Major Project Report

on

Solar Energy Forecasting

Submitted by

Garimella Amruthavalli - 20bec014

Rujula Deependra Karanjkar - 20bec041

S Padmalatha - 20bec042

Rakshitha Y - 20bcs107

Under the guidance of

Dr. Prakash Pawar

Assistant Professor,

Electronics and Communication Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

22/03/2024

Certificate

This is to certify that the project, entitled **Solar Energy Forecasting**, is a bonafide record of the Major Project coursework presented by the students whose names are given below during <2024> in partial fulfilment of the requirements of the degree of Bachelor of Technology in Computer Science and Engineering.

Roll No	Names of Students
20bec014	Garimella Amruthavalli
20bec041	Rujula Deependra Karanjkar
20bec042	S Padmalatha
20bcs107	Rakshitha Y

Dr. Prakash Pawar
(Project Supervisor)

Contents

List of Figures **ii**

1 Introduction **1**

2 Related Work **2**

3 Data and Methods **7**

4 Results and Discussions **10**

5 Conclusion **17**

References **18**

List of Figures

1 Flowchart for the procedure method 9
2 Flowchart for the procedure method 9
3 DC and AC power output for both the Plants 10
4 Line Plots based on Sensor data for both the Plants 11
5 Solar Irradiation for both the Plants 12
6 Module Temperature for both the Plants 12
7 Ambient Temperature for both the Plants 12
8 RNN Day-wise Actual vs Predicted Values for Plant 1 13
9 LSTM Day-wise Actual vs Predicted Values for Plant 2 14
10 RNN Week-wise Actual vs Predicted Values for Plant 1 15
11 LSTM Week-wise Actual vs Predicted Values for Plant 2 15
12 Comparison of Models for Plant 1 16
13 Comparison of Models for Plant 2 16

1 Introduction

The utilization of renewable energy sources, particularly solar energy, has gained significant attention in recent years due to its potential to address the global energy crisis and reduce greenhouse gas emissions. Solar power plants play a crucial role in harnessing solar energy and converting it into electricity. Solar power plants harness this energy through photovoltaic cells, converting sunlight directly into electricity. These plants play a crucial role in our transition to a sustainable energy future, providing clean and renewable power to homes, businesses, and industries. However, the efficiency of solar power generation is heavily influenced by various factors, including weather conditions, equipment performance, and environmental parameters like temperature and sunlight intensity. Understanding and predicting these variables are essential for optimizing the operation of solar plants and ensuring a reliable energy supply.

Several methods are employed to predict photovoltaic power generation (PVPG) accurately for solar energy forecasting. These methods include physical models, statistical models, and machine learning (ML) models each offering unique advantages and challenges. The physical model relies on mathematical equations that describe the physical state of the PV system, primarily based on global irradiance on solar cells. While physical models can achieve acceptable forecasting accuracy under stable weather conditions, their performance may falter when faced with significant weather fluctuations. In contrast, statistical models establish mathematical relationships between random and non-random variables, forming the basis for inferences and predictions. Machine learning (ML) models, a subset of statistical models, have gained prominence in recent years for their versatility and adaptability. These models involve data preparation, algorithm training, and model generation to make and refine predictions.

In this project, we employ a machine learning-based approach for solar energy prediction. By leveraging historical data from solar power plants, including measurements of DC and AC power, environmental parameters, and energy yield, we train predictive models to forecast future energy output. These models utilize algorithms such as linear regression, decision trees, and random forests to analyze data patterns and make accurate predictions. Through rigorous evaluation and optimization, we aim to develop robust forecasting models that can contribute to the efficient operation of solar power plants and facilitate the integration of solar energy into the grid.

2 Related Work

Forecasting solar energy production: A comparative study of machine learning algorithms: Younes Ledmaoui [1] This study was aimed to address the need for accurate energy production forecasting and monitoring in a solar installation of 24 KWc. Through the utilization of machine learning algorithms, including SVR, ANN, DT, RF, GAM and XGBoost, the performance was evaluated in predicting energy production. This process included estimation of required power and energy output. They calculated the number of PV modules which was found to be 56 photovoltaic panels of 430 Wp, along with the current and voltage at the maximum power point and the power of the real photovoltaic generator. Based on the predictions, the system was installed which monitored the daily production for a year, which was turned into a database. The study took into account two crucial parameters; temperature and irradiance. Based on the analysis, the ANN algorithm emerged as the most accurate model for energy predictions in the specific scenario.

Solar power generation forecasting using ensemble approach based on deep learning and statistical method:Mariam AlKandari [2] introduced a hybrid model that integrated machine-learning techniques with the Theta statistical method to improve the precision of forecasting future solar power generation from renewable energy plants. The machine learning models included long short-term memory (LSTM), gate recurrent unit (GRU), AutoEncoder LSTM (Auto-LSTM) and a newly proposed Auto-GRU. In order to enhance the accuracy of the proposed Machine Learning and Statistical Hybrid Model (MLSHM), two strategies, namely structural diversity and data diversity, were incorporated. To combine the prediction of the ensemble members in the proposed MLSHM, four methods were combined; simple averaging approach, weighted averaging using linear approach and using non-linear approach, and combination through variance using inverse approach. The proposed MLSHM scheme was validated on two real-time series datasets. Experimental results indicated that the proposed MLSHM, utilizing all the combination methods, achieved superior accuracy compared to the predictions of traditional individual models. Results demonstrated that a hybrid model combining machine-learning methods with statistical methods outperformed a hybrid model that only combines machine-learning models without statistical methods.

Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge: Xing Luo [3]

They proposed a deep learning based framework for accurate PVPG forecasting. This paper considers the specific domain knowledge of PV and proposes a physics-constrained LSTM (PC-LSTM) to forecast the hourly day-ahead PVPG. Real-life PV datasets are adopted to evaluate the feasibility and effectiveness of the models. To theoretically identify highly correlated feature variables, a two-stage hybrid method is proposed, integrating filter and wrapper processes. Additionally, a sensitivity analysis of PV data is performed to ensure suitable input for Machine Learning (ML) models. The results indicate that the proposed PC-LSTM model possesses stronger forecasting capability than the standard LSTM model.

Solar Power Forecasting Using Deep Learning Techniques :MEFTAH Elsaraiti and Adel Merabet [4] discussed a method for predicting the generated power, in the short term, of photovoltaic power plants, by means of deep learning techniques. A deep learning technique based on the Long Short Term Memory (LSTM) algorithm is evaluated with respect to its ability to forecast solar power data. In this study, the MATLAB software (R2019b) was used for the training process of the LSTM. To evaluate the effectiveness of the proposed method, a case study was performed using a data set that includes one-year data.For each day, data was selected only during daylight hours, from 8 am to 5 pm. The original photo-electric data were collected at 5-minute intervals and included 43800 measurements.The task of predicting photovoltaic power for the coming days at 30-minute intervals was considered. The results of the suggested model were compared using the Multi-Layer Perceptron (MLP) algorithm. When compared to the MLP algorithm, the prediction performance of the suggested LSTM model offered more effective values in all performance parameters like Mean Absolute Error (MAE), Mean Absolute Percentage Error(MAPE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R2).

A Comprehensive Review on Ensemble Solar Power Forecasting Algorithms: Negar Rahimi [5] reviewed recently published studies on solar irradiance forecasting using ensemble models including competitive and cooperative forecasting methods.In this article, the different model inputs, and their effects on the prediction of solar radiation have been discussed. Common inputs are solar irradiance, atmospheric and module temperature, wind speed and direction, and humidity. Among these, solar irradiance is most positively correlated with PV output. Artificial Neural Network (ANN) and Space Vector Machine (SVM) are widely used with ensemble models (WD-ANN, EMD-BPNN and W-SVM) due to their ability in solving complex and non-linear forecasting models. These ensemble approaches leverage a combination of statistical, physical, and machine learning techniques to account for the dynamic nature of solar irradiance and mitigate the impact of fluctuations on solar power systems. EMD is more powerful, and it has more accuracy than WD in solar forecasting methods.They concluded that ensemble models perform better than standalone ones. Hybrid models have a more complex structure, but they provide better accuracy.

Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques: K. Anuradha [6] presented a machine learning-based approach for solar power generation analysis which accurately forecasted power generated across India's states based on environmental data generation. The dataset used had daily average temperature in Celsius, distance from solar noon, wind speed, wind direction, sky cover, and humidity and then the power generated. The available dataset was based on hourly weather parameter values. To convert the data mean values per day, the average of the 24-hour data was used. Support Vector Machine, Random Forest, Linear Regression are various ML Models used in this paper. Certain error statistics such as mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), relative MBE (rMBE), mean percentage error (MPE), and relative RMSE were used to assess the model's validity (rRMSE). The proposed models are SVR, LR, and RF. By a wide margin, the proposed method outperformed other popular methods, such as Random forest.

The research, **Predicting Solar Energy Generation with Machine Learning based on AQI and Weather Features: Arjun Shah,Varun Viswanath [7]** determined solar power generation utilizing machine learning models based on weather data and AQI(Air Quality Index). It evaluated the performance of these models in terms of R2 score and simultaneously results in the lowest Mean Absolute Error(MAE) and Root Mean Square Error(RMSE). The UNISOLAR Solar Generation Dataset which includes two years of Photovoltaic solar energy generation data collected at an interval of 15 minutes was used. Weather data like apparent temperature, air temperature, dew point temperature, wind speed, wind direction, and relative humidity were also provided by the dataset. AQI data is utilized, to improve the input features of the model, which was captured by a station near the location of the solar generation plant from which the original data was sourced. The chief models employed were Linear Regression, Lasso, Ridge, ElasticNet, ensemble models like RandomForest and XGBoost, and deep learning models like ConvL-STM2D. However, on noticing the skewness of the solar energy generation data which had a high number of zeros, they switched to a zero-inflated model which immediately yielded a higher accuracy of solar prediction with a lower mean standard error.

Forecasting Solar Energy Production Using Machine Learning: C.Vennila [8] validated the forecasting made by the ensemble model for optimal prediction of power generation using PV plants. The study considered two case studies, where the former was simulated for smaller PV farms of 1000 PV cells and larger PV farms of 100000 PV cells. In this the used data had five-minute resolution. It employed the traditional LSTM model for the analysis of feature subsets because of its capacity to address time series forecasting issues. The data was classified into a single classifier and the training data was used to classify the samples. Aggregation was done once it was completed. This was used for metric evaluation with several comparison techniques. The performance of the ensemble model that integrated all of the combination strategies to standard individual models suggested that the ensemble model outperformed the conventional individual models. According to the findings, the hybrid model that made use of both machine learning and statistics outperformed a model that made sole use of machine learning in its performance.

The study, **Solar Power Forecasting Using the Artificial Neural Networks : Mohamed Abuella [9]** utilized a Global Energy Forecasting Competition 2014 (GEFCOM2014) dataset which included forecasting in the domains of electric load, wind power, solar power, and electricity prices.TTwo distinct testing scenarios were conducted, aiming to forecast solar power generation on an hourly basis for September 2013 and May 2014, each trained separately due to the disparity in available historical data. Evaluation of forecast accuracy and model performance was conducted through diverse methodologies, including the assessment of plots and graphs, computation of Root Mean Square Error (RMSE), determination of correlation coefficients (R) between forecasts and actual measurements, and comparison against alternative models such as Multiple Linear Regression (MLR) Analysis. The findings underscored the superiority of the artificial neural networks (ANN) model over MLR and persistence models, with performance contingent upon robust training and data quality. The study discerned that forecasts during clear sky conditions yielded superior accuracy compared to cloudy periods, underscoring the pivotal role of precise weather forecasts in enhancing solar power predictions. The study posited that augmenting historical data could further enhance model performance, offering avenues for future research and refinement in solar power forecasting methodologies.

Intelligent solar photovoltaic power forecasting: Keaobaka D. Potia, Raj M. Naidooa, Nsilulu T.Mbungu, Ramesh C. Bansalc [**10**] introduced an intelligent solar power forecasting model employing the day-ahead methodology. This innovative approach integrated numerical weather prediction (NWP) models sourced from open weather maps with power plant specifications to predict the power output from photovoltaic (PV) power plants. The predicted output was then incorporated into an optimal control strategy for PV plants equipped with battery storage systems. By leveraging optimal algorithms, the model facilitated PV power plant curtailment during instances of over-generation and minimized reliance on conventional power sources like generators during periods of under-generation. The forecasting model, simulated using data from a 16.8 kW PV power plant, showcased its ability to accurately forecast the output power of PV plants utilizing real-time weather data.

The study, **Solar Irradiance Forecasting Using Deep Learning Techniques: Hammad Ali Khan** [**11**] investigated deep learning techniques for very short-term solar irradiance forecasting, specifically tailored for the city of Karachi. Leveraging a dataset from 2019 acquired from the NSRDB, consisting of 35,040 samples with a 15-minute time resolution, the research explored the effectiveness of RNN, GRU, LSTM, and TCN models. Results showcased the superior performance of the LSTM model, exhibiting the highest R-squared value and lowest RMSE among the architectures considered. These findings underscored the significance of accurate forecasting models in optimizing renewable energy generation and grid management, with potential applications across various sectors.

Deep Learning Enhanced Solar Energy Forecasting with AI- Driven IoT :Hangxia Zhou [**12**] proposed a hybrid deep learning approach for solar energy forecasting, integrating weather categories into the model. Unlike traditional methods, this approach incorporated correlation analysis and clustered calculations to enhance model generalization and prediction accuracy. The training algorithm utilized CNN to extract data features effectively, while an attention mechanism was applied to the LSTM model to focus on important features. The dataset was over a period from October 2014 to September 2018, with a 7.5-minute time interval. Training data comprised records from 2014 to 2016, while testing used data from 2017 to 2018. Remote sensors captured PV module temperature, current, voltage, frequency, phases, and PV power every 7.5 minutes. Results demonstrated superior performance over traditional methods like MLP, LSTM, and ALSTM, as evidenced by lower RMSE, MAPE, and MAE. The method excelled in short-term and very short-term forecasting, albeit with diminishing advantage in longer-term forecasts. Future research aimed to extend the framework to broader datasets and apply it to diverse time series data applications beyond energy forecasting.

3 Data and Methods

Solar Power Generation Data

The dataset utilized in this project spans a period of 34 days, from May 15, 2020, to June 17, 2020, capturing data from two solar power plants located in India. It comprises of two sets of files, each set includes a power generation dataset and a sensor readings dataset. The power generation data is collected at the inverter level, with each inverter connected to multiple lines of solar panels. Meanwhile, the sensor data is gathered at the plant level, featuring a single array of optimally placed sensors within the plant. Observations are recorded at 15-minute intervals, providing detailed insights into the performance and environmental conditions of the solar power plants.

The power generation dataset includes crucial parameters such as timestamp, unique identifiers for each inverter/source, direct current (DC) power, alternating current (AC) power, daily yield (total energy generated per day), and total yield (cumulative energy generated since installation). On the other hand, the weather sensor dataset contains information on ambient temperature, module temperature (temperature of the solar panel/module), and solar irradiation, all recorded at the same timestamp intervals as the power generation data. This dataset serves as the foundation for analyzing and forecasting solar energy generation, offering valuable insights into the relationship between environmental factors and power output.

Methodology

The initial work involved a comprehensive exploration of the datasets, encompassing generation and weather data for both the plants. Essential libraries were installed, and the datasets were meticulously examined, laying the groundwork for subsequent analyses. Further a statistical summary of the all the files in the dataset was presented, offering insights into key metrics such as mean, standard deviation, and minimum and maximum values. Plots were constructed to visualize the distribution of the different column variables in all the files.

The generation and sensor data were processed separately. Data cleaning involved removing unnecessary columns and grouping essential ones. To enable time-based analysis, the 'DATE TIME' column in both power generation and sensor data was converted to datetime format. Subsequently, the power generation and sensor data were merged based on the 'DATE TIME' column, creating a unified dataset for further analysis. Descriptive analysis was conducted to identify missing values, and heatmaps were plotted for visualization. Bar plots were generated for both plants, revealing stable outputs in Plant 1 and varying outputs among inverters in Plant 2.

Further analysis included the construction of scatter plots to examine the relationship between power generation data features, focusing on AC and DC power for each plant. Similar procedures were applied to sensor data, where line plots were used to analyze module temperature, ambient temperature, and solar irradiation. Through analysis and correlation studies, the necessary features were derived for feature engineering. The final selected input features included 'DAILY YIELD', 'TOTAL YIELD', 'AMBIENT TEMPERATURE', 'MODULE TEMPERATURE', 'IRRADIATION', 'DC POWER', and 'AC POWER', which were utilized for evaluating different machine learning models.

The training and the validation set were split wherein we utilized 80 percent of the training data as the training set and 20 percent as the validation set. Finally, five models linear regression, decision tree, random forest, recurrent neural network and LSTM models were trained on the data. The models were evaluated based on four metrics, mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and r-square (R2).

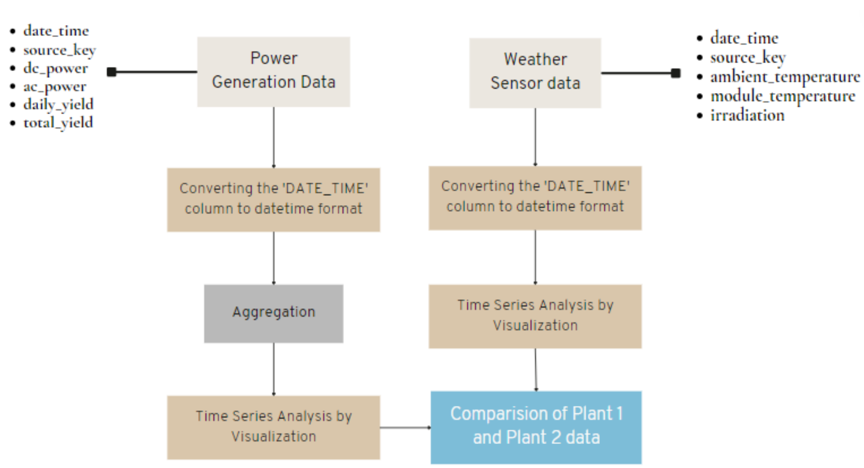

Figure 1. Flowchart for the procedure method

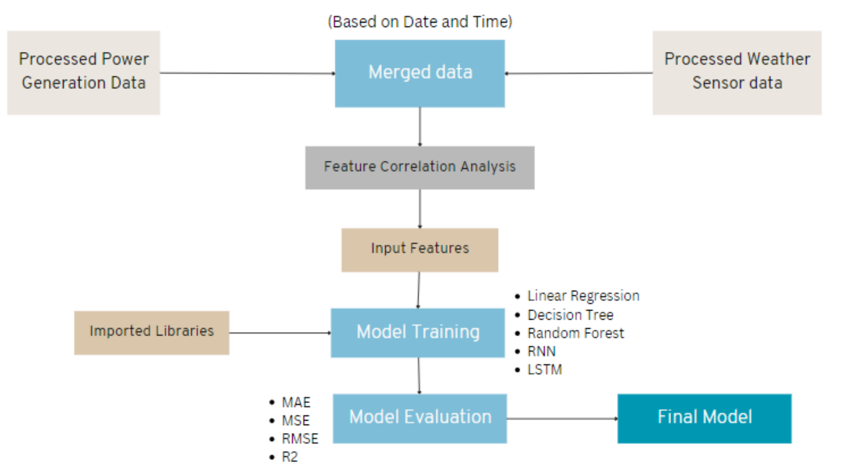

Figure 2. Flowchart for the procedure method

4 Results and Discussions

The results revealed insightful findings regarding the power generation and sensor data from both solar power plants. Observations on the power generation data illustrated that power output correlates with sunlight presence, typically starting around 0540hrs and ending at 1800hrs. Plant 2 displayed more scattered AC and DC power values, suggesting potential issues with the solar modules. The DC power output from Plant 2's solar modules were significantly lower, approximately ten times less than that of Plant 1. However, despite differences in DC power output, both plants exhibited similar levels of AC power, with the Plant 2 displaying more erratic behavior.

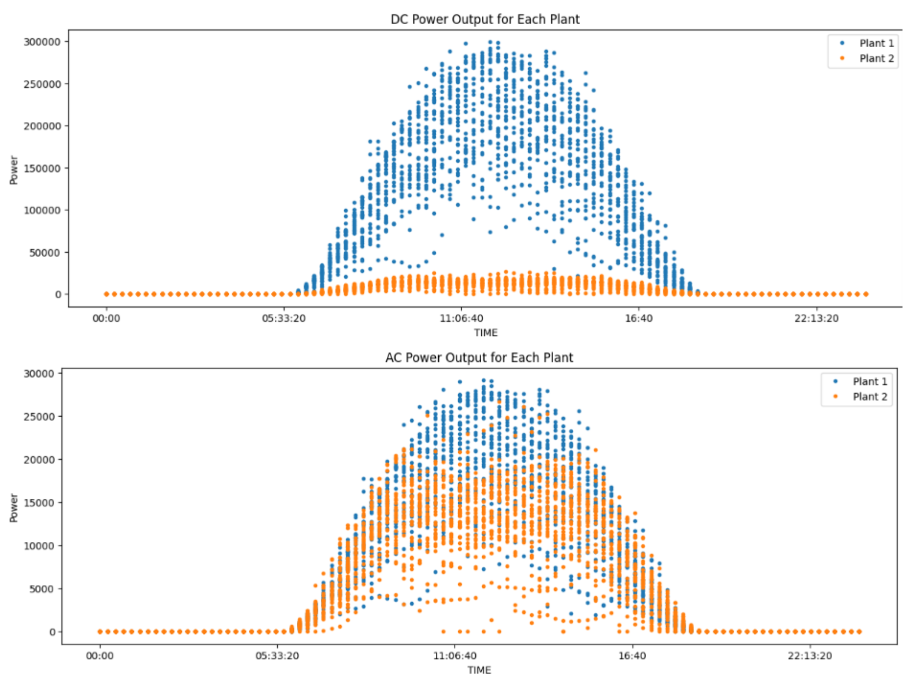

Figure 3. DC and AC power output for both the Plants

Visualization of sensor data through line plots revealed some additional insights. While mean solar irradiation values were comparable between the two plants, overall Plant 1 maintained a slightly lower mean module temperature and ambient temperature than Plant 2. This suggested that Plant 1 may be located in a colder region with less temperature fluctuation, potentially contributing to its more reliable performance. Furthermore, despite similar levels of sunlight received daily, Plant 2 exhibited slightly more erratic irradiation values, implying a potentially cloudier location compared to Plant 1.

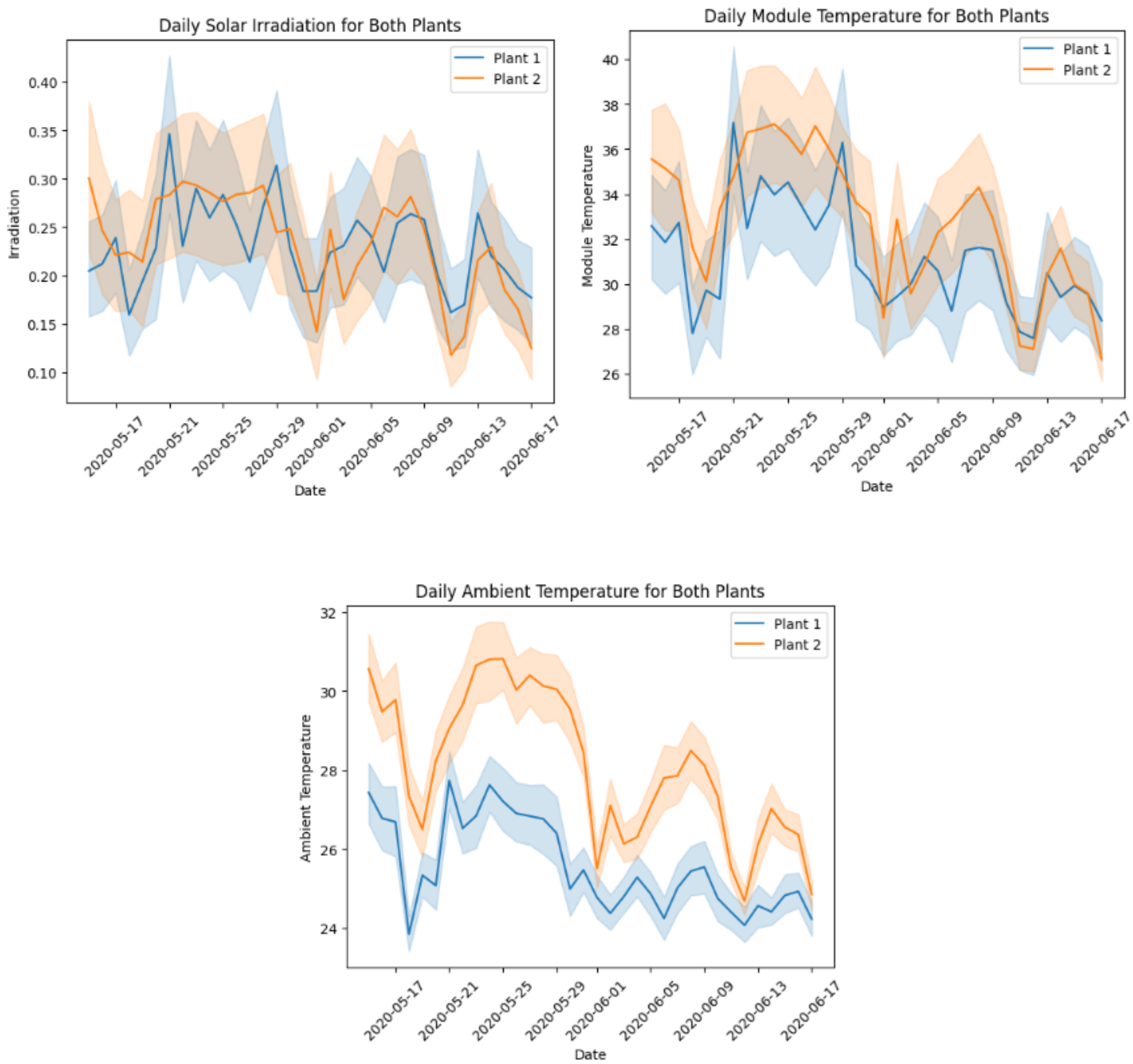

Figure 4. Line Plots based on Sensor data for both the Plants

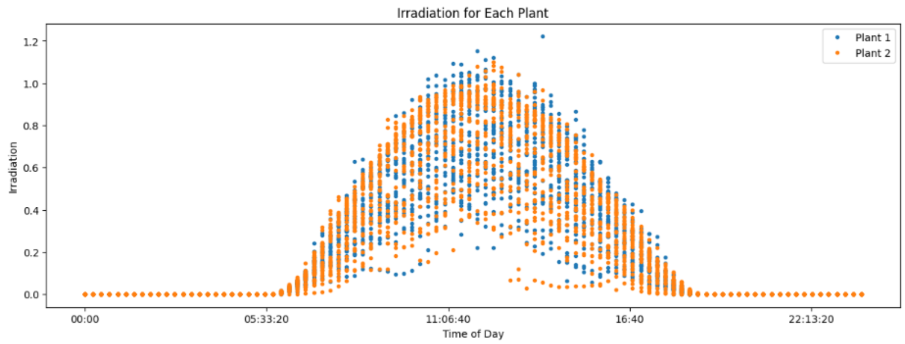

Figure 5. Solar Irradiation for both the Plants

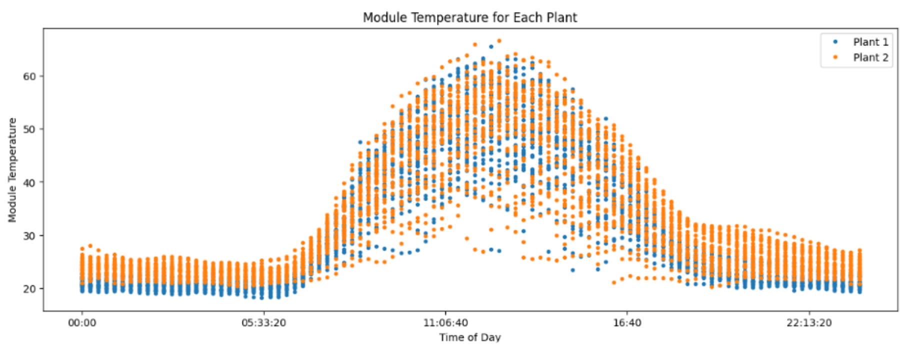

Figure 6. Module Temperature for both the Plants

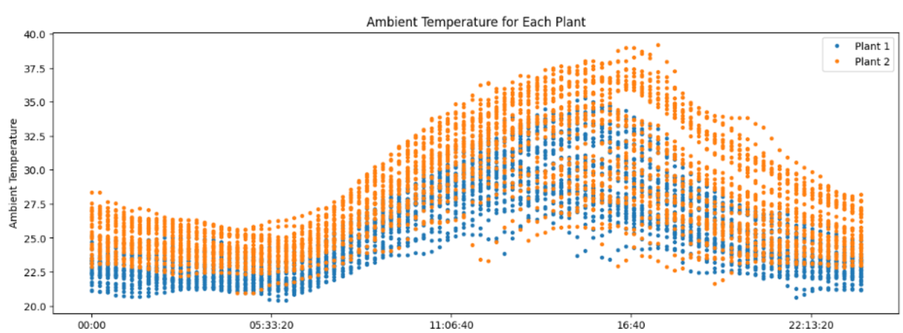

Figure 7. Ambient Temperature for both the Plants

Regarding model performance, the evaluation metrics indicated that Plant 1's power generation is more predictable compared to Plant 2. Plant 1 consistently demonstrated tighter regression and lower RMSE values across all evaluated algorithms, suggesting higher overall system efficiency. These findings underscore the importance of considering environmental factors and location-specific conditions when analyzing solar power generation data and designing predictive models for optimal performance.

Among the evaluated algorithms, the Recurrent Neural Network emerges as the top performer for Plant 1 and LSTM for Plant 2, exhibiting the lowest Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and highest R-squared (R2) value, indicating its effectiveness in predicting power generation from the solar plant's data.

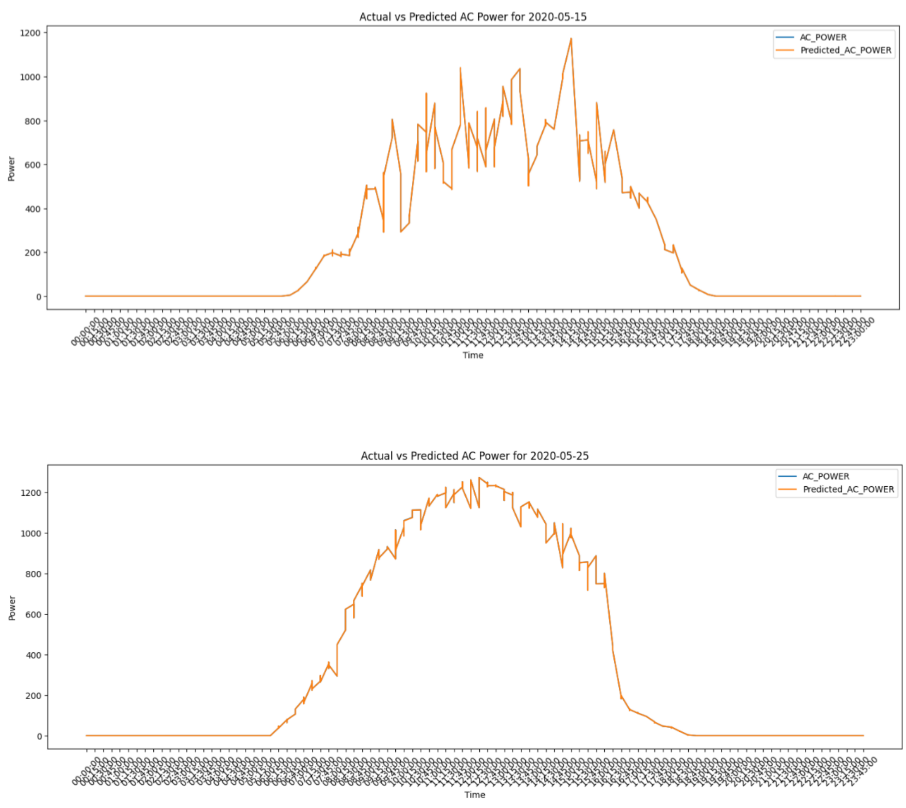

Figure 8. RNN Day-wise Actual vs Predicted Values for Plant 1

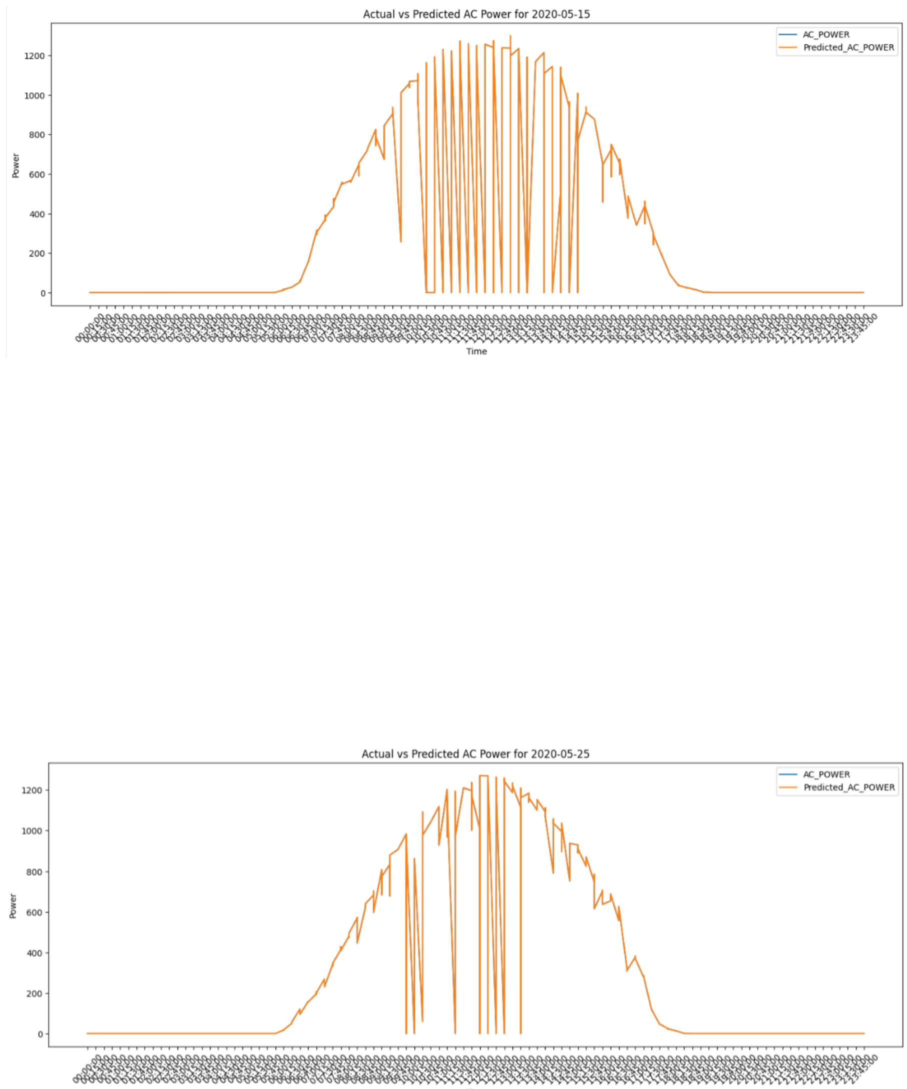

Figure 9. LSTM Day-wise Actual vs Predicted Values for Plant 2

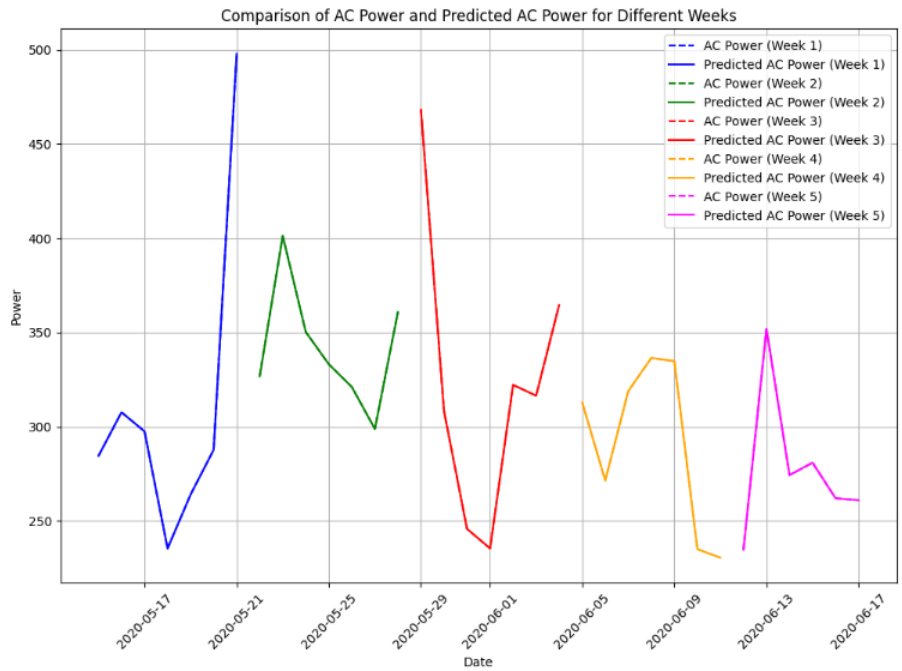

Figure 10. RNN Week-wise Actual vs Predicted Values for Plant 1

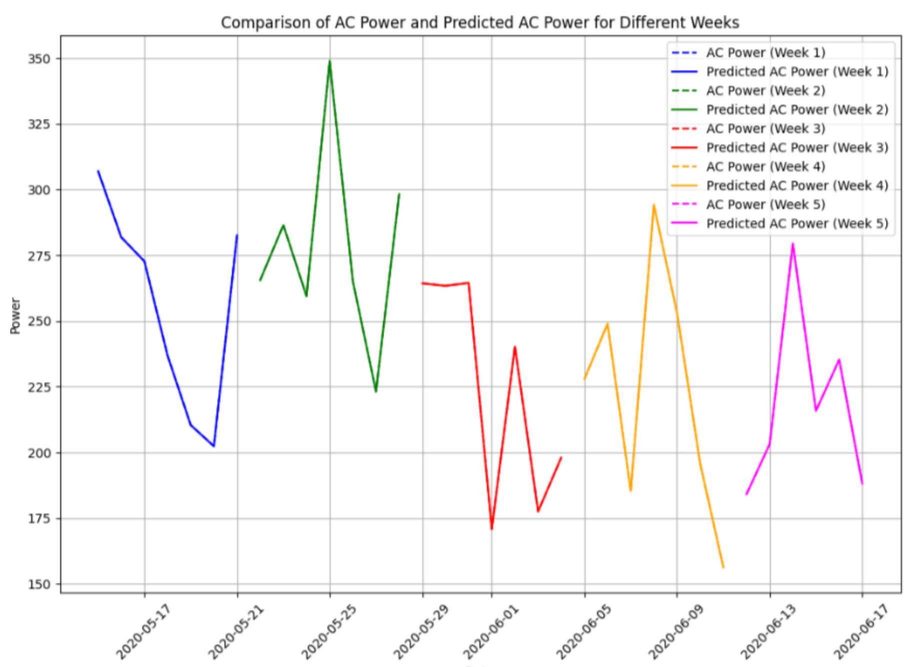

Figure 11. LSTM Week-wise Actual vs Predicted Values for Plant 2

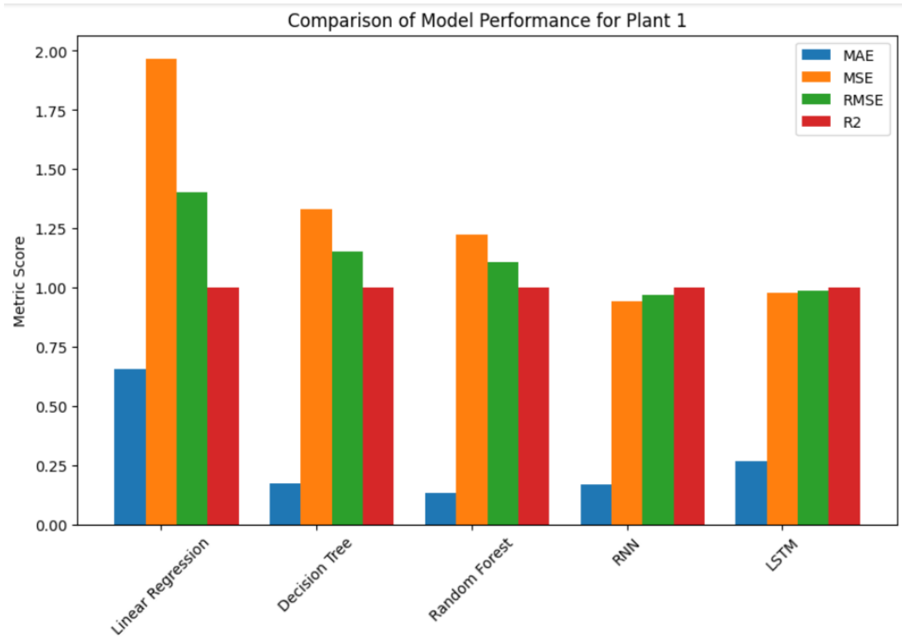

Figure 12. Comparison of Models for Plant 1

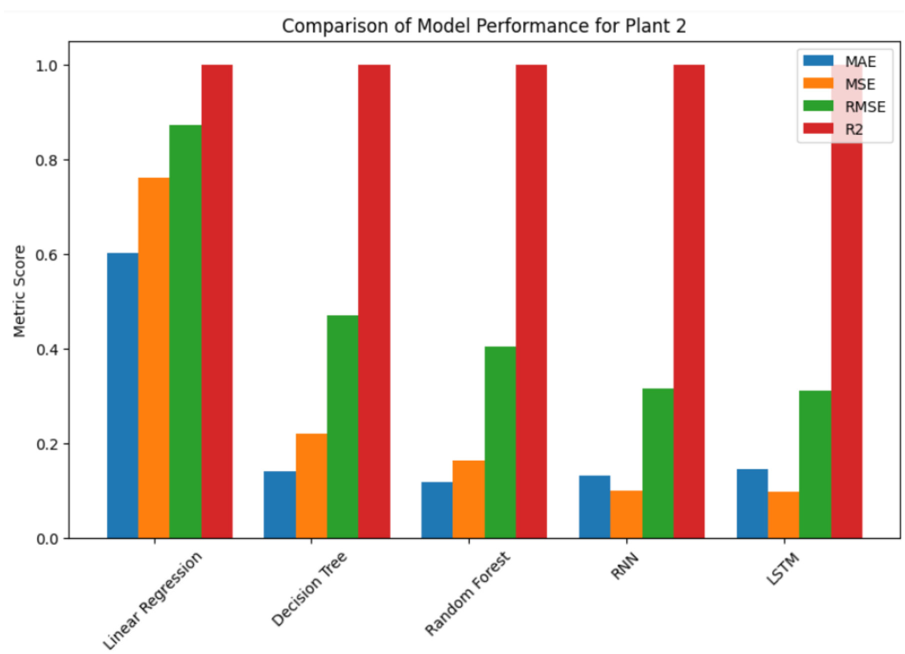

Figure 13. Comparison of Models for Plant 2

5 Conclusion

This project presents a comprehensive analysis of power generation and sensor data from two solar power plants in India, aiming to forecast energy output using machine learning algorithms. The results highlight the effectiveness of different models in predicting power generation for each plant. Specifically, the Recurrent Neural Network (RNN) model demonstrates superior performance for Plant 1, while the Long Short-Term Memory (LSTM) model outperforms others for Plant 2. Furthermore, the insights gained from this project underscore the potential of solar energy predictions to drive positive change in renewable energy management. Accurate forecasting can optimize energy production, reduce dependency on non-renewable resources, and contribute to a more sustainable future. By harnessing advanced machine learning techniques and leveraging data-driven insights, solar power plants can enhance operational efficiency, reduce costs, and pave the way for widespread adoption of renewable energy technologies.

Future endeavors in this domain could focus on integrating additional data sources, leveraging advanced machine learning techniques, fine-tuning model hyper parameters, implementing hybrid models, developing real-time forecasting systems, and conducting validation studies to further enhance the accuracy and applicability of solar energy forecasting models. Ultimately, this project underscores the potential of data-driven approaches in advancing renewable energy technologies and fostering sustainability in the face of evolving energy needs and environmental challenges.

References

1. **Younes Ledmaoui**, Forecasting solar energy production: A comparative study of machine learning algorithms, Sciencedirect, August 2023

2. **Mariam AlKandari**, Solar power generation forecasting using ensemble approach based on deep learning and statistical method, November 2019

3. **Xing Luo**,Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge, Sciencedirect, March 2021

4. **Meftah Elsaraiti**, Solar Power Forecasting Using Deep Learning Techniques, IEEE Access, March 2022

5. **Negar Rahimi**, A Comprehensive Review on Ensemble Solar Power Forecasting Algorithms, Journal of Electrical Engineering and Technology, January 2023

6. **K. Anuradha**, Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques, 2021

7.**Arjun Shah,Varun Viswanath**, Predicting Solar Energy Generation with Machine Learning based on AQI and Weather Features, July 2023

8. **C.Vennila**, Forecasting Solar Energy Production Using Machine Learning, Hindawi, April 2022

9. **Mohamed Abuella**, Solar Power Forecasting Using the Artificial Neural Networks, IEEE

10. **Keaobaka D. Potia**, Intelligent solar photovoltaic power forecasting, Sciencedirect, September 2023

11. **Hammad Ali Khan**, Solar Irradiance Forecasting Using Deep Learning Techniques, 2023

12. **Hangxia Zhou**, Deep Learning Enhanced Solar Energy Forecasting with AI-Driven IoT, Hindawi, June 2021