A Novel Approach in Machine Learning for Solar Energy Prediction System

Supriya Vaddi [1], Fathima Sumreen[2], Vaishnavi Narayanam[3], Akshara Patlannagari[4], Peruru Ananya[5]

[1] Associate Professor, Department of Information Technology, GNITS, Hyderabad, India

[2, 3, 4,5] Department of Information Technology, GNITS, Hyderabad
supriya-vaddi@gnits.ac.in

Abstract

India, a rapidly growing economy with a population exceeding 1 billion, faces a significant demand for energy. Despite the increasing population, power generation in the country has also seen a rise. To address India's escalating energy requirements', harnessing solar energy emerges as the most suitable solution, taking advantage of the country's favorable geographical location. However, effectively utilizing the solar energy generated presents challenges. While organizations invest effort in generating solar energy, they often neglect the importance of utilizing it optimally. This oversight can result in financial losses for organizations, as they invest substantial amounts in setting up solar energy systems but fail to monitor their utilization. The main objective of this project is to analyze the usage of solar energy and predict the amount of energy that our college, G. Narayanamma Institute of Technology and Sciences, would generate. The project employs four distinct algorithms, namely Multiple Linear Regression, Decision Trees, Gradient Boost, and XGBoost. We have evaluated these algorithms and identified the most accurate model based on their performance.

Keywords: Machine Learning, Solar Energy, Multiple Linear Regression, Decision Trees, Gradient Boost, XG Boost.

1 Introduction

Solar power is a renewable and environmentally friendly energy source that holds significant potential for combating climate change and achieving sustainable energy practices. Accurate forecasting of solar energy production is crucial as it plays a vital role in optimizing the integration of solar power into the power grid and ensuring a reliable energy supply. Precise forecasts enable effective management of energy generation, storage, and distribution, thereby enhancing grid stability and reducing reliance on traditional energy sources. Utilities need to plan effectively to maintain grid stability as solar power penetration increases. This study focuses on forecasting solar energy production for a period of fourteen days or two weeks. Our research specifically examines the prediction of solar energy using carefully selected small-to-medium-sized solar panels installed at our institution, G. Narayanamma Institute of Technology and Science. We analyze the energy generated by Polycrystalline solar panels that were reinstalled in 2019.

The combined capacity of the 81 solar panels is 30 kWh, with each panel connected to a 160-kW inverter. Rooftop solar energy is gaining popularity, and precise forecasting is essential to fully leverage the benefits it offers. Our study employs machine learning techniques that utilize publicly available weather predictions to forecast solar intensity. By utilizing these predicted sun intensity values, we can estimate the amount of solar energy generated based on the assumption that intensity directly correlates with energy production. This approach offers several sustainable economic advantages, including potential subsidies on electricity bills through surplus energy supply to the grid. Such benefits contribute to the economic viability, sustainability, and growth of the energy sector.

2 Literature Survey

The paper titled "A Hybrid Approach of Solar Power Forecasting Using Machine Learning" [1] investigates solar power forecasting by employing multiple machine learning models and analyzing various weather parameters. The accuracy of these models is evaluated using historical weather data. This approach combines pipelining, clustering, classification, and regression algorithms to develop a unique PV forecasting model. By grouping data points with similar weather conditions and utilizing segmented regression, the model leverages the next-day weather forecast to enhance prediction accuracy. The hybrid model, which integrates clustering, classification, and regression techniques, surpasses the performance of the random forest regression model, demonstrating improved forecasting capabilities under specific weather conditions.

In another study [2], a mathematical approach is introduced to predict solar energy generation for the upcoming seven days. This model takes into account factors such as ambient temperature, solar irradiance, panel efficiency, and plant component efficiency by considering weather data and plant specifications. The analysis provides insights into how temperature, irradiance, plant design, and efficiency affect the output of a photovoltaic (PV) system. However, the model's vulnerability to uncertainties arising from weather conditions is a concern.

Additionally, a research paper [3] presents an artificial neural network model for solar power forecasting. The study conducts a sensitivity analysis of different input variables to determine the optimal selection and compares the model's performance with multiple linear regression and persistence models. The objective is to generate hourly solar power forecasts for a month-long timeframe. The European Center for Medium-Range Weather Forecasts (ECMWF) serves as the source for 12 independent variables used to predict the target variable, solar power.

3 Dataset

Data from the college portal 'Solaris' was collected between May 2022 and January 2023 to monitor the generation of solar energy. The solar energy produced by the panels was recorded at 5-minute intervals on a daily basis, along with various attributes such as voltage, current, total power, frequency, and power grid total apparent power. The collected data was then aggregated on an hourly basis. In conjunction with the solar energy dataset, weather data from the same time intervals was obtained using the Visual Crossing open weather data API.

4 Methodology

The proposed system aims to forecast solar energy production for the next 14 days. To build the system, various algorithms including Multiple Linear Regression, Decision Tree, Gradient Boost, and XG Boost were compared.

Figure 1 illustrates the system's architecture, which involves the use of solar panels to generate energy that is then sent to an inverter. Data related to the solar panels is stored by a data logger and sent to the solar energy prediction system along with weather data obtained from an external website. The system creates a machine-learning model and displays the forecasted solar energy for the selected date on the user interface (UI). Users interact with the system through the UI.

Several assumptions were made for the analysis. Since the college collects data only from one building with solar panels, it is assumed that the energy generated by panels in other buildings is similar, as they were installed at the same time. The accuracy of the weather data collected from the Visual Crossing Weather API is also crucial for the accuracy of the outcomes.

In the data pre-processing phase, irrelevant attributes such as inverter status code, inverter serial number, data logger serial number, alert details, alert code, and grid power factor were removed. The 5-minute time intervals in the solar dataset were converted to 1-hour intervals to align with the weather data. Outlier analysis was conducted to identify and exclude days with fewer than the average number of recorded time intervals in a day. The common time intervals during which solar energy was generated, specifically from 6:00:00 to 19:00:00 were identified. Finally, the weather dataset was merged with the solar dataset.

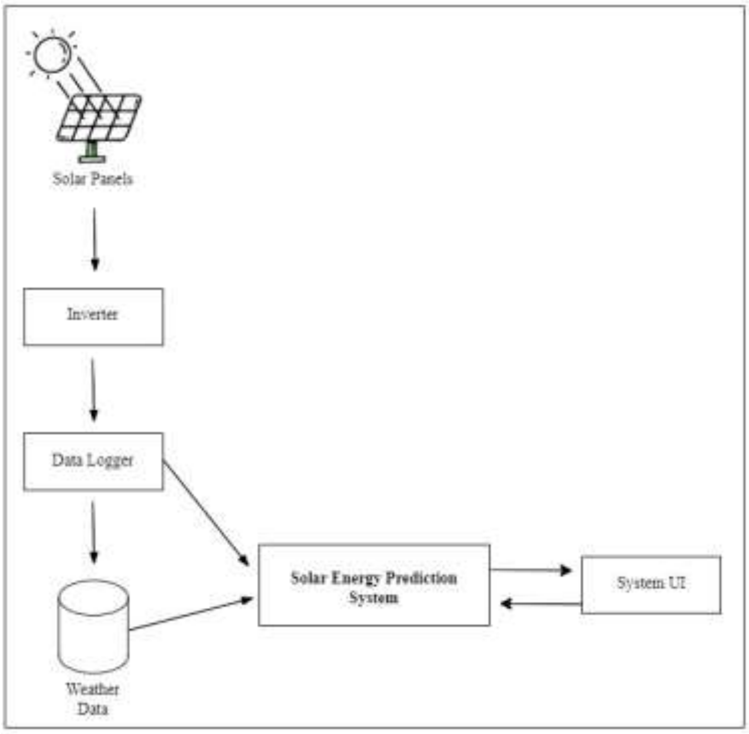

Figure 1: Architecture for Solar Energy Prediction System

Feature engineering: In our analysis, we utilized both correlation matrix and mutual information score to identify the relevant attributes. The correlation matrix allowed us to measure the linear relationships between each attribute, while the mutual information score helped us identify non-linear relationships. After conducting these analyses, we determined that the important attributes for our study included time, temperature, dew, humidity, wind direction, cloud cover, visibility, and solar radiation.

Model selection: To predict the daily solar power generation, we employed various machine learning models, namely multiple linear regression, decision tree, gradient boost, and XGBoost. These models were selected based on their compatibility with the dataset and their ability to forecast solar energy up to 14 days in advance. Ultimately, we chose XGBoost as our preferred model due to its highest accuracy, reaching approximately 81%.

Performance metric: We evaluated the performance of our models using two commonly used evaluation metrics for regression models: R-squared and Mean-Square Error. These metrics allowed us to assess the goodness of fit and the accuracy of predictions. After considering these metrics, we selected the model with the highest R-squared value and the lowest mean-squared error.

Prediction: Our model is designed to forecast the daily solar power generation for the next 14 days, starting from the current day. This prediction is based on obtaining forecasted weather data through an API call. Additionally, we developed a user interface using HTML, CSS, and Flask to effectively display the forecasted results.

5 Results and Discussion

To assess the models' performance, we initially divided the dataset into 75% for training and 25% for testing. We then utilized the R-squared value and Mean-Square error to evaluate the models' performance. The R-squared value, as defined by equation 1, quantifies the proportion of variance in the dependent variable that can be attributed to the independent variable.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \overline{y_i})^2} \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1}$$

Mean Squared Error was additionally used to compare the models further. The mathematical formula of Mean Square Error is given below by equation 2. Mean Squared Error assesses the average squared difference between the observed and predicted values.

The following table summarizes our obtained result.

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2 \tag{2}$$

Table 1. Result from multiple algorithms

S. No	Algorithm	R Squared Value	Mean Square Error
1	Multiple Linear Regression	0.7704	393.6623
2	Decision Tree	0.7881	363.3103
3	Gradient Boost	0.7721	390.7736
4	XGBoost	0.8168	314.4238

According to the provided R-squared values, the XGBoost algorithm demonstrates the highest R-squared value, indicating the best fit to the data compared to the other algorithms considered. The Decision Tree algorithm performs the second best, followed by Multiple Linear Regression and Gradient Boost. The superior performance of XGBoost can be attributed to its capability to enhance weaker learners through effective parallel computing methods.

6 Conclusion & Future scope

The paper is aimed at predicting the solar energy generated using machine learning algorithms. We found out that XGBoost gives the best performance with an accuracy of 81%. However, this study emphasizes on the different weather attributes and their importance in solar energy generation and one of the vital limitations of the project is that it depends upon the weather forecasted by the third-party website. Hence more the reliability of the weather data, higher is the accuracy of the project. The major limitation of this project is the limited dataset. Accuracy can be further improved with the help of larger dataset. This will also ensure better prediction results. The project can be extended by using a hybrid model to categorize similar weather patterns. We also believe that model fitting upon the clusters of data can lead to a better and higher accuracy in performance. Automating the whole system can make it effective for real real-time.

References

1. A. Bajpai and M. Duchon, "A Hybrid Approach of Solar Power Forecasting Using Machine Learning," 2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC), 2019.
2. D. Solanki, U. Upadhyay, S. Patel, R. Chauhan and S. Desai, "Solar Energy Prediction using Meteorological Variables," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018.
3. M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," 2015 North American Power Symposium (NAPS), 2015.
4. M. Yesilbudak, M. Çolak and R. Bayindir, "A review of data mining and solar power prediction," 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, UK, 2016, pp. 1117-1121, doi: 10.1109/ICRERA.2016.7884507.
5. K. A. Dotche, A. A. Salami, K. M. Kodjo, Y. P. C. D. Blu and Y. E. J. Diabo, "Evaluating Solar Energy Harvesting using Artificial Neural Networks: A Case study in Togo," 2019 II International Conference on High Technology for Sustainable Development (HiTech), Sofia, Bulgaria, 2019, pp. 1-5, doi: 10.1109/HiTech48507.2019.9128285.
6. A. Dhage, A. Kakade, G. Nahar, M. Pingale, S. Sonawane and A. Ghotkar, "Recommendation and Prediction of Solar energy consumption for smart homes using machine learning

algorithms," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-5, doi: 10.1109/AIMV53313.2021.9670909.

7. Rahul, A. Gupta, A. Bansal and K. Roy, "Solar Energy Prediction using Decision Tree Regressor," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 489-495, doi: 10.1109/ICICCS51141.2021.9432322.
8. A. Tuohy et al., "Solar Forecasting: Methods, Challenges, and Performance," in IEEE Power and Energy Magazine, vol. 13, no. 6, pp. 50-59, Nov.-Dec. 2015, doi: 10.1109/MPE.2015.2461351.
9. K. Pun, S. M. Basnet and W. Jewell, "Solar Power Prediction in Different Forecasting Horizons Using Machine Learning and Time Series Techniques," 2021 IEEE Conference on Technologies for Sustainability (SusTech), Irvine, CA, USA, 2021, pp. 1-7, doi: 10.1109/SusTech51236.2021.9467464.
10. Kim, Seul-Gi, Jae-Yoon Jung, and Min Kyu Sim. 2019. "A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning" Sustainability 11, no. 5: 1501.
11. Mohammed, Azhar & Aung, Zeyar. (2016). Ensemble Learning Approach for Probabilistic Forecasting of Solar Power Generation. Energies. 9. 1017. 10.3390/en9121017.
12. Gensler-Janosch, A., et al. “Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks.” 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016.
13. Hammer, A., et al. “Short-Term forecasting of solar radiation: a statistical approach using satellite data.” Solar Energy, vol. 67, no. 1-3, 1999, pp. 139–150., doi:10.1016/s0038-092x(00)00038-4.
14. Mellit, Adel. “Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review.” International Journal of Artificial Intelligence and Soft Computing, vol. 1, no. 1, 2008, p. 52., doi:10.1504/ijaisc.2008.021264.
15. Perera, Kasun S., et al. “Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey.” Data Analytics for Renewable Energy Integration Lecture Notes in Computer Science, 2014, pp. 81–96., doi:10.1007/978-3-319-13290-7 7
16. Srivistava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. ”Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” Journal of Machine Learning Research 15 (2014).