Solar irradiance prediction using Satellite Data and Machine Learning

MT2313 : Gaurav Prakash Lute,[1] Internal Guide : Dr. Bhalchandra Pujari[1]
External Guide : Dr. Abhaykumar Tiwari[2], Dr. Deepashri Saraf[2]
[1]Department of Scientific Computing, Modeling and Simulation, SPPU, Pune, India.
[2] Spatialty.Ai Bangalore, India.

Introduction

This research developed machine learning models to predict Global Horizontal Irradiance (GHI) a key factor in solar energy planning. Using Satellite data (2016–2025), we explored two goals: accurate daily GHI prediction and forecasting for March 2025. Advanced models like XGBoost, LightGBM, ARIMA, and Prophet were trained using engineered features such as lag values and cyclical patterns. Results reveal that XGBoost excelled in precision. This intelligent forecasting pipeline offers a scalable, low-cost solution for improving solar resource management in real-world energy systems.

Research objectives

The present study investigates the following objectives:

- **Objective 1:** To develop predictive models that estimate daily Global Horizontal Irradiance (GHI) based on historical satellite meteorological features.
- **Objective 2:**To forecast GHI for March 2025 using time series approaches such as ARIMA and Prophet.

The ultimate goal is to enhance short-term solar radiation forecasting accuracy using data-driven models, which can aid in efficient solar energy management.

Dataset Description

The dataset used in this project is sourced from the Satellite Dataset (Prediction of Worldwide Energy Resources) API. The selected location is **Pune, India** (Lat: 18.52°N, Lon: 73.86°E), covering the period from February 2016 to February 2025.

Target variable:

- ALLSKY_SFC_SW_DWN (Global Horizontal Irradiance – GHI)

Features extracted:

- Temperature at 2m, Relative Humidity, Wind Speed at 10m
- Cloud Cover, Surface Pressure, Precipitation
- Derived: Day of Year, Month, Lag Features, Rolling Mean

Methodology

Solar irradiance Prediction Workflow

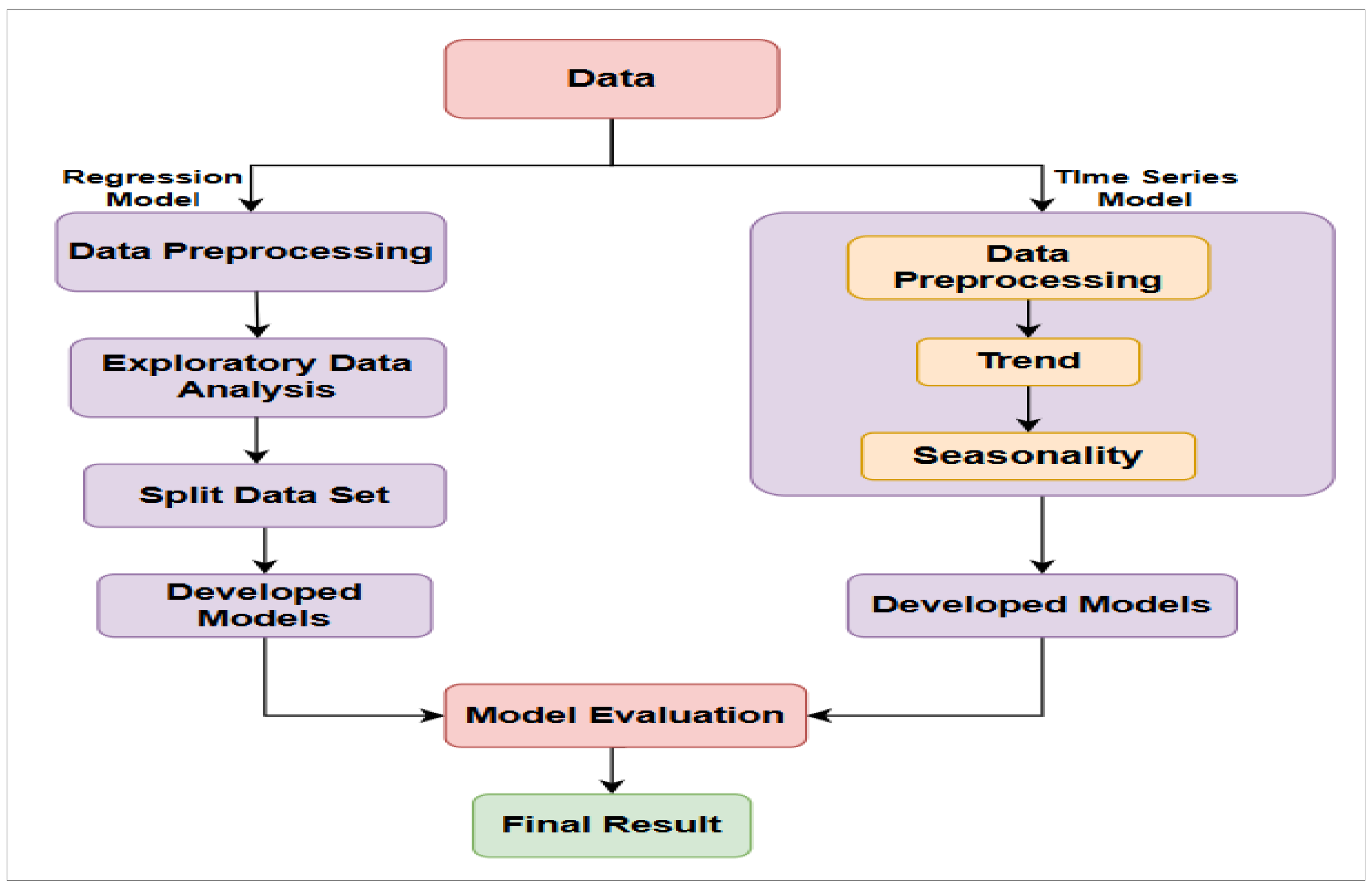


Figure 1. Overall workflow of solar irradiance prediction: from data preparation to model evaluation and forecasting.

Daily Irradiance Data

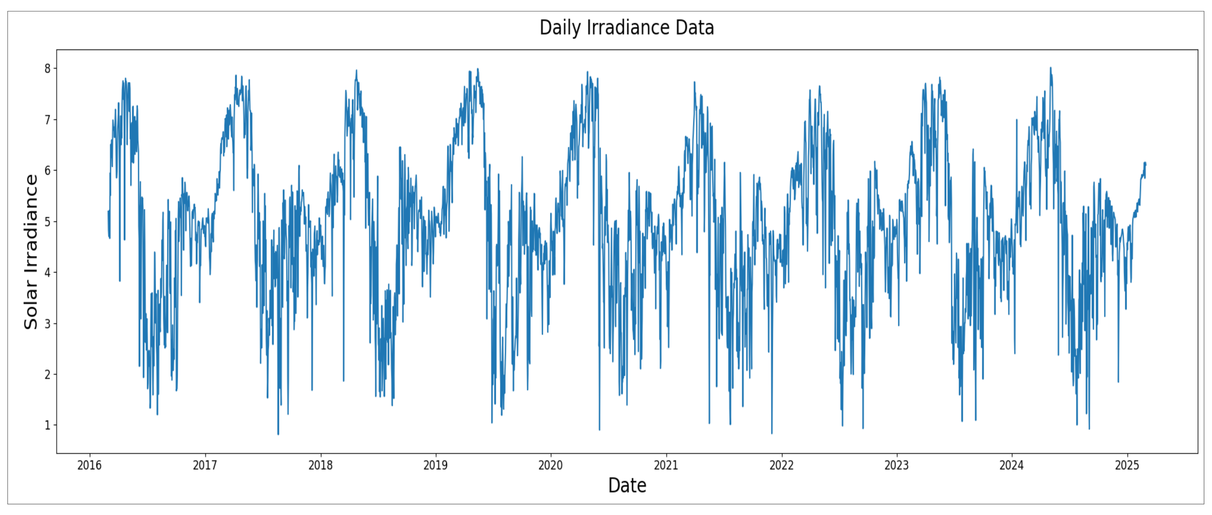
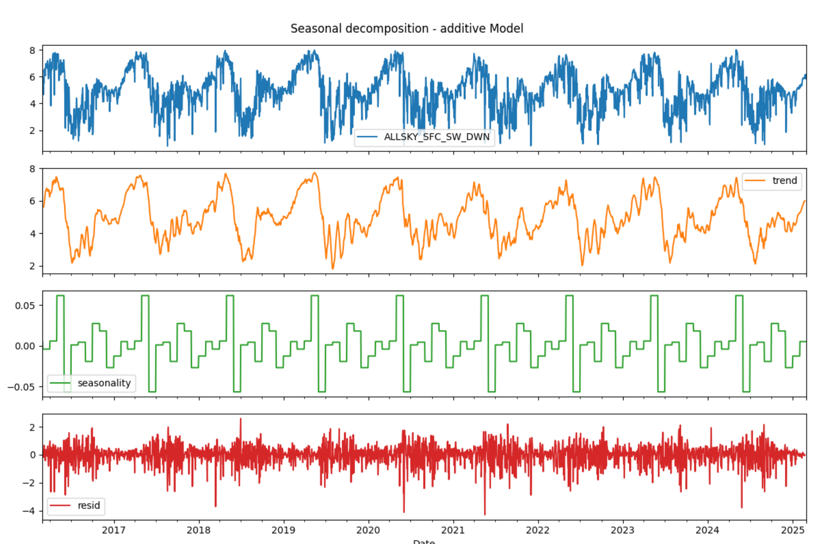


Figure 2. This line plot illustrates the daily variation of all-sky surface downward shortwave radiation over time, showing seasonal patterns and overall trends.

- **Trend** – Long-term upward or downward movement in GHI values.
- **Seasonality** – Repeating patterns caused by seasonal solar cycles.
- **Residual** – Random noise or irregular variations not explained by trend or seasonality.



Understanding Global Horizontal Irradiance (GHI)

Explanation: Global Horizontal Irradiance (GHI) represents the total solar energy received on a horizontal surface. It is the sum of:

- **Direct Irradiance** – sunlight reaching the surface directly from the sun.
- **Diffuse Irradiance** – sunlight scattered by clouds, air molecules, and particles.

Formula:

$$\mathrm{GHI} = \mathrm{DNI} \cdot \cos(z) + \mathrm{DHI}$$

Where:

- DNI: Direct Normal Irradiance
- DHI: Diffuse Horizontal Irradiance
- z: Solar zenith angle

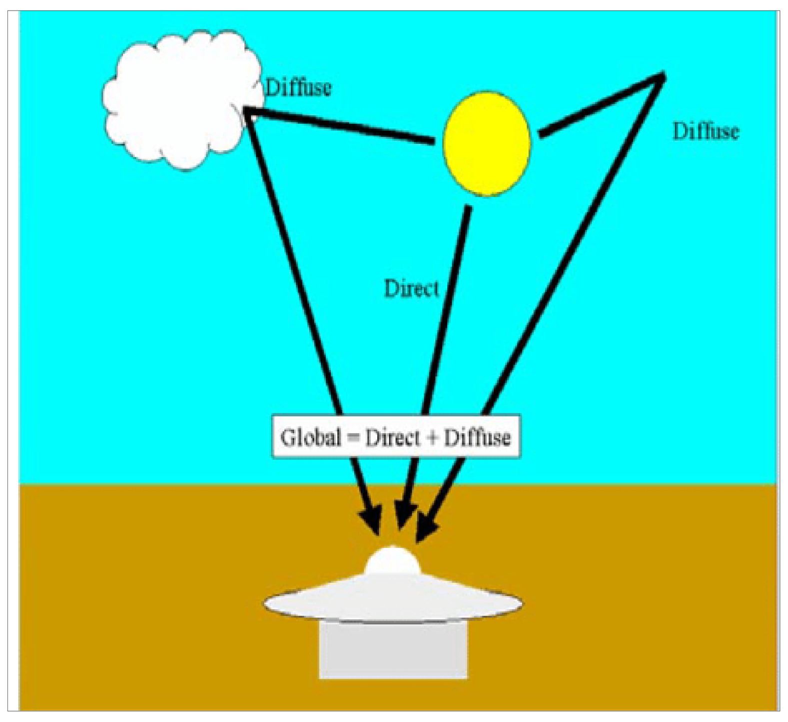


Figure 4. GHI consists of both direct and diffuse radiation.

Result and Discussion

Prediction Results (1-Month vs 3-Month)

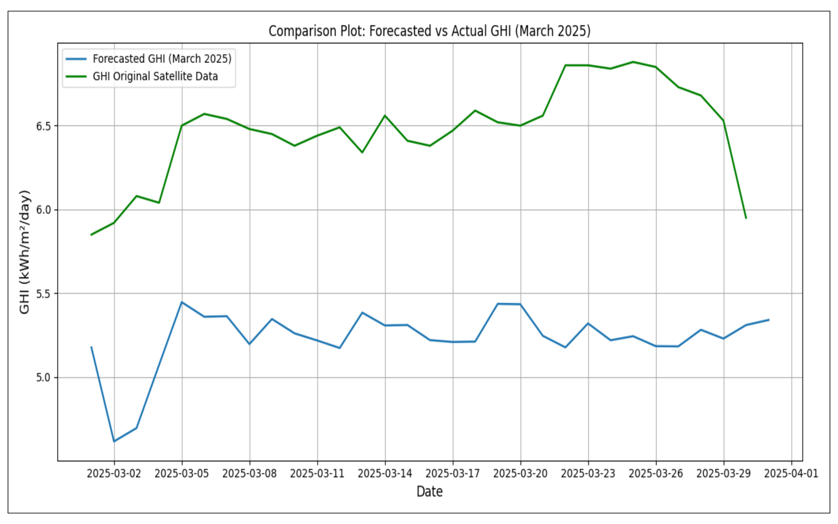


Figure 5. March 2025 Prediction (XGBoost)

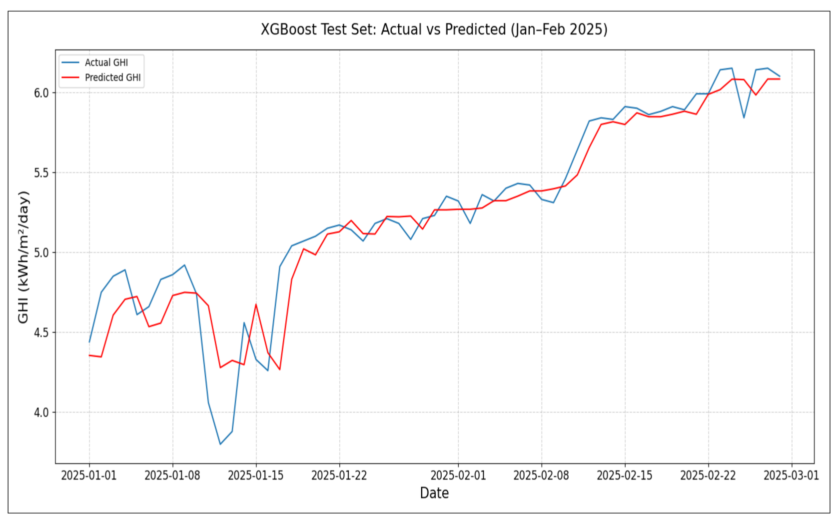


Figure 6. Three-Month Prediction (XGBoost)

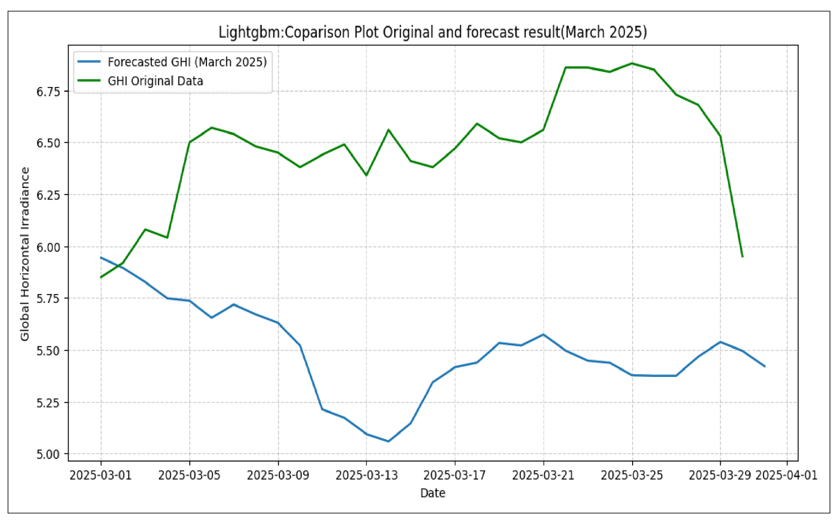


Figure 7. March 2025 Prediction (LightGBM)

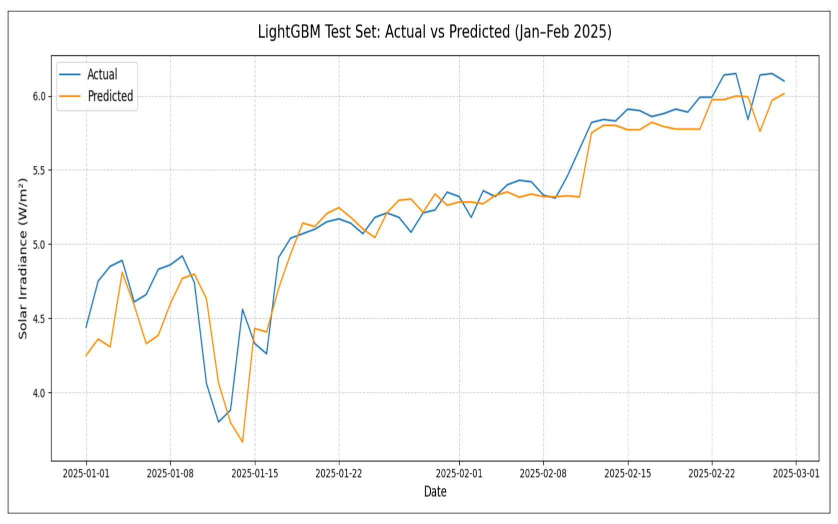


Figure 8. Three-Month Prediction (LightGBM)

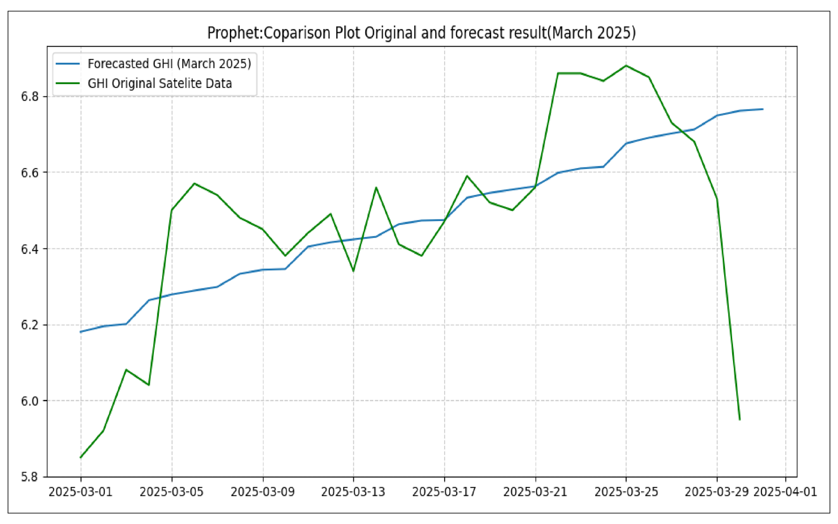


Figure 9. March 2025 Prediction (Prophet)

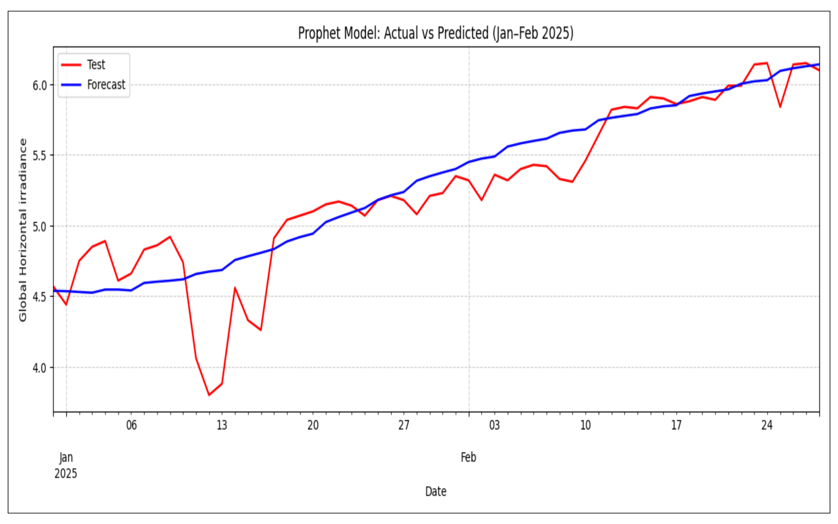


Figure 10. Three-Month Prediction (Prophet)

Model Evaluation Results

Objective 1: ML-based GHI Prediction

Model	MSE	R^2
Random Forest	0.39	0.98
Gradient Boosting	0.43	0.98

Performance metrics for ML models

Objective 2: Forecasting Models

Model	RMSE/MAPE	R^2
XGBoost	0.192 (RMSE)	0.891
LightGBM	0.050 (RMSE)	0.856
ARIMA	0.035 (RMSE)	–
Prophet	0.03 (MAPE)	–

Forecasting performance for March 2025

Conclusion

Machine learning models, especially XGBoost and Lightgbm, proved effective in accurately predicting solar irradiance from satellite data.Prophet Model provided seasonal insights but were less precise for short-term forecasts. This framework is well-suited for real-world solar energy planning and can be extended with deep learning and real-time deployment.

References