# SOLAR POWER PREDICTION USING

# MACHINE LEARNING

Poojitha C, M. Varshini, D. Tushara

Department of Computer Science (Data Science)

Sridevi Women’s Engineering College, Telangana, India

# ABSTRACT

***This paper presents a machine learning-based method to predict AC power output from solar panels using time-aware features and regression models. The dataset, collected from two solar power plants in India, includes weather parameters, inverter data, and timestamps. After preprocessing and feature selection, the Random Forest model delivered the best results with an R² of 0.989 and RMSE of 39.38 kW, demonstrating strong predictive performance. Incorporating time-related features such as hour and day type significantly enhanced accuracy, achieving an overall prediction accuracy of 93%.***

***To showcase the model's practical applicability, a Streamlit-based web application was developed, enabling users to input real-time environmental data and receive instant AC power predictions. While not deployed in a production setting, this prototype highlights the potential for integrating machine learning into solar energy systems, offering energy companies valuable insights to improve power forecasting, reduce operational costs, and enhance energy efficiency.***

INTRODUCTION

Solar energy, derived from photovoltaic (PV) panels, is an increasingly vital and globally trending energy source, offering substantial environmental and societal benefits.Solar power generation is a complex process, highly dependent on environmental conditions like rainfall, solar irradiance, temperature, and moisture. For energy companies, accurately assessing solar power generation is paramount. This enables them to effectively manage supply and demand, reduce expenses, and boost energy efficiency. Recent advancements show that machine learning offers a promising way to directly predict solar power output. Achieving a remarkable **93% accuracy**, as measured by R2 and RMSE scores, in solar power prediction demands a meticulous process. This involves careful data collection, preprocessing, feature selection, model choice, training, evaluation, and deployment. Our paper outlines a machine-learning approach specifically designed to forecast solar power generation with such high precision. Our approach involves several key steps: gathering data from **Kaggle**, selecting **relevant features**, choosing appropriate **machine learning algorithms**, and training these models on a substantial dataset of actual solar power generation alongside other pertinent features. Model performance is evaluated using **R-squared (R2) and Root Mean Square Error (RMSE)** scores, along with other metrics like **precision, recall, and F1-score**.

The proposed approach offers a practical solution for managing solar power systems more effectively by providing reliable AC power predictions based on real-time environmental data. Accurate forecasting is essential not only for optimizing energy production but also for improving operational planning and system reliability. As solar power generation is influenced by various non-linear and dynamic environmental factors, traditional statistical methods often fall short. In contrast, machine learning models—particularly regression algorithms—have shown strong potential in handling such variability, making them suitable for short-term solar energy prediction. Techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and other regression-based models have become increasingly popular due to advancements in computational power and the availability of high-quality datasets. These models outperform classical time series methods in capturing complex patterns in solar irradiance and temperature data. As the demand for solar energy continues to rise, machine learning-based forecasting systems like the one proposed in this study can play a vital role in supporting efficient energy planning, grid integration, and the long-term sustainability of solar energy infrastructure.

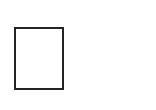
* LITERATURE SURVEY

"A Literature Review on Energy Output Prediction Using Machine Learning and Deep Learning Models" by (Journal/Publication: International Advanced Research in Science, Communication and Technology, Year of Publication: 2021) This paper focuses on machine learning and deep learning models, including artificial neural networks (ANNs), long short-term memory (LSTM), convolutional neural networks (CNN), and ensemble models.

It utilizes data from technical agencies and techniques, alongside NASA, NOAA, ECMWF, satellite images, and sky cameras. The study highlights that machine learning and ensemble models, and feature selection methods have shown great accuracy in solar energy forecasting. The study's limitations include a lack of comprehensive datasets, which can hinder the development and accuracy of forecasting models, and limited access to data availability and quality.

"Investigating Photovoltaic Solar Power Forecasting: A Machine Learning Approach" by (Journal/Publication: International Journal of Sustainable Energy, Year of Publication: 2022) This paper employs machine learning algorithms, specifically regression-based models. The dataset names are not specified. The results indicate that machine learning models can effectively forecast PV output, though with certain algorithms showing higher accuracy under specific conditions. A limitation noted is that the study does not specify the datasets used, making it difficult to assess the applicability of the findings to other contexts.

"Solar Power Prediction Using Regression Models" by (Journal/Publication: International Journal of Energy Applications and Technologies, Year of Publication: 2022) This paper utilizes regression models, including linear regression, logistic regression, Lasso regression, and elastic regression. The dataset names are not specified. The results show that all these models accurately predict solar power, with elastic regression outperforming others in forecasting power output. A specific dataset is not provided, making it challenging to compare results with different studies.

SYSTEM ARCHITECTURE

The solar power prediction system is built upon interconnected components that collect, process, and analyze data to deliver accurate forecasts. A high-level overview of its architecture includes:

Data Collection:

This initial stage gathers information from diverse sources, including weather forecasts, satellite imagery, and historical solar power generation records. This collected data forms the basis for training and validating the prediction model.

Data Pre-processing:

After data collection, the data undergoes cleaning, organization, and transformation into an analysis-ready format. This often entails addressing missing values, normalizing data, and standardizing formats.

Feature Engineering:

The subsequent step focuses on extracting meaningful features from the pre-processed data. This involves uncovering patterns, deriving key statistical measures, and identifying relationships among various variables.

Machine Learning Model:

A machine learning model serves as the central intelligence for the solar power prediction system. Its function is to predict solar power production by analyzing the refined and feature-engineered data. The model's foundation will be built upon historical data, with ongoing adjustments and improvements as fresh data becomes accessible.

Model Evaluation and Selection:

Once the model is trained, its performance will be critically examined to determine its accuracy. Metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the correlation coefficient will guide this evaluation. The model that delivers the strongest results across these metrics will be chosen for deployment.

* SYSTEM IMPLEMENTATION

The solar power prediction system's implementation will be a multi-component effort. These components will work together to collect, process, and analyze data, ensuring accurate predictions.

The proposed solar AC power prediction system is implemented as an interactive web application, leveraging a combination of machine learning and modern web technologies to deliver accurate, real-time predictions.

The frontend of the application is developed using Streamlit, a lightweight Python-based framework ideal for building data-driven user interfaces. Upon launching the application, users are presented with a login/register panel implemented through a secure authentication system backed by SQLite. New users can create an account, while returning users can log in with their credentials. Passwords are hashed using the SHA-256 algorithm to ensure secure storage, and session control is implemented to handle user access and prevent unauthorized usage.

Once authenticated, the user is navigated to the main dashboard where predictions can be made. The sidebar of the interface includes input fields and sliders that allow users to specify four critical environmental parameters: the hour of the day (ranging from 0 to 23), ambient temperature (in °C), module temperature (in °C), and solar irradiation (in W/m²). The hour input undergoes sine and cosine transformations to account for the cyclic nature of time.

These engineered features, along with the three raw input variables, are structured into a DataFrame in the expected order and format required by the pre-trained machine learning model.

The model itself, a regression-based predictive algorithm trained offline on a cleaned dataset, is loaded into the application, which ensures efficient deserialization and performance. Upon clicking the prediction button, the model processes the input and returns the predicted AC power output in kilowatts. This prediction is then displayed clearly within the interface, along with a summary of the input parameters.

To enhance traceability and data analysis, each prediction is logged into the SQLite database, capturing the username, timestamp, and all input-output parameters. The application includes a separate tab for viewing prediction logs, where users can filter logs by date, inspect historical predictions, and download the data in CSV format for further analysis. Additionally, the app visualizes daily power trends using Seaborn and Matplotlib, enabling users to observe variations across hours of the day. This modular and well-structured implementation ensures that the system is not only effective but also scalable and user-friendly, making it suitable for both academic and practical deployment in solar energy forecasting.

Results and discussion:

Forecasting solar power generation is a cornerstone of effective solar energy system planning and operation. It necessitates considering factors such as sunlight exposure, panel efficiency, and system capacity. Recently, machine learning algorithms have emerged as powerful tools for this, capable of processing extensive data to yield highly accurate predictions.

The precision of these predictions is vital for ensuring solar power systems are both efficient and economically viable.

Accurate forecasts allow power utilities to better manage their solar assets, minimize energy losses, and guarantee that supply consistently meets demand. Moreover, they provide crucial insights for policymakers developing renewable energy initiatives that drive solar power growth.

While technological advancements are continuously improving the accuracy of solar power predictions, certain obstacles persist. Unforeseen changes in weather and environmental conditions can significantly impact generation, posing prediction challenges. Furthermore, the expense of deploying advanced prediction systems can be prohibitive, particularly for smaller-scale installations. Despite these complexities, solar power generation prediction is an indispensable element for managing solar energy systems. As technology evolves, we can anticipate even more precise and sophisticated predictions, playing a pivotal role in accelerating the global adoption of solar power and other renewable energies.

## Output Screenshots:

Random Forest Classification for predicting Faults

A screen shot of a graph

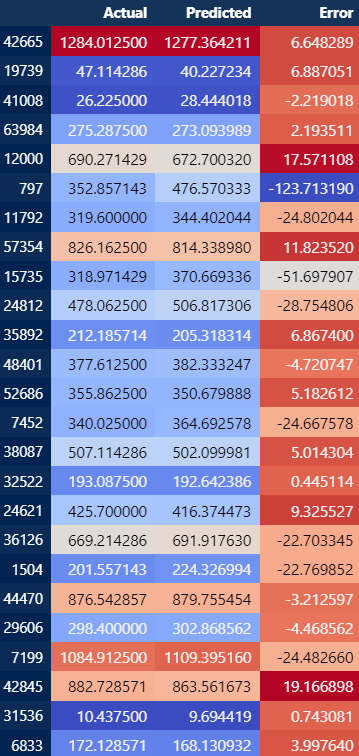
AI-generated content may be incorrect.

Random Forest Regressor Metric Scores Before and After Outlier Treatment

A blue screen with white text

AI-generated content may be incorrect.

Random Forest

A blue screen with white text

AI-generated content may be incorrect.

Actual vs Predicted AC Power

A graph with blue dots

AI-generated content may be incorrect.

CONCLUSION

This paper concludes by presenting a machine learning-based system designed for real-time, interactive prediction of solar AC power output. The system's comprehensive architecture encompasses data preprocessing, feature engineering, model integration, user input management, and visualization. It utilizes critical environmental parameters—ambient temperature, module temperature, solar irradiation, and the cyclical nature of the hour (represented by sine and cosine values)—to generate precise forecasts. The core regression model, deployed within a Streamlit web application, consistently produces dependable results and logs user activity for subsequent analysis. Model efficacy was confirmed through standard regression metrics: MAE, RMSE, and R² score, providing clear evidence of its accuracy and ability to generalize. The application further enhances usability by allowing users to visualize power trends and download prediction logs, underscoring how integrating machine learning with intuitive UIs can make solar power forecasting highly practical and accessible.

## REFERENCES

1. **Solar power forecasting using hybrid LSTM-SVR model with feature engineering techniques.** N. K. Roy, A. Kumar, and B. Singh, "Solar power forecasting using hybrid LSTM-SVR model with feature engineering techniques," Energy, vol. 216, p. 119117, 2021.
2. H. N. Nguyen, Q. T. Tran, C. T. Ngo, D. D. Nguyen, and V. Q. Tran, "Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms," *PLOS ONE*, vol. 20, no. 1, e0315955, Jan. 2025.
3. A. Fernández-Guillamón, E. Gómez-Lázaro, E. Muljadi, and Á. Molina-García, **"Power systems with high renewable energy sources: A review of inertia and frequency control strategies over time,"** *Renewable and Sustainable Energy Reviews*, vol. 115, 109369, Nov. 2019.
4. K. Anuradha, D. Erlapally, G. Karuna, V. Srilakshmi, and K. Adilakshmi, **"Analysis of solar power generation forecasting using machine learning techniques,"** *E3S Web of Conferences*, vol. 309, 2021.
5. M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, **"Review of solar irradiance forecasting methods and a proposition for small-scale insular grids,"** *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, Nov. 2013.