Short-Term Solar Energy Forecasting: Comparative Study of Statistical and Machine Learning Models

Author: Abhas Kumar

Institution: Birla Institute of Technology, Mesra

E-Mail: abhassingh246@gmail.com

1. Introduction

Accurate renewable energy forecasting is essential for grid stability and energy management. Solar generation exhibits strong daily seasonality and nonlinear variability driven by meteorological factors. This study compares statistical and machine learning models for short-term (hourly to 7-day ahead) forecasting.

2. Dataset Description

The dataset contains multi-year solar energy measurements recorded at 15-minute intervals, along with meteorological variables. The data was resampled to hourly resolution for computational feasibility. A rolling 60-day training window and 7-day test window were used.

3. Methodology

3.1 Persistence Model

Forecast equation: $\hat{y}(t) = y(t-1)$

This model captures strong short-term autocorrelation and serves as a baseline benchmark.

3.2 SARIMA Model

General SARIMA formulation:

$$\Phi(B^s)\, \varphi(B)\, (1-B)^d\, (1-B^s)^D\, y_t = \Theta(B^s)\, \theta(B)\, \varepsilon_t$$

Configuration used: (1,0,1)(1,0,1,24). This captures daily seasonal structure under linear assumptions.

3.3 Random Forest

A nonlinear ensemble regression model trained on cyclical hour encoding and meteorological features to capture complex weather-energy interactions.

3.4 Neural Network (MLP)

Neural network mapping:

$$\hat{y} = f(W_2\, \sigma(W_1 X + b_1) + b_2)$$

A feedforward network trained on 24-hour input windows to learn nonlinear temporal dependencies.

4. Results

Model	MAE	RMSE
Persistence	1067	1902
SARIMA (1,0,1)(1,0,1,24)	1555	2892
Random Forest	1168	2214
MLP Neural Network	887	1419

Figure 1: 7-Day Forecast Comparison

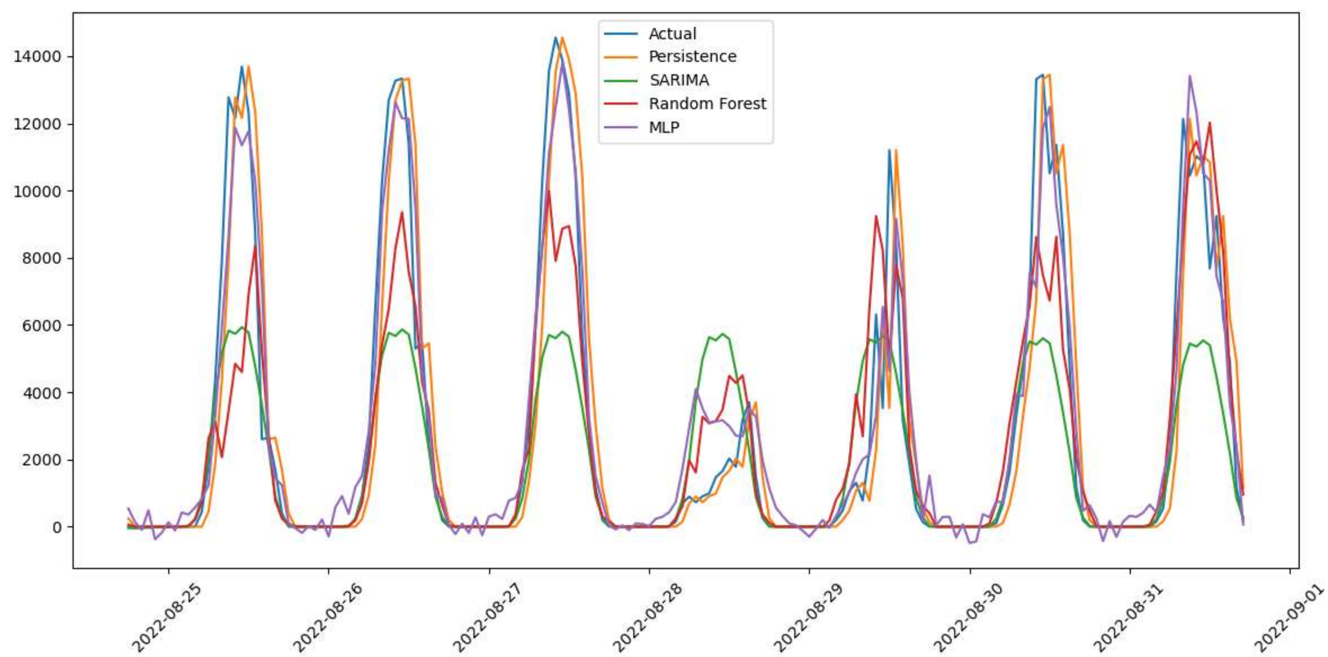
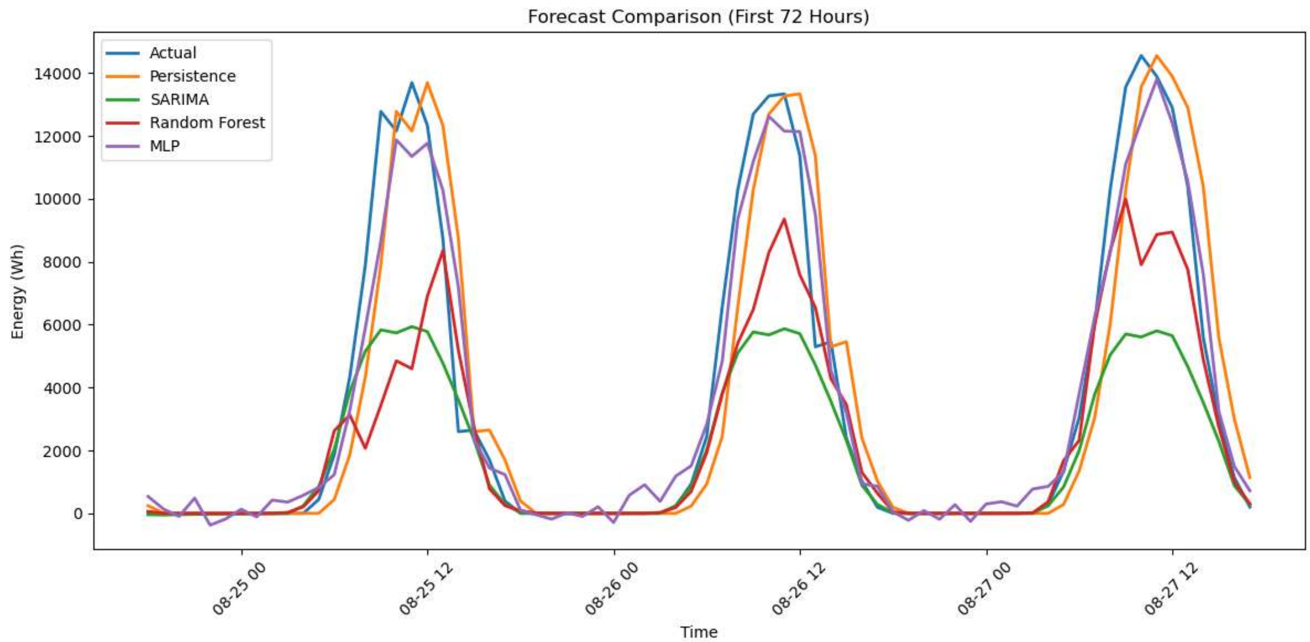


Figure 2: Zoomed 72-Hour Forecast Comparison



5. Discussion

The persistence model performs strongly due to high short-term autocorrelation. SARIMA underperforms because solar generation is highly nonlinear and weather-dependent. Random Forest captures nonlinear feature interactions but lacks explicit sequential modeling. The MLP neural network achieves the best performance by learning nonlinear temporal relationships across daily windows.

6. Conclusion and Future Work

This study demonstrates that nonlinear neural models outperform classical statistical approaches for short-term solar forecasting. Future work may include LSTM/GRU architectures, multi-horizon forecasting, and incorporation of weather forecast inputs.