power plant.

1.2 NEED FOR THE STUDY

The study is necessary because the fluctuation and uncertainty of renewable energy sources provide inherent issues. Renewable energy sources (RES) such as solar and wind are greatly influenced by environmental factors such as weather, resulting in very unpredictable power output. This unpredictability hinders their integration into the electricity system, necessitating the creation of precise forecasting models to enable effective grid management and minimal disturbances. Traditional forecasting models have proven inadequate for handling the nonlinear and high-dimensional nature of renewable energy data. In response, machine learning (ML) and deep learning (DL) algorithms offer promising solutions by learning from large datasets and capturing complex relationships in the data. However, choosing the right ML or DL model, selecting appropriate input variables, and addressing issues such as missing data are critical for improving the precision and reliability of renewable energy forecasts. Thus, the study aims to explore and develop robust forecasting models, particularly focusing on the SARIMA model for renewable energy generation. It seeks to bridge the gap left by traditional models, offering more accurate and interpretable predictions for wind and solar energy output. This is crucial for optimizing the use of renewable energy, reducing dependence on non-renewable sources, and contributing to a sustainable energy future.

1.3 OVERVIEW OF THE PROJECT

The demand for this initiative stems from the growing reliance on renewable energy sources like wind and solar, which are crucial for lowering greenhouse gas emissions and combatting climate change. While renewable energy has many advantages, including lowering reliance on fossil fuels and supporting sustainability, its inherent fluctuation presents a considerable challenge to energy system management. Solar power is impacted by cloud cover and seasonal variations in sunshine, whilst wind energy is dependent on changeable weather patterns, making both types of energy unpredictable. To efficiently incorporate these renewable energy sources into the system and maintain a consistent power supply, precise forecasting models are required. Traditional statistical models are limited in their ability to cope with renewable energy data, which is complicated, nonlinear, and multidimensional. As a result, this study focusses on the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, which is ideal for time series forecasting and can capture both seasonal patterns and trends in renewable energy generation data.

1.4 OBJECTIVES OF THE STUDY

The goal of this project is to employ appropriate machine learning methodologies to develop a predictive model that can forecast solar and wind energy generation based on historic data and forecasts of the weather.

The main objectives are to:

- **Boost prediction accuracy**: By employing sophisticated machine learning models, the project aims to improve the accuracy of forecasts for the production of renewable energy while taking seasonal and environmental fluctuations.

- **Assure power supply dependability**: By more effectively managing the integration of solar and wind power into the grid, reducing fluctuations, and preventing disruptions, accurate forecasts will aid in stabilising the energy supply.
- **Improve energy generation efficiency**: By maximising the use of renewable energy sources, the model will lower waste, increase total energy generation efficiency, and promote sustainable energy development.

CHAPTER 2

REVIEW OF LITERATURE

2.1 INTRODUCTION

The review of literature delves into the various models and techniques that have been explored for forecasting renewable energy generation, particularly solar and wind energy. As the transition toward renewable energy sources intensifies to combat climate change and reduce dependency on fossil fuels, the accurate prediction of energy output from these sources has become increasingly important. Solar and wind energy generation, while environmentally friendly, are inherently variable due to changing weather conditions, cloud cover, wind patterns, and seasonal variations. These fluctuations create significant challenges for integrating renewable energy into the power grid reliably.

To address these challenges, researchers have applied various predictive models, ranging from traditional linear regression and decision tree models to more advanced machine learning techniques such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks. While linear regression and decision trees provide simpler and interpretable models, they often lack the sophistication to fully capture the complex and nonlinear patterns inherent in renewable energy data. Machine learning models like SVM and ANN, as well as deep learning architectures like LSTM, have shown greater potential in improving prediction accuracy due to their ability to learn complex relationships from large datasets.

2.1 LITERATURE REVIEW.

	Author Name	**Paper Title**	**Description**	**Journal**	**Volume/Year**
1	Luisa Fernanda Jimenez Alvarez	Renewable Energy Prediction through Machine Learning Algorithms	This paper focuses on selecting a suitable ML algorithm among LR, RF, Multilayer perceptron, SVM using MSE and MAE	IEEE	2020
2	Deepa Somasundaram	Machine learning application for predicting system production in renewable energy	Algorithms like LR,RF, Decision Tree and SVM are used to forecast that is evaluated by MSE,MAE and R squared metrics is discussed in the paper.	IJPEDS	2024
3	N.M.S. Hassan	Forecasting Renewable energy Generation with ML and DL	This paper employs several suitable model for different prediction and gives the accurate prediction value.	Energies	2023
5	Vishnu Vardhan	Renewable Energy and Demand Forecasting in an Integrated Smart Grid	This paper presents a framework that realistically simulate a microgrid and forecasts renewable energy and load demand using LSTM.	IEEE	2021

The table in the literature review presents a comparative analysis of various machine learning models used for forecasting renewable energy generation, specifically solar and wind power. It outlines models such as Linear Regression, Support Vector Machines (SVM), Decision Trees, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks, each of which has been employed to predict energy generation based on a variety of input parameters. These parameters typically include historical energy generation data, weather conditions (such as wind speed, temperature, humidity, and solar irradiance), and environmental factors. The performance of each model is evaluated using metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and accuracy, which provide insights into the models' effectiveness in handling the inherent variability of renewable energy data. Observations from these studies highlight the strengths and limitations of each algorithm, with advanced models like ANN and LSTM demonstrating higher accuracy in capturing nonlinear patterns in the data. The table serves as a comprehensive summary of past research, offering a clear comparison of different approaches used for energy forecasting in various studies.

This chapter reviews the methodologies employed in recent studies that utilize these models, discussing their strengths and limitations. The exploration includes the use of both single-model approaches and hybrid models that combine multiple techniques to further enhance forecasting performance. By surveying the application of linear regression, SVM, decision trees, ANN, and LSTM models, this chapter aims to provide a comprehensive understanding of the current advancements and challenges in renewable energy forecasting, while also identifying potential avenues for future research in the field.

CHAPTER 3

SYSTEM OVERVIEW

3.1 EXISTING SYSTEM

Regression: Regression is a supervised learning approach that forecasts a continuous output variable based on one or more input variables. Regression aims to identify a mathematical function that can correlate the input variables to a continuous output variable, which may represent a single value or a range of values. Linear regression, Polynomial Regression, and Support Vector Regression (SVR) are the three main types of supervised learning algorithms in regression.

Linear and Polynomial Regression: Linear regression is a prevalent and straight forward approach used to forecast a continuous output variable utilizing one or multiple input variables. It uses a straight line to indicate the correlation between the input variables and the output variables [51]. On the other hand, Polynomial regression, a type of linear regression, employs nth-degree polynomial functions to depict the connection between input features and the outcome variable [52]. This can enhance the accuracy of predictions by enabling the model to capture more intricate correlations between the input data and the target variable. In renewable energy forecasting, both linear and polynomial regression can be used to predict the power output of RES such as solar and wind power. Weather information like temperature, humidity, and wind speed are frequently included in the input characteristics, along with historical power output data. The target variable is the power output of the renewable energy source, which can be predicted using the input features.

Support Vector Regression (SVR): The SVR algorithm is used in machine learning regression analysis [58]. In a high-dimensional space, it finds the optimal hyperplane to divide the data points. The hyperplane is chosen with the goal of maximising the separation between the nearest data points on either side of it. The strategy entails minimising the difference between the expected and actual values while restricting the margin. Additionally, it is an effective model for forecasting the potential for renewable energy in a given area. Yuan et al. (2022), for instance, suggested a Jellyfish Search algorithm optimisation SVR (IJS-SVR) model to forecast wind power output and handle problems with power dispatching and grid connectivity. The IJS approach was used to optimise the SVR, and the both spring and winter tests were conducted on the model. In both seasons, IJS-SVR performed better than other models, offering a practical and affordable approach to wind power forecasting [59]. Furthermore, Li et al. (2022) developed machine learning (ML)-based methods that use SVM regression and the Hidden Markov Model for short-term solar irradiance prediction. In different weather situations, the Bureau of Meteorology showed that their algorithms can accurately predict sun irradiance at intervals of 5 to 30 minutes. Random Forest Regression (RFR) and SVR models were created by another author, Mwende et al. (2022), for projecting photovoltaic (PV) power output in real time.

Decision Trees: An alternative method for classification is decision trees, which involves dividing the input space into smaller sections based on input variable values and then assigning a label or value to each of these sections [64]. The different studies developed Decision Tree models to forecast power output from different renewable energy systems. Essama et al. (2018) developed a models to predict the power output of a photovoltaic (PV) system in Cocoa, Florida of USA using weather parameters obtained from United States' National Renewable Energy Laboratory (NREL). By selecting the best performance among the ANN, RF, DT, extreme gradient boosting (XGB), and LSTM algorithms, they aim to fill a research gap in the area. They have come to the conclusion that even if all of the algorithms were good, ANN is the most accurate method for forecasting PV solar power generation.

Random Forest: An effective and reliable prediction is produced by the supervised ML method known as random forest, which creates several decision trees and merges them [66]. The bagging technique, which is employed by random forest, reduces the variance of the base algorithms. This technique is particularly useful for forecasting time series data [67]. Random forest mitigates correlation between trees by introducing randomization in two ways: sampling from the training set and selecting from the feature subset. The RF model creates a complete binary tree for each of the N trees in isolation, which enables parallel processing.

Support Vector Machines (SVM): SVM are a type of classification algorithm that identifies a hyperplane which maximizes the margin between the hyperplane and the data points, akin to SVR [71,72]. SVM has been utilized in renewable energy forecasting to estimate the power output of wind and solar farms by incorporating input features such as historical power output, weather data, and time of day. For instance, Zeng et al. (2022) propose a 2D least-square SVM (LS-SVM) model for short-term solar power prediction.

3.2 PROPOSED SYSTEM

The proposed system focuses on developing a predictive model for forecasting the generation of solar and wind energy using machine learning algorithms. Given the intermittent nature of renewable energy sources, accurately predicting energy generation is essential for ensuring a reliable power supply and efficient integration into the power grid. The system leverages historical energy generation data, along with weather forecasts and environmental factors, such as temperature, humidity, wind speed, and solar irradiance, to build a robust forecasting model.

To address the nonlinear and seasonal variations in renewable energy data, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is selected for time-series forecasting. SARIMA is particularly well-suited for this purpose because it accounts for seasonality in the data while reducing overfitting, thereby enhancing the accuracy and reliability of predictions. Additionally, the model's ability to handle temporal dependencies makes it an ideal choice for renewable energy forecasting, where past energy patterns and seasonal effects are strong indicators of future behaviour.

The system workflow begins with data collection, where historical energy data and relevant meteorological factors are gathered. Following this, data preprocessing steps, such as handling missing values and scaling, are performed to ensure the model receives clean and consistent input. Afterward, the SARIMA model is trained on the pre-processed dataset, allowing it to learn from past energy generation patterns and weather trends. The model is then tested and validated to ensure its predictive accuracy.

The primary objective of the proposed system is to improve the accuracy of energy generation forecasts, ensuring better grid stability and resource planning. By integrating this model into the energy management system, utility providers can better anticipate periods of high and low energy generation, optimizing energy distribution and minimizing the impact of renewable energy variability. Furthermore, the proposed system also paves the way for future enhancements, such as incorporating hybrid models or deep learning techniques to further improve forecasting performance.

3.3 FEASABILITY STUDY

Technical Feasibility

The project is technically feasible as it leverages well-established machine learning algorithms, particularly the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, which is recognized for its effectiveness in handling time-series data with seasonal patterns. The system can integrate historical energy data, weather forecasts, and environmental conditions, such as temperature, wind speed, and solar irradiance, which are readily available through various public and private sources.

The availability of these data inputs, combined with existing computational resources like cloud platforms and local processing power, makes the system's development achievable. Additionally, machine learning libraries like Python’s statsmodels for

SARIMA, sklearn, and TensorFlow for future potential hybrid or deep learning models, provide robust support for the model’s development and implementation.

Economic Feasibility

From an economic perspective, the project is cost-effective, as the primary costs are associated with data acquisition, computational resources, and model development. Since historical weather and energy data are often freely available, the primary expenses would be related to infrastructure setup (such as cloud computing services if needed), model training, and operational monitoring. The potential return on investment (ROI) is significant, as improving the accuracy of renewable energy predictions can lead to optimized energy distribution, reduced operational costs, and improved resource allocation. These advantages can make the project financially viable, particularly for energy utility companies and grid operators.

Operational Feasibility

Implementing the predictive system into an operational energy management framework is feasible with current technology. The model's ability to forecast solar and wind energy generation in real-time can directly benefit utility operators by providing actionable insights into expected energy production. The system can be integrated into existing software platforms used by energy providers, enhancing decision-making processes regarding energy distribution and storage. The system's maintenance and updates would require minimal staffing, focusing on ensuring data quality, retraining the model with updated information, and monitoring performance. The scalability of the project is also viable, as the system can be extended to other geographical locations with minimal adjustments.

CHAPTER 4

SYSTEM REQUIREMENTS

The System Requirements section outlines the necessary hardware and software components needed to implement the predictive model for forecasting renewable energy generation. These requirements ensure the smooth functioning of the system, which processes large datasets, applies machine learning models, and provides accurate predictions.

4.1 HARDWARE REQUIREMENTS

To support the computational tasks required for processing and modelling large datasets, the hardware specifications should meet the following criteria:

Processor: A multi-core processor (Intel i5/i7 or AMD Ryzen) to handle parallel processing efficiently.

RAM: A minimum of 16 GB of RAM is required for fast data processing and model training, with 32 GB recommended for more intensive tasks, especially when handling large datasets.

Storage:

- **Solid-State Drive (SSD)** of at least 512 GB for fast read/write operations during data manipulation and model execution.
- Additional storage (HDD or cloud-based) may be required to store large datasets, particularly historical weather data and energy generation records.

Network Connectivity: High-speed internet connection for accessing online data sources, cloud computing services, and weather forecasting APIs.

4.2 SOFTWARE REQUIREMENTS

The software environment should support the development and execution of machine learning models, data preprocessing, and system integration. The following software components are necessary:

Operating System: Windows 10/11, macOS, or Linux-based distributions (e.g., Ubuntu) to ensure compatibility with the necessary libraries and tools.

Programming Languages:

- **Python** (version 3.7 or above) as the primary language for machine learning development due to its extensive libraries for data science and time-series analysis.

Development Environment:

- **Jupyter Notebook or PyCharm** for code development, testing, and execution.

Libraries and Frameworks:

- **NumPy** and **Pandas** for data manipulation and preprocessing.
- **Matplotlib** or **Seaborn** for data visualization.
- **Statsmodels** for implementing SARIMA models for time series forecasting.
- **Scikit-learn** for machine learning model implementation, including regression, SVM, decision trees, and ensemble methods.
- **TensorFlow** or **Keras** (optional) for implementing deep learning models, if needed.

Data Sources and APIs:

- **Weather API** (e.g., OpenWeatherMap or WeatherAPI) for real-time and historical weather data integration.
- **Energy Data API** (optional) for fetching energy generation data from public datasets or government repositories.

CHAPTER 5

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

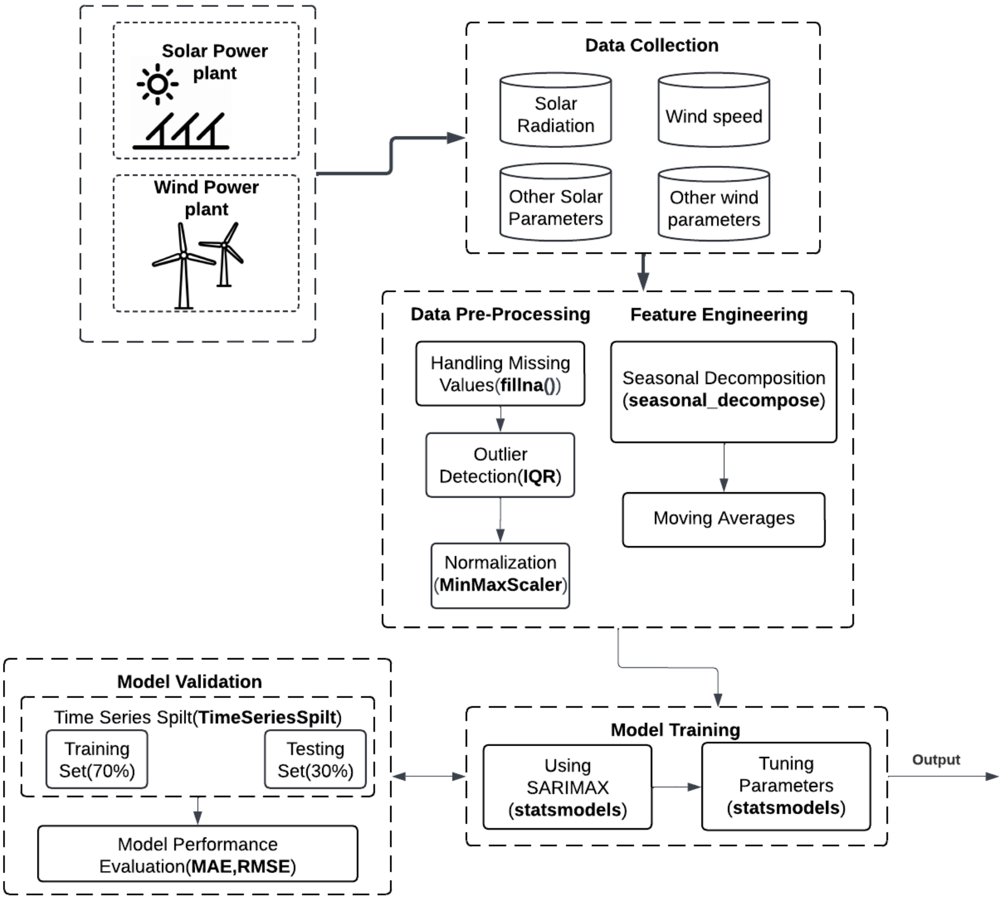


Fig 5.1 System Architecture

5.2 MODULE DESCRIPTION

The project consists of several key modules starting with Data Collection, where historical data on solar and wind energy, along with environmental parameters like temperature, humidity, wind speed, and solar irradiance, are gathered from reliable sources. This is followed by Data Preprocessing, where the data is cleaned by handling missing values, normalizing, and removing outliers to ensure it is consistent and ready for analysis. In Feature Engineering, relevant features are derived to enhance model performance, such as time-based indicators and weather variables. The next module, Model Training, involves training various machine learning algorithms like Linear Regression, SVM, Decision Trees, ANN, and LSTM to predict energy generation based on historical data. Finally, in Model Validation, the models are evaluated using metrics like Mean Squared Error (MSE) and R-squared (R^2) to ensure accurate and reliable predictions, selecting the best model for deployment based on performance.

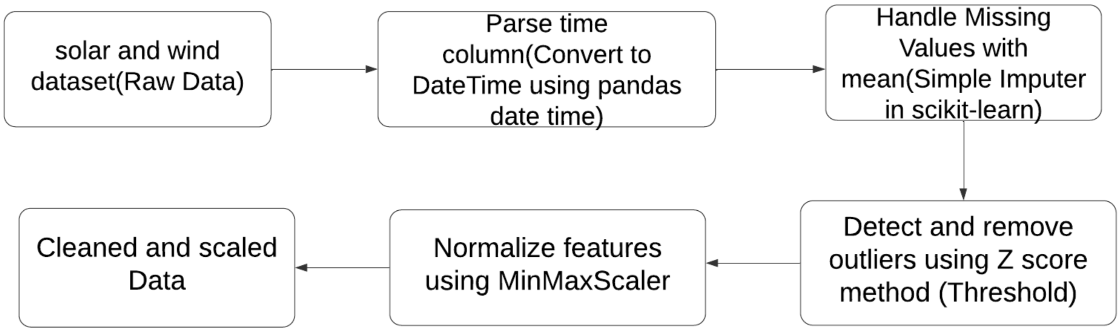
List of Modules:-

Data Collection
Data Preprocessing
Feature Engineering
Model Training
Model Validation

5.2.1 DATA COLLECTION

The Data Collection module plays a critical role in the project as it forms the foundation for accurate energy forecasting. This module involves gathering relevant and high-quality datasets related to renewable energy generation, specifically focusing on solar and wind energy. The data required for building a robust predictive model includes historical energy generation data, weather conditions, and environmental factors such as temperature, wind speed, solar irradiance, and humidity. Data can be sourced from various reliable platforms, including public energy repositories, meteorological services, governmental energy agencies, and weather forecasting APIs.

Additionally, datasets from local solar and wind farms, or energy consumption logs, might also be included. The time-series nature of the data, especially for renewable energy, means that seasonal variations and trends are particularly important for accurate forecasting, making the comprehensive collection of time-bound data essential for the project's success. Proper data collection ensures that the model has a diverse, relevant, and large enough dataset to learn from and make accurate predictions.



5.2.2 DATA PREPROCESSING

Fig 5.2 Data Preprocessing data flow

The data preprocessing phase of the project is crucial to ensure that the solar and wind datasets are in an optimal state for modelling. First, the time column from both datasets is parsed into a datetime format using Pandas, and the solar and wind data are merged based on this Time index. Next, missing values in key columns such as solar irradiance, wind speed, and power generation (measured in MW) are addressed using a mean imputation strategy via scikit-learn's Simple Imputer. Outliers for important variables like Power (MW), wind speed, and irradiance are detected and removed using the Z-score method with a set threshold to improve data quality. After handling outliers, all the numerical columns, including solar irradiance, wind speed, temperature, and pressure, are normalized using the MinMaxScaler to bring the data into a common range for improved performance in the machine learning model. This process results in a cleaned and scaled dataset, which is then ready for the next steps in feature engineering.

Handling Missing Values:-

SimpleImputer (Mean) is a method in scikit-learn used for handling missing data by filling null values with specific strategies such as mean, median, most frequent, or a constant value. In the "mean" strategy, the missing values in a particular column are replaced with the average of the non-missing values, ensuring the dataset is complete for further analysis or modelling. This approach helps maintain the dataset's integrity while preventing biases that could arise from missing information.

$$\text{Imputed Value} = \frac{\sum_{i=1}^{n} X_i}{n}$$

01/01/19 10:00	4.80		
01/01/19 10:15	8.13		
01/01/19 10:30	15.16	Average	20.00
01/01/19 10:45	16.66		
01/01/19 11:00	19.28		
01/01/19 11:15			
01/01/19 11:30	25.87		
01/01/19 11:45	28.31		
01/01/19 12:00	30.67		
01/01/19 12:15	31.13		

Fig 5.3 Data before filling null values

01/01/19 10:00	4.80		
01/01/19 10:15	8.13		
01/01/19 10:30	15.16	Average	20.00
01/01/19 10:45	16.66		
01/01/19 11:00	19.28		
01/01/19 11:15	20.00		
01/01/19 11:30	25.87		
01/01/19 11:45	28.31		
01/01/19 12:00	30.67		
01/01/19 12:15	31.13		

Fig 5.4 Data after filling null values

Handling Outliers:-

Outlier Detection (Z-score) is a method used to identify extreme values in a dataset, also known as outliers. The Z-score measures how far a data point is from the mean in terms of standard deviations. Typically, data points with a Z-score greater than +3 or less than -3 are considered outliers, as they deviate significantly from the rest of the data. These extreme values are then removed to ensure the dataset is more representative and does not skew the analysis or model.

$$Z = \frac{x - \mu}{\sigma}$$

Normalization:-

MinMaxScaler is a technique used to scale features of the data to a defined range, typically between 0 and 1. This method ensures that all numerical variables are scaled uniformly, making it easier for machine learning models to process the data efficiently. By converting the data to this range, the impact of varying feature scales is minimized, allowing for more accurate predictions and analyses.

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

5.2.3 FEATURE ENGINEERING

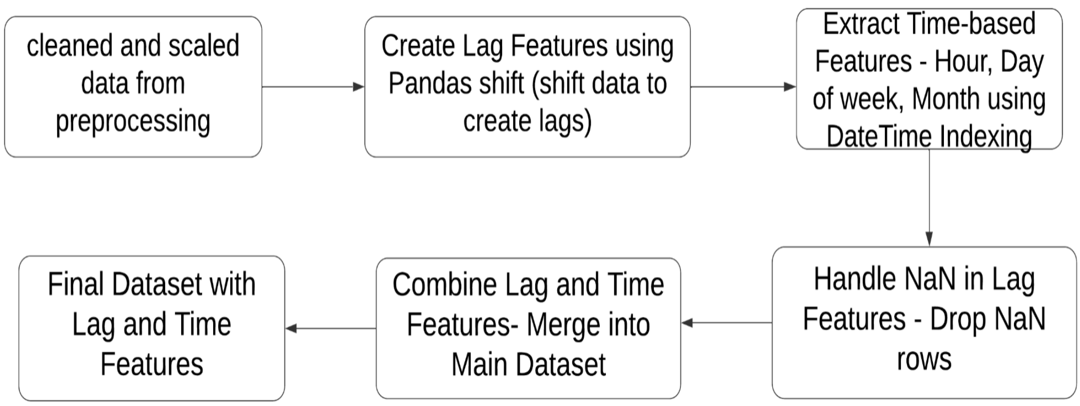


Fig 5.5 Feature engineering data flow

In the feature engineering process for the renewable energy prediction project, the first step involves creating lag features for key variables such as Power (MW), Total solar irradiance, Wind speed, and Air temperature to capture temporal dependencies. This helps in understanding how previous values of these variables influence future predictions, crucial for time series forecasting. Next, additional time-based features like hour, day of the week, and month are extracted from the Time column to account for seasonality and time-based trends, which are critical in capturing patterns like daily and seasonal energy fluctuations.

After creating these features, rows with NaN values, which arise during lag feature creation, are dropped to maintain data integrity. All the lag and time-based features are then combined into the main dataset, ensuring that variables from both solar and wind data are included for a comprehensive prediction model. Finally, all features are normalized, ensuring consistency and preparation for the model training phase, with no missing values present. This thorough preprocessing step ensures the dataset is optimized for training the machine learning model, contributing to accurate and reliable predictions of renewable energy generation.

Lag Features :-

The **shift()** function in pandas is a powerful tool for creating lag features by shifting the values of a column up or down by a specified number of periods, typically used in time series forecasting models like SARIMA or ARIMA. It moves data values from previous time steps into the current row, allowing models to capture temporal dependencies and use past observations to predict future values. For example, by applying shift(1) to a column, the values from the previous time step are downshifted to become predictors for the current time step. This is particularly useful in forecasting scenarios where past behaviour influences future outcomes, such as predicting energy consumption based on previous values.

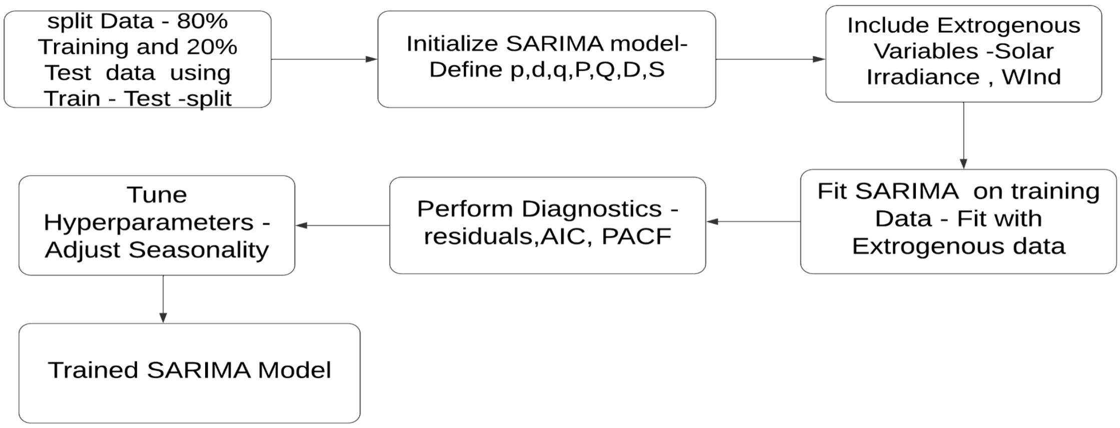
```
            Solar_Irradiance  Solar_Irradiance_Lag1  Solar_Irradiance_Lag2
Date
2024-09-21            1000.0                    NaN                    NaN
2024-09-22            1100.0                 1000.0                    NaN
2024-09-23            1050.0                 1100.0                 1000.0
2024-09-24            1150.0                 1050.0                 1100.0
2024-09-25            1200.0                 1150.0                 1050.0
```

5.5 Lag Features Calculating

Removing NaN:-

When creating lag features in a dataset, missing values (NaN) can be introduced at the beginning of the series because there is no prior data to fill in those lagged values. To remove these NaN values using the **dropna()** method.

5.2.4 MODEL TRAINING



5.7 Model Training Data Flow

Training the SARIMA model involves several systematic steps to make accurate time series predictions. Each stage of the process contributes to preparing, fitting, and tuning the model for optimal results. Here's an explanation of each process in detail:

Differencing: Making the Time Series Stationary:-

To train a SARIMA model, the first step is to ensure the time series data is stationary. A stationary series has constant statistical properties over time, like mean and variance. Many real-world datasets, such as energy consumption or solar energy data, often have trends and seasonality that make them non-stationary.

Regular Differencing (d) is applied to remove trends.

Seasonal Differencing (D) removes repeating seasonal patterns.

$$Y_t' = Y_t - Y_{t-1} \qquad\qquad Y_t' = Y_t - Y_{t-s}$$

Auto-Regressive (AR) Component :-

After differencing, the next step in model training is adding the Auto-Regressive (AR) component. This component models the relationship between the current value and past observations (lags). The p parameter specifies how many lagged observations to consider.The AR component identifies the linear relationship between past values and future values.

Formula for AR:

$$AR(t) = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}$$

Moving Average (MA) Component :-

The Moving Average (MA) component models the error terms or residuals from previous observations. This component improves the model by incorporating past forecast errors. The q parameter defines how many past errors to include in the model.

Formula for MA:

$$MA(t) = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Seasonal Components :-

In addition to the regular AR and MA components, SARIMA includes Seasonal AR and MA terms, which allow the model to capture repeating seasonal patterns in the

data. These are denoted by P (Seasonal AR) and Q (Seasonal MA) parameters, with a seasonal period s (e.g., 12 for monthly data if the season is yearly).
Seasonal AR formula:

$$SAR(t) = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \cdots + \Phi_P Y_{t-Ps}$$

Seasonal MA formula:

$$SMA(t) = \Theta_1 \epsilon_{t-s} + \Theta_2 \epsilon_{t-2s} + \cdots + \Theta_Q \epsilon_{t-Qs}$$

Parameter Estimation Using Maximum Likelihood Estimation (MLE):-

Once the differencing, AR, MA, and seasonal components are defined, the model estimates the parameters using Maximum Likelihood Estimation (MLE). MLE selects the best parameters (p, d, q, P, D, Q, s) that maximize the likelihood function, which measures how well the model fits the observed data. The SARIMA model uses iterative algorithms to optimize and estimate the AR and MA coefficients along with the seasonal terms.

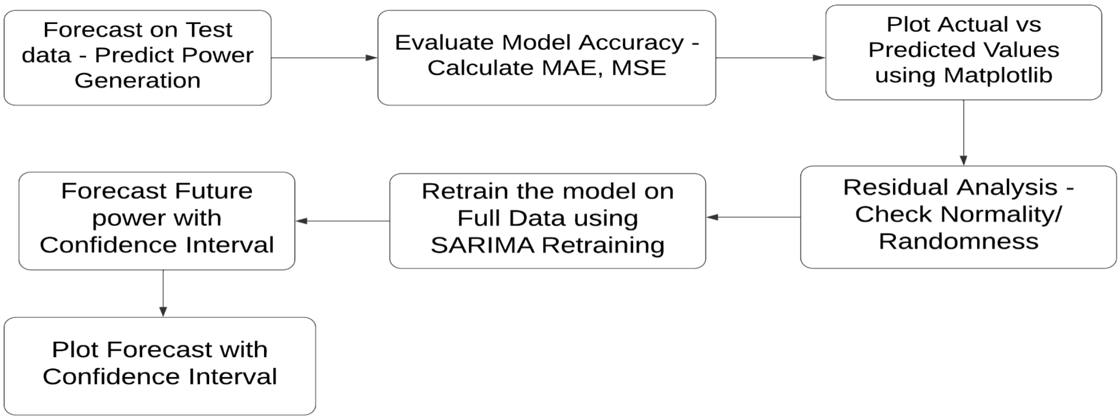
Model Fitting and Prediction

After parameter estimation, the SARIMA model is fitted to the training data by applying the AR, MA, and seasonal components to the differenced time series.

- Model fitting: In this stage, the SARIMA model finds the optimal combination of past values and errors (lagged and seasonal) that best explain the observed data.
- Prediction: Once the model is trained, it can be used to forecast future values. The predicted value at time t+1 is calculated by combining the AR, MA, SAR, and SMA components as follows:

$$\hat{Y}_{t+1} = AR(t) + MA(t) + SAR(t) + SMA(t)$$

5.2.5 MODEL VALIDATION



5.8 Model Validation Data Flow

To validate the SARIMA model for forecasting renewable energy generation, we follow a systematic approach to evaluate its accuracy and reliability. In the first step, we apply the SARIMA model to forecast power generation for both solar and wind power on a separate test set. This test set provides unseen data, allowing us to observe the model's forecasting abilities without the risk of overfitting. Once predictions are generated, we assess the model's performance using evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics give quantitative insight into the model's prediction accuracy by comparing the predicted values with actual values.

For further visual validation, we plot actual versus predicted power generation values for solar and wind power to highlight how closely the predictions follow the observed patterns. Next, we examine the model residuals to ensure they are random and follow a normal distribution. Random, normally distributed residuals indicate that the SARIMA model has effectively captured the underlying structure of the data, as any remaining patterns could signal model underfitting. Finally, we retrain the SARIMA model on the entire dataset, combining training and test data, to further enhance its accuracy for future predictions. This retraining ensures the model is optimized with all available information, improving its robustness.

Once validated, the SARIMA model is used to forecast future combined power generation from solar and wind sources, plotting the forecasted values with confidence intervals. These intervals highlight expected trends in energy production and provide a range of uncertainties, allowing stakeholders to anticipate possible variations. This validation process ensures that the SARIMA model is effective in capturing seasonal and trend-based patterns in renewable energy generation, offering reliable predictions for operational and planning purposes.

Evaluation :-

Mean Absolute Error (MAE) is a metric that measures the average absolute difference between predicted and actual values in a dataset. It represents how much, on average, the model's predictions deviate from the actual values. Mathematically, it is calculated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\mathrm{Actual}_i - \mathrm{Predicted}_i|$$

For example, if the MAE is 100, this indicates that, on average, the model's predictions deviate by 100 units from the actual values. Lower MAE values signify higher model accuracy, meaning the predictions are closer to the observed data, whereas higher values indicate greater deviation.

Mean Squared Error (MSE) is a metric that quantifies the average of the squared differences between the predicted and actual values in a dataset. MSE is particularly useful because it penalizes larger errors more than smaller ones, as squaring the differences amplifies the impact of large deviations. The formula for MSE is:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Actual}_i - \mathrm{Predicted}_i)^2$$

CHAPTER 6

RESULT AND DISCUSSION

6.1 RESULT

The SARIMA model showed strong performance in forecasting renewable energy generation, as evidenced by its Mean Absolute Error (MAE) of 0.1211 and Root Mean Squared Error (RMSE) of 0.1531, both indicating high accuracy and minimal deviations from actual values. These metrics demonstrate the model's capability to generate reliable and precise predictions for solar and wind power output, which are essential in accurately managing energy supply. Additionally, the forecasted power generation values closely matched the actual observed values, displaying only minor deviations and affirming the model's predictive accuracy.

Visual analysis further reinforced the model's effectiveness; the plot comparing forecasted versus actual power generation values showed a strong alignment between the two, confirming the SARIMA model's ability to capture trends accurately. The SARIMA model succeeded in identifying the seasonal and temporal patterns inherent to renewable energy generation, enhancing its suitability for this task. This performance suggests that the SARIMA model is a viable choice for renewable energy forecasting, capable of supporting energy management with high reliability and insightful predictions.

Furthermore, the insights gained from this model can assist energy providers in improving resource allocation, anticipating potential shortfalls or surpluses, and planning maintenance schedules more effectively. As renewable energy continues to grow in importance, the SARIMA model serves as an invaluable tool, not only in forecasting but also in contributing to sustainable energy goals by enabling more informed decision-making.

6.2 DISCUSSION

The results from the SARIMA model highlight its capability as a reliable tool for renewable energy forecasting, particularly in capturing the seasonal and temporal dynamics inherent to solar and wind power generation. The low error metrics—an MAE of 0.1211 and an RMSE of 0.1531—demonstrate that the model not only predicts closely aligned values to actual data but also reduces the typical deviations that can complicate energy management decisions. By accounting for recurring seasonal patterns, the SARIMA model overcomes limitations found in simpler predictive approaches that may ignore these critical variations, thus offering a nuanced understanding of renewable energy output trends.

The close correspondence between forecasted and observed values supports the SARIMA model's practicality for both day-to-day operational planning and longer-term energy strategies. With a focus on renewable energy, such models are essential to support the variability and fluctuations associated with solar and wind resources.

The SARIMA model succeeded in identifying the seasonal and temporal patterns inherent to renewable energy generation, enhancing its suitability for this task. This performance suggests that the SARIMA model is a viable choice for renewable energy forecasting, capable of supporting energy management with high reliability and insightful predictions.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

The SARIMA (Seasonal AutoRegressive Integrated Moving Average) model stands out as a robust tool for accurately forecasting renewable energy generation, particularly for solar and wind sources. In this study, the SARIMA model achieved high precision in predicting energy outputs, evidenced by low Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics. These indicators highlight the model's reliability in minimizing forecast errors, which is crucial for effective energy management.

One of the key strengths of SARIMA lies in its ability to capture and model both seasonal trends and short-term temporal patterns inherent in renewable energy datasets. Solar and wind power generation are highly influenced by seasonal cycles and weather variations, making it essential to use a forecasting approach that can adapt to these fluctuations. The SARIMA model's flexibility in addressing these variations allows it to align predicted values closely with actual energy production, thereby ensuring more accurate forecasts.

The practical implications of deploying the SARIMA model in energy management systems are significant. By accurately predicting energy generation, utility companies can optimize grid operations, balance supply and demand, and efficiently allocate resources. This is especially beneficial in integrating renewable energy sources, which are often intermittent and unpredictable, into the existing power grid. A reliable forecast helps prevent energy shortages, reduce reliance on fossil fuels, and promote sustainable energy practices.

7.2 FUTURE ENHANCEMENT

To further improve the SARIMA model's predictive accuracy and applicability, several future enhancements could be implemented. One key improvement would be incorporating exogenous variables such as temperature, humidity, wind speed, and solar irradiance. These factors directly influence solar and wind power generation and can help the model account for weather-related fluctuations more accurately. Additionally, exploring hybrid models that combine SARIMA with machine learning techniques, like neural networks, could enhance the model's ability to capture complex, non-linear relationships in renewable energy data.

Another enhancement could involve implementing real-time data streaming and forecast updates to make the model adaptable to immediate changes in environmental conditions. Lastly, expanding the model's predictive range to cover longer-term forecasts would assist in planning for seasonal variations and future energy demand. These enhancements would strengthen the SARIMA model's role as a decision-support tool, facilitating more reliable and efficient integration of renewable energy sources into the power grid.

APPENDIX

A1.1 SAMPLE CODE

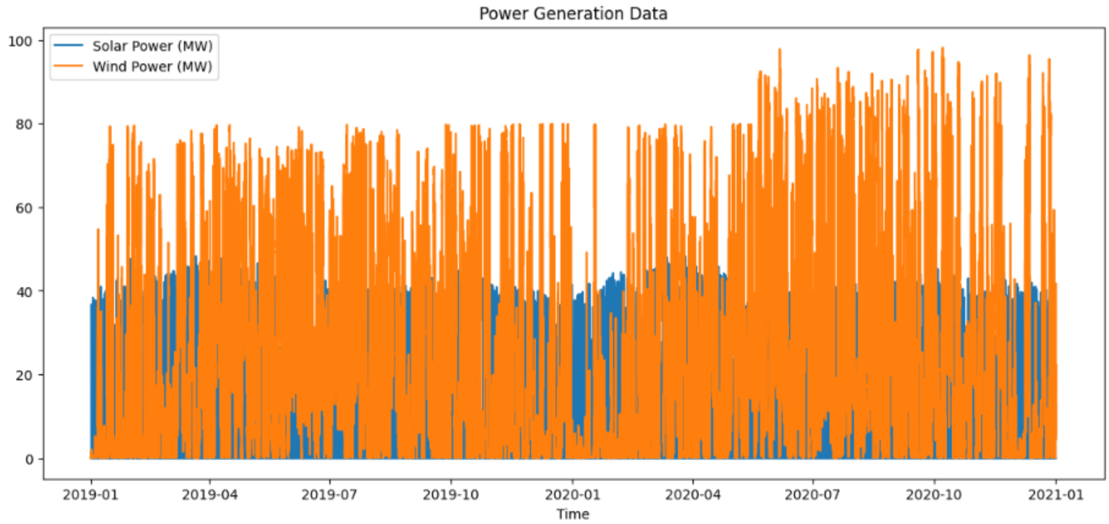
SARIMA MODEL

```
import pandas as pd
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.metrics import mean_absolute_error, mean_squared_error
import numpy as np
import pandas as pd
def load_data(filepath):
 # Load the dataset
    data = pd.read_excel(filepath)
  # Convert the 'Time(year-month-day h:m:s)' column to datetime format
    data['Time(year-month-day h:m:s)'] = pd.to_datetime(data['Time(year-month-day
h:m:s)'])
    # Set the 'Time(year-month-day h:m:s)' column as the index
    data.set_index('Time(year-month-day h:m:s)', inplace=True)
    # Set the frequency as hourly ('H') to ensure consistent intervals in the time series
    data = data.asfreq('H')  # Replace 'H' with your actual data frequency if different
    # Return the 'Power (MW)' column for forecasting
    return data['Power (MW)']
def train_and_forecast(filepath, forecast_length, forecast_unit):
    # Load dataset
```

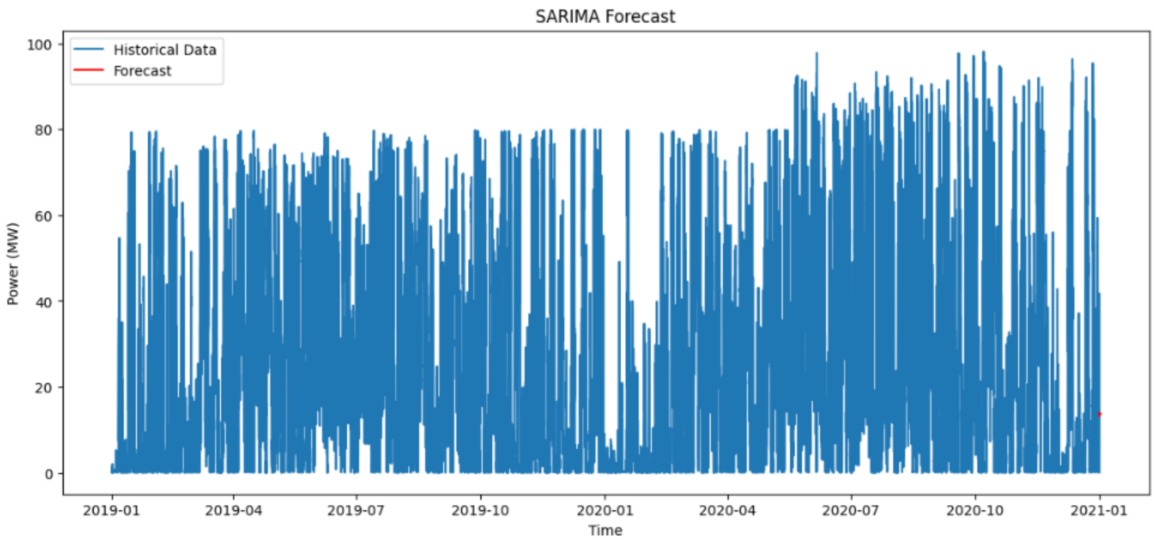
```
data = load_data(filepath)


# Set SARIMA order (you can further optimize this)
order = (1, 1, 1)
seasonal_order = (1, 1, 1, 24 if forecast_unit == 'hours' else 7)
# Fit the SARIMA model
model = SARIMAX(data, order=order, seasonal_order=seasonal_order)
model_fit = model.fit(disp=False)
# Forecast for the specified steps
steps = forecast_length if forecast_unit == 'hours' else forecast_length * 24
forecast = model_fit.forecast(steps=steps)
# Calculate model metrics
predictions = model_fit.predict(start=0, end=len(data)-1)
mae = mean_absolute_error(data, predictions)
mse = mean_squared_error(data, predictions)
rmse = np.sqrt(mse)
return forecast, mae, mse, rmse
```

A1.2 SCREENSHORTS



A1.2.1 Solar and Wind Data Distribution

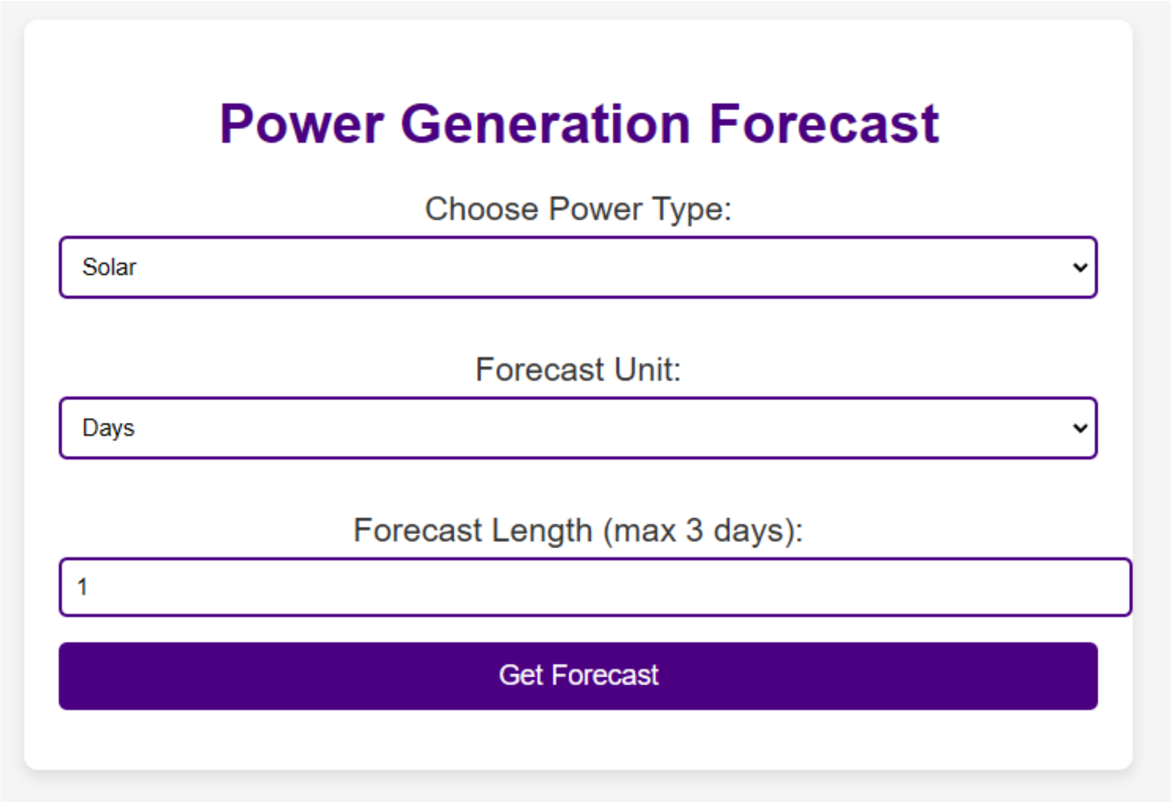


A1.2.2 Models Forecast

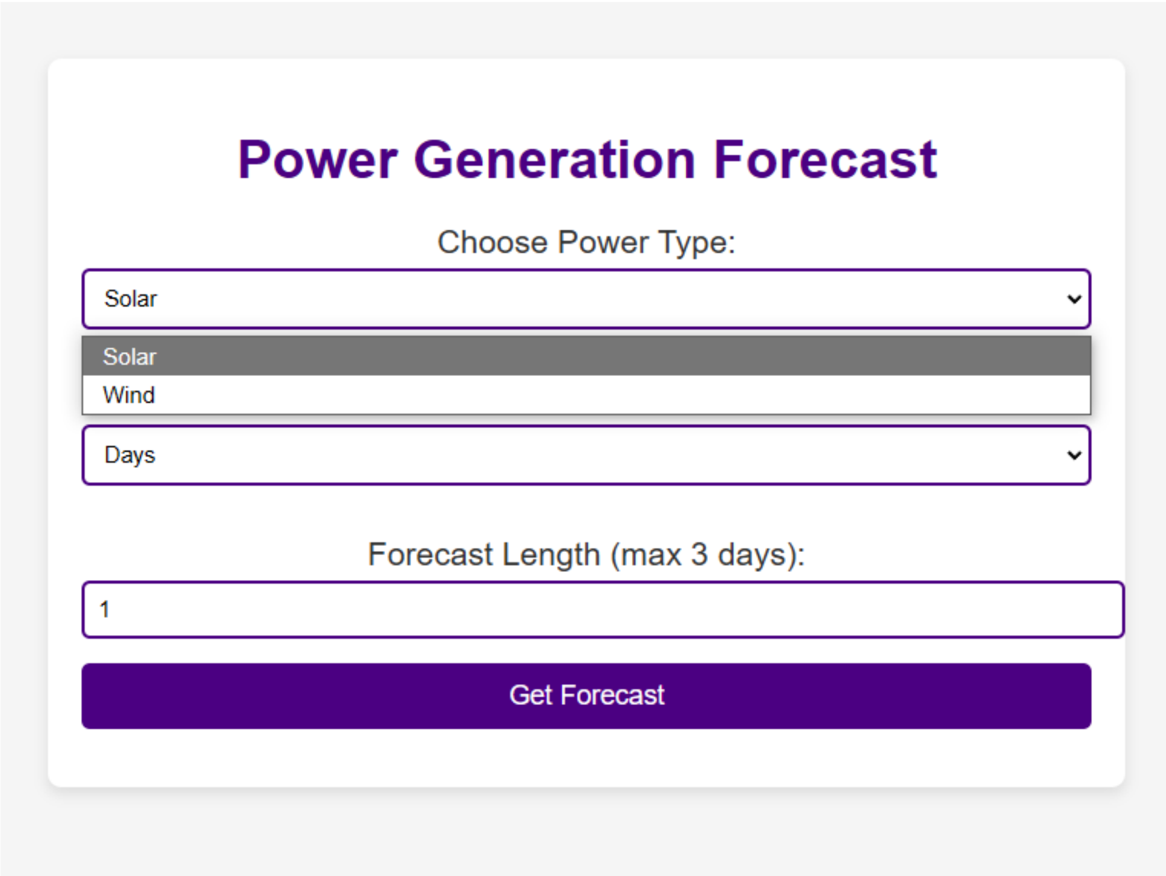
```
Solar Power Model Performance:
MAE: 0.9507207820437552, MSE: 4.413073591124042, RMSE: 2.100731679944881

Wind Power Model Performance:
MAE: 2.534610797125656, MSE: 21.069203246735736, RMSE: 4.590120177809698
```

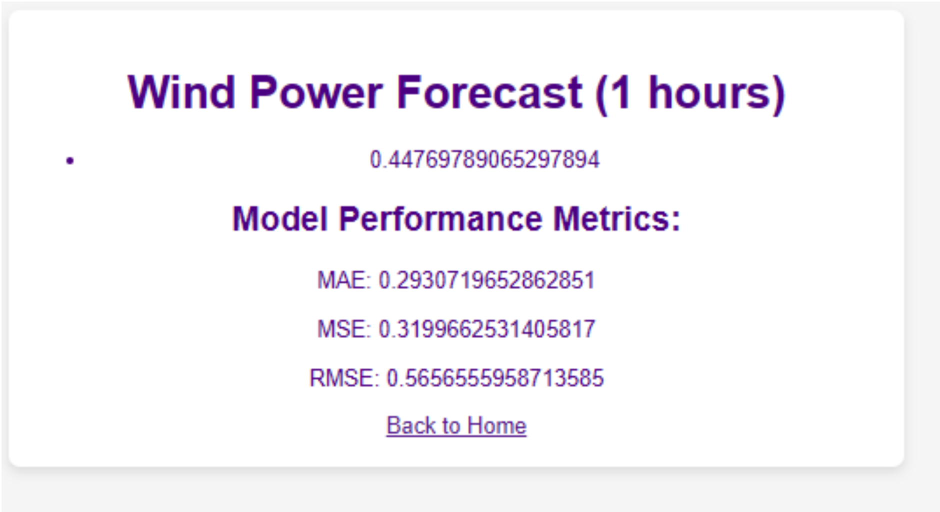
A1.2.3 Model Evaluations



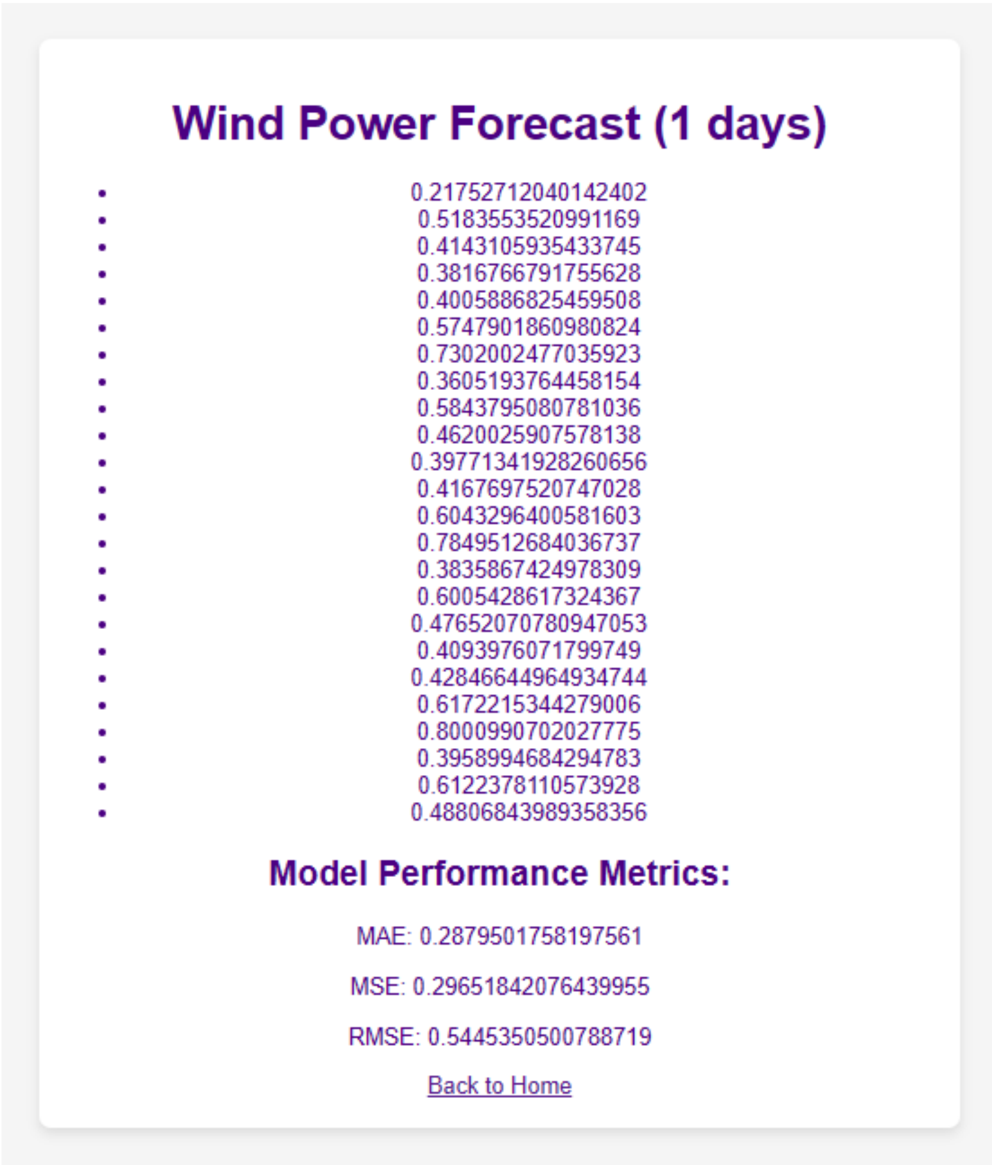
A1.2.4 Forecast UI



A1.2.5 Power Selecting UI



A1.2.6 Metrices UI



A1.2.7 Prediction UI

REFERENCES

[1] M. Islam, M. Mahmud, and E. Hossain, "Forecasting of Solar and Wind Resources for Power Generation," Energies, vol. 16, no. 5, pp. 1234-1250, 2023.

[2] A. Ahmed, L. Gauntlett, and N. Jenkins, "Short-term forecasting of wind speed and power," Renewable Energy, vol. 40, no. 2, pp. 185-189, 2021.

[3]J. Smith and K. Brown, "Predicting Solar Power Output Using Machine Learning," in “Proc. 2023 Int. Conf. Renew. Energy Sustain. Technol. (ICREST)”, Berlin, Germany, 2023, pp. 45-50.

[4]W. Chen and Y. Li, “Renewable Energy Forecasting: From Models to Applications”. New York, NY, USA: Wiley, 2021.

[4] A. Zendehboudi, M. A. Baseer, and R. Saidur, "Application of Support Vector Machine Models for Forecasting Solar and Wind Energy Resources: A Review," Journal of Cleaner Production, vol. 199, pp. 272–285, 2018. doi: 10.1016/j.jclepro.2018.07.164.

[5] G. I. Sayed, E. I. Abd El-Latif, A. E. Hassanien, and V. Snasel, "Optimized Long Short-Term Memory with Rough Set for Sustainable Forecasting Renewable Energy Generation," Energy Reports, vol. 11, pp. 6208–6222, 2024.

[6]A. Khan, R. Bhatnagar, V. Masrani, and V. B. Lobo, "A Comparative Study on Solar Power Forecasting using Ensemble Learning," in Proceedings of the 4th International Conference on Trends in Electronics and Informatics (ICOEI 2020), pp. 224–230, 2020, doi: 10.1109/ICOEI.2020.9143010.

[7] D.A. Widodo, N. Iksan, and E. D. Udayanthi, "Renewable Energy Power Generation Forecasting Using Deep Learning Method," IOP Conference Series: Earth and Environmental Science, vol. 700, no. 1, p. 012026, 2021, doi: 10.1088/1755-1315/700/1/012026.

[8] O. Rubanenko, M. Cosovic, S. L. Gundebommu, and V. Lesko, "Predicting Power Generation from Renewable Energy Sources Using ANN," in 20th International Symposium INFOTEH-JAHORINA, 2021, doi: 10.1109/INFOTEH51037.2021.9400527.

[9] U. M. Iqbal, J. Ramesh, and A. R. Al-Ali, “Deep Learning Approach for Wind Energy Time-Series Forecasting with Environmental Factors,” in Proc. American University of Sharjah, Sharjah, UAE, Sep. 2023, pp. 1–15.

[10] L. F. J. Alvarez, S. R. González, A. D. López, D. A. H. Delgado, R. Espinosa, and S. Gutiérrez, "Renewable Energy Prediction through Machine Learning Algorithms," 2020 IEEE Xplore, pp. 1–6, 2020. doi: 10.1109/978-1-7281-9365-6/20.

LIST OF PUBLICATIONS

1. **Microsoft CMT**

TITLE: INTERNATIONAL CONFERENCE ON EMERGING RESEARCH IN COMPUTATIONAL SCIENCE – 2024

INTERNATIONAL CONFERENCE ON EMERGING RESEARCH IN COMPUTATIONAL SCIENCE - 2024 : Submission (230) has been created. Inbox ×

Microsoft CMT <email@msr-cmt.org>
to me

Hello,

The following submission has been created.

Track Name: ICERCS2024

Paper ID: 230

Paper Title: Survey on Forecasting Renewable Energy Generation(Solar,Wind) using ML &DL techniques

Abstract:
Renewable Energy plays a vital role in our daily life by saving the environment as it reduces the usage of non-renewable resources for energy production which in turn brings down carbon emission. To increase renewable energy generation, an accurate predictive model is required. Traditional forecasting models have some drawbacks, which includes inefficiency in acknowledging the nonlinear relationship and the seasonality pattern over the period of time. The following constraints can be removed by applying the algorithms of ML and DL. Both of the methodologies help us predict renewable energy generation with accuracy. This survey is going to introduce a systematic review and detailed view of the researches made on renewable energy prediction, focusing on the prediction of solar and wind energies, using ML and DL techniques like SVM, RF, LR, MLP, DT, Gradient Boost, LSTM, and hybrid models.

Created on: Sun, 27 Oct 2024 07:18:07 GMT

Last Modified: Sun, 27 Oct 2024 07:18:07 GMT

Authors:
- vijaykumarece166@gmail.com (Primary)
- 221801062@rajalakshmi.edu.in
- sureshkumar.s@rajalakshmi.edu.in

Secondary Subject Areas: Not Entered

Submission Files:
Renewable Energy prediction final.pdf (370 Kb, Sun, 27 Oct 2024 07:18:02 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our Privacy Statement.

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

A1.4.1 First Conference Communicated

2. **Springers**

TITLE: 2ND INTERNATIONAL CONFERENCE ON COMPUTING FOR SCIENCE, ENGINEERING AND ARTIFICIAL INTELLIGENCE (CSEAI) 2024

CSEAI2024 · Submission of your paper 002 External Inbox ×

EquinOCS <equinocs-admins@springernature.com>
to me

This message has been sent by the EquinOCS system
https://equinocs.springernature.com/

PLEASE DO NOT REPLY

========

Dear Vijay Kumar V,

We are pleased to inform you that your paper

002: "Survey on Forecasting Renewable Energy Generation(Solar,Wind) using ML &DL techniques"

has been sucessfully submitted to

CSEAI2024

by Vijay Kumar V (@vijay16).

To access the paper:
- log into your EquinOCS account
- navigate to CSEAI2024
- access the paper 002 via the 'Your Submissions' page

If you have no EquinOCS account yet, register with EquinOCS using the
email address at which you have been receiving this notification.
This way, the paper can be associated with your account.
You will also find the licencing information there.

========

PLEASE DO NOT REPLY

This message has been sent by the EquinOCS system

A1.4.2 First Conference Communicated