Predicting Photovoltaic Power from Solar Irradiance using Supervised Machine Learning Models

Pragyan S. Shukla[1], Rushil S. Johal[2], Sharmin Abdullah[3] |pshukl@gmu.edu[1], rjohal@gmu.edu[2], sabdul9@gmu.edu[3]
Department of Computer Science[1], Department of Computational and Data Sciences[2, 3]| George Mason University

ABSTRACT

In this study, using two distinct datasets, we analyze the long-term performance of various supervised machine-learning (ML) regression models in predicting the photovoltaic (PV) power of solar panels from solar irradiance throughout 2017. Using primarily Python in JupyterLab, we systematically and visually compare models using error metrics to determine which provides the most accurate predictions for each dataset. Our results identify the best-fit model with the least error by simulating different models' performance compared to the actual PV Power outputs while providing future advancements in PV Power forecasting.

Figure 1: Solar Tech Lab in Milan, Italy (left) and Residential Rooftop Solar Panel in Utrecht, Netherlands (right)

INTRODUCTION

- Due to rising consumer interest and falling solar panel prices, PV systems drive renewable energy demand to reduce carbon emissions, mitigate climate change, and replace more costly, unreliable non-renewable sources (e.g. natural gas, fossil fuel) [1]
- PV systems enhance energy management and support grid stability to accelerate sustainable energy integration
- Our study addresses intermittent solar energy challenges during this integration (e.g. if PV systems are worth the ROI of installation and maintenance) by balancing power generation and load demand through reliable PV prediction [2]

METHODOLOGY

- Data Sets:
  Data #1 from IEEE DataPort and was part of a previously conducted research in Milan, Italy. [5]
  - **Parameters:** PV Power (W), Ambient Temperature (°C), Global Horizontal Irradiance (W/m²), Global Plane of Array Irradiance (W/m²), Wind Speed (m/s), Wind Direction (degrees) over the Time (CET)

  Data #2's PV Power variable from one Solar Panel system (ID005) from 175 systems provided in Zenodo. Weather variables are from Solcast AP Toolkit. [6]
  - **Parameters:** PV Power (W), Ambient Temperature (°C), Global Horizontal Irradiance (W/m²), Global Tilt Irradiance (W/m²), Direct Normal Irradiance (W/m²), Wind Speed (m/s), Wind Direction (degrees) over the Time (UTC)
- Cleaning data through filling in missing values, which was done by taking averages of known data points near missing values
- Frequency of data for was 30-minute (Data #1) and 5-minute (Data #2) intervals
- Outliers were eliminated by deleting data points (Data #1 only)
- Exploratory Data Analysis:
  - Scatterplots of All Irradiances and PV Power vs. Time
  - Correlation Matrix of All Variables
- Machine Learning Application:
  - **Data #1:** Linear, Polynomial, Decision Tree, SVR, and Random Forest
  - **Data #2:** All of the above in addition to Lasso, Ridge, and XG Boost
- Performance Analysis: MAE, MSE, RMSE, nRMSE, and R^2

EXPLORATORY DATA ANALYSES

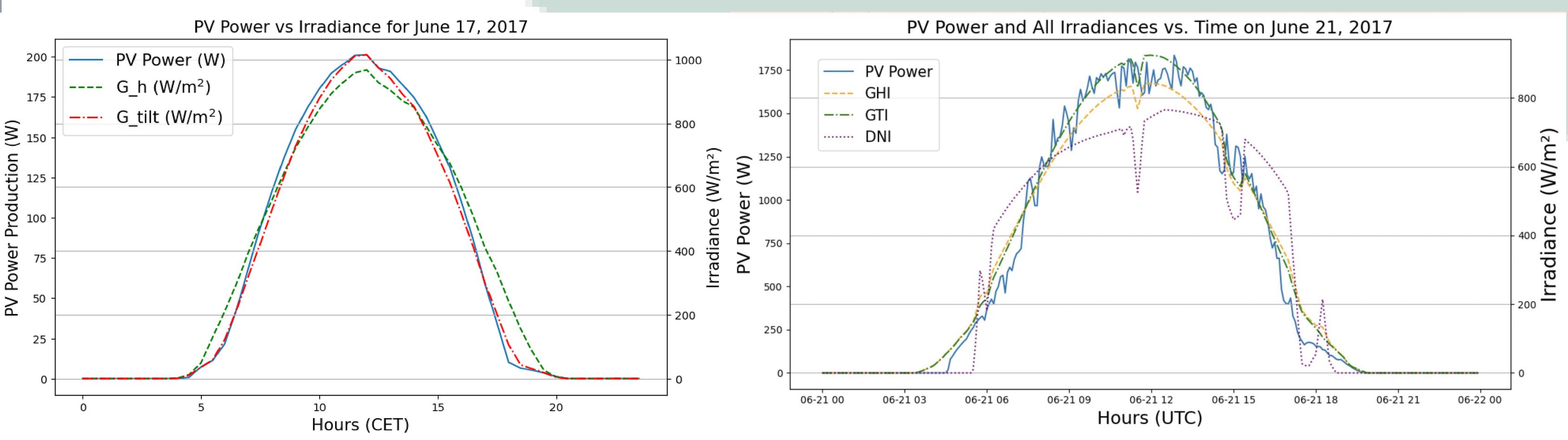

Figure 2: Line Plot of PV Power vs. Irradiance vs. Time (Data #1 left and Data #2 right)

EXPLORATORY DATA ANALYSES CONTINUED

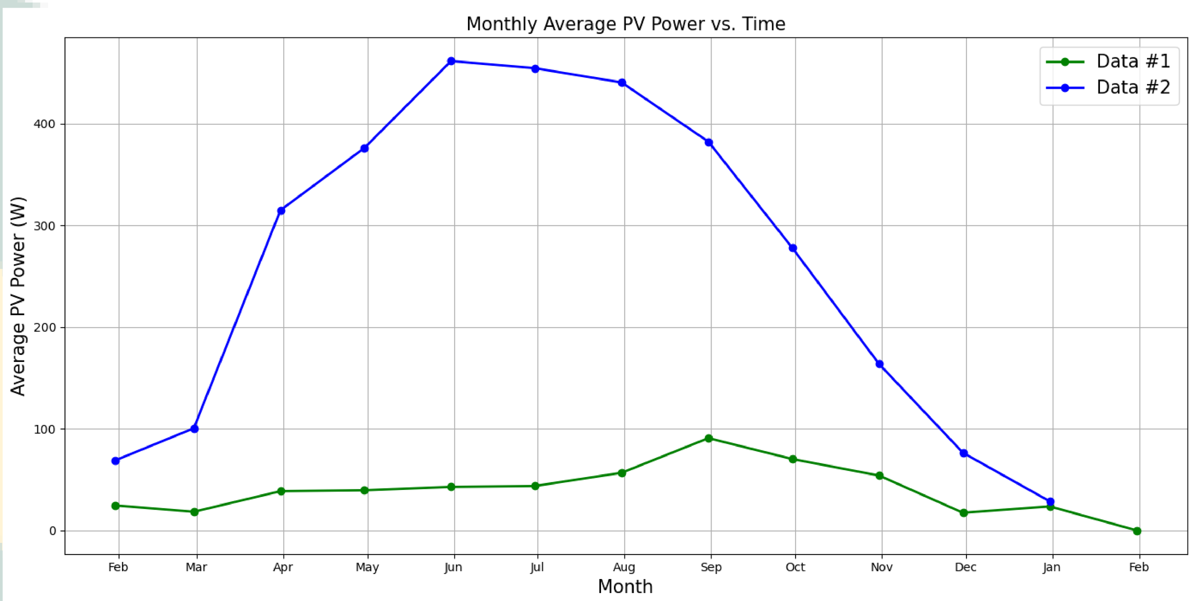

Figure 3: Monthly Averages of Real PV Power in Data # 1 and Data #2

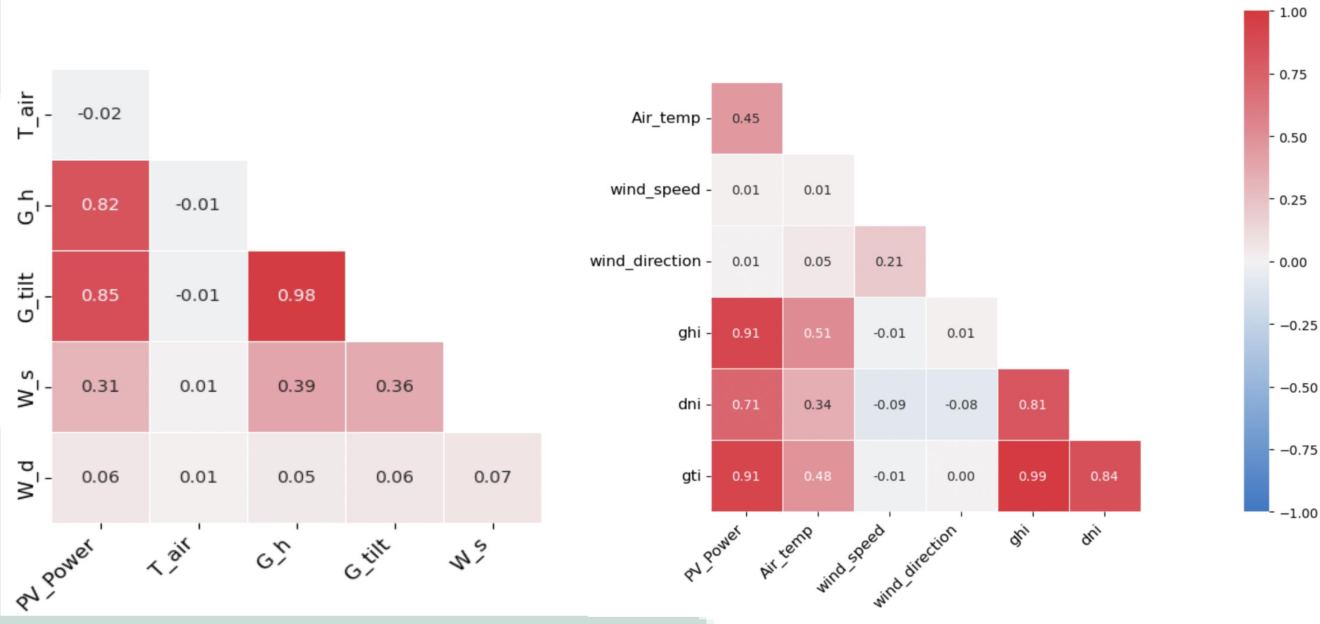

Figure 4: Correlation Matrix of All Variables throughout 2017 (Data # 1 left and Data #2 right)

OUTCOMES

- Calculated the correlation coefficient of each variable across Data #1 and #2 (*Fig 4);* PV Power vs. all the types of Irradiances (**Data #1**: G_tilt, G_h and Data #2: ghi, dni, gti) has the strongest linear relationship
- Plotted scatterplots of Actual vs. Predicted PV Power for each respective ML regression model (**Data #1** tested five models, **Data #2** tested eight models)
- While visualizing each model, we also calculated its error metrics (*Table 1*)
- By producing a predicted PV Power value for each model through training the actual values with "test_size=0.2" and "random_state=42", we plotted all models throughout 2017 to visually compare their behavior (*Fig. 5*)

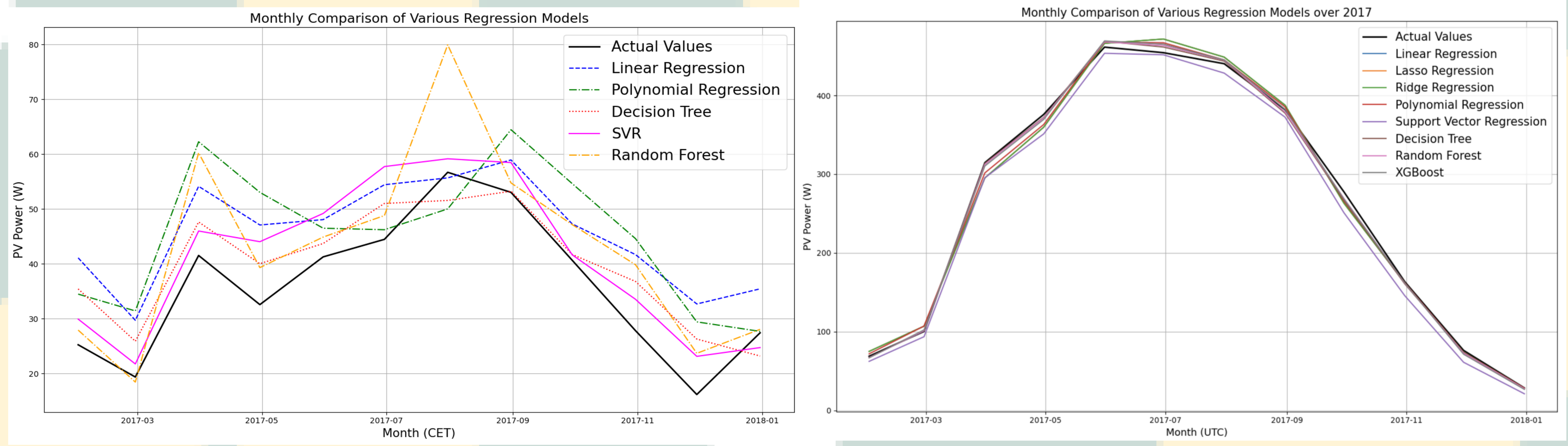

Figure 5: PV Power Prediction of All Regression Models (Data #1 left and Data #2 right)

ERROR METRICS

Model	MAE (W)		MSE (W)		RMSE (W)		nRMSE (%)		R^2	
	Data #1	Data #2	Data #1	Data #2	Data #1	Data #2	Data #1	Data #2	Data #1	Data #2
Linear	16.102	89.416	1807.218	37492.492	42.511	193.630	118.318	6.815	0.592	0.844
Polynomial	18.225	89.274	1945.138	37129.900	44.104	192.691	122.750	6.782	0.561	0.846
Decision Tree	17.349	98.176	2143.802	59025.886	46.301	242.952	128.866	8.551	0.516	0.755
Support Vector Regression	10.077	89.416	1719.425	37492.492	41.466	193.630	115.408	6.815	0.612	0.844
Random Forest	21.113	82.146	21799.186	39135.170	147.645	197.826	410.928	6.963	-3.924	0.837
Lasso Regression	N/A	89.414	N/A	37492.658	N/A	193.630	N/A	6.815	N/A	0.844
Ridge Regression	N/A	89.416	N/A	37492.492	N/A	193.630	N/A	6.815	N/A	0.844
XG Boost	N/A	80.599	N/A	37708.433	N/A	194.187	N/A	6.835	N/A	0.843

Table 1: Performance Metrics of All Regression Models

CONCLUSION

Data #1 (yellow):
- SVR has the lowest error and highest R2; SVR is the best-fit model in PV Power prediction (*Table 1*)
- All evaluation metric values have small error and variance
- SVR's regression line matches most closely with the actual PV Power values (*Fig 5*)

Data #2 (green):
- Polynomial has the highest R^2 and lowest value for all evaluation metrics except for MAE; Polynomial is the best-fit model in PV Power prediction (*Table 1*)
- All evaluation metric values have significant error but small variance, except for Decision Tree, but Lasso's and Ridge's values are very close to Linear
- Polynomial's regression line matches closely with the actual PV Power values (*Fig. 5*)

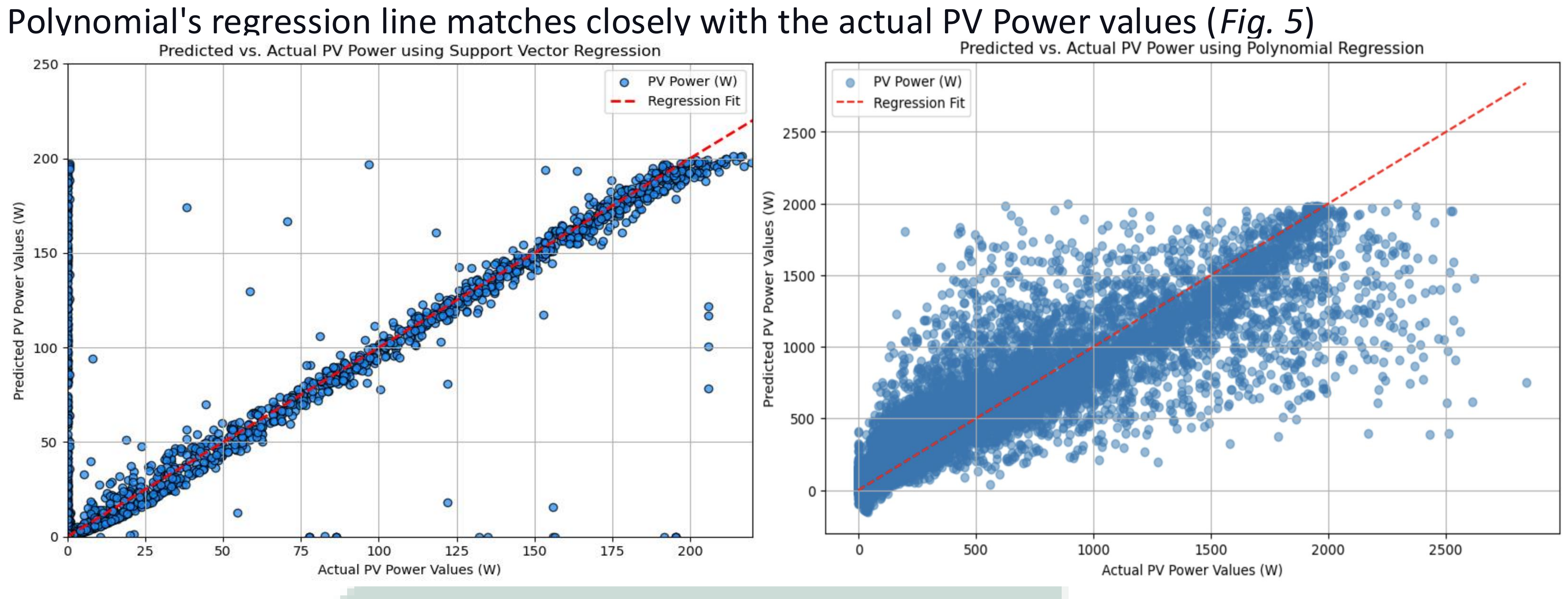

Figure 6: Predicted vs. Actual PV Power (Data #1 left using SVR and Data #2 right using Polynomial Regression

FURTHER DISCUSSION

- Investigate the effectiveness of time-series forecasting models (e.g. ARIMA and SARIMA) for improving PV power predictions over 1-5 year forecasting intervals
- Implement deep learning models (e.g., artificial neural networks) to enhance prediction accuracy; capture nonlinear relationships and temporal dynamic in PV power using other weather variables (humidity, temperature, pressure, etc.) [3]
- Apply probabilistic forecasting techniques (e.g. Bayesian Neural Networks) to provide a range of possible PV Power outputs and their respective probabilities, aid in risk management for more accurate, precise forecasts [4]

ACKNOWLEDGEMENTS

Thank you to Dr. Kerin Hilker-Balkissoon and Ms. Frances Troy's support and the funding for this study from the College of Science's RISE Research Scholars Program, COS 300, and the College of Engineering and Computing.

CONTRIBUTIONS

Data #1 (Milan, Italy) - Pragyan S. Shukla[1] [5]
Data #2 (Utrecht, Netherlands) – Rushil S. Johal[2] [6]
Conceptualization & Supervision – Sharmin Abdullah, Ph.D[3]

REFERENCES

[1] Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. *Sustainability, 15*(9), 7087. https://doi.org/10.3390/su15097087

[2] Du, W., Peng, S.-T., Wu, P.-S., & Tseng, M.-L. (2024). High-Accuracy Photovoltaic Power Prediction under Varying Meteorological Conditions: Enhanced and Improved Beluga Whale Optimization Extreme Learning Machine. *Energies, 17*(10), 2309. https://doi.org/10.3390/en17102309

[3] Leva, S., Nespoli, A., Pretto, S., Mussetta, M., & Ogliari, E. G. C. (2020). PV Plant Power Nowcasting: A Real Case Comparative Study With an Open Access Dataset. *IEEE Access, 8,* 194428–194440. https://doi.org/10.1109/access.2020.3031439

[4] Mahmud, K., Azam, S., Karim, A., Zobaed, S., Shanmugam, B., & Mathur, D. (2021). Machine Learning Based PV Power Generation Forecasting in Alice Springs. *IEEE Access, 9,* 46117–46128. https://doi.org/10.1109/access.2021.3066494

[5] Mussetta, M. (2020, September 23). Photovoltaic Power and Weather Parameters. *Ieee-Dataport.org.* https://ieee-dataport.org/open-access/photovoltaic-power-and-weather-parameters

[6] Visser, L., Elsinga, B., AlSkaif, T., & van Sark, W. (2022, August 1). Open-Source Quality Control Routine and Multi-Year Power Generation Data of 175 PV Systems. *Zenodo.* https://doi.org/10.5281/zenodo.10953360