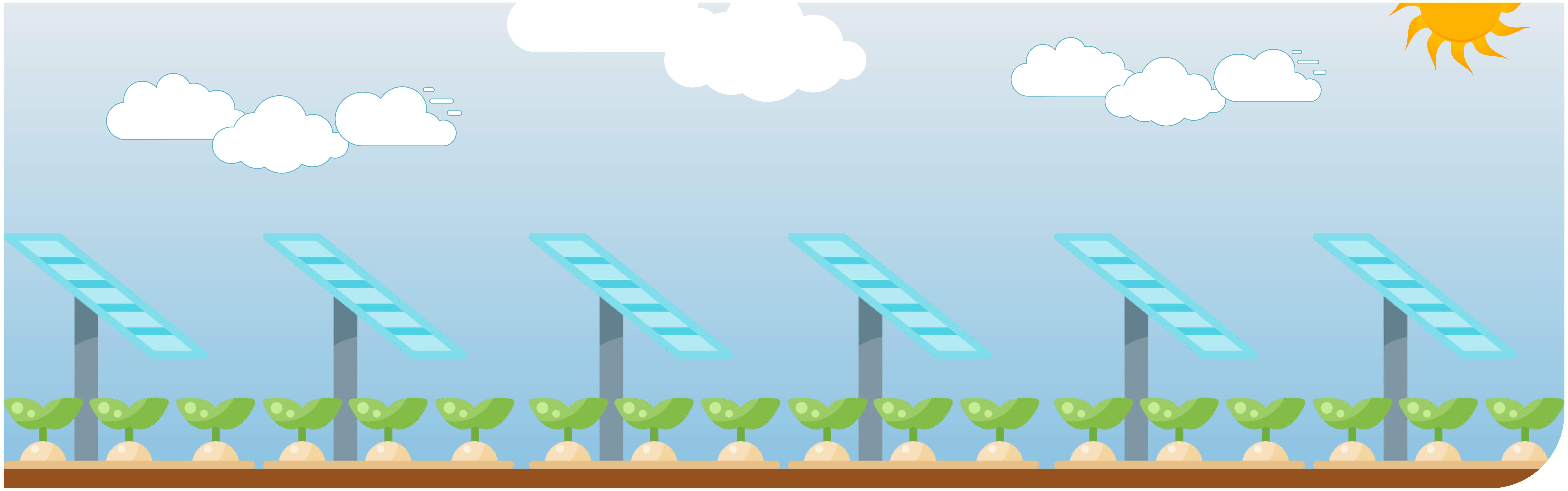
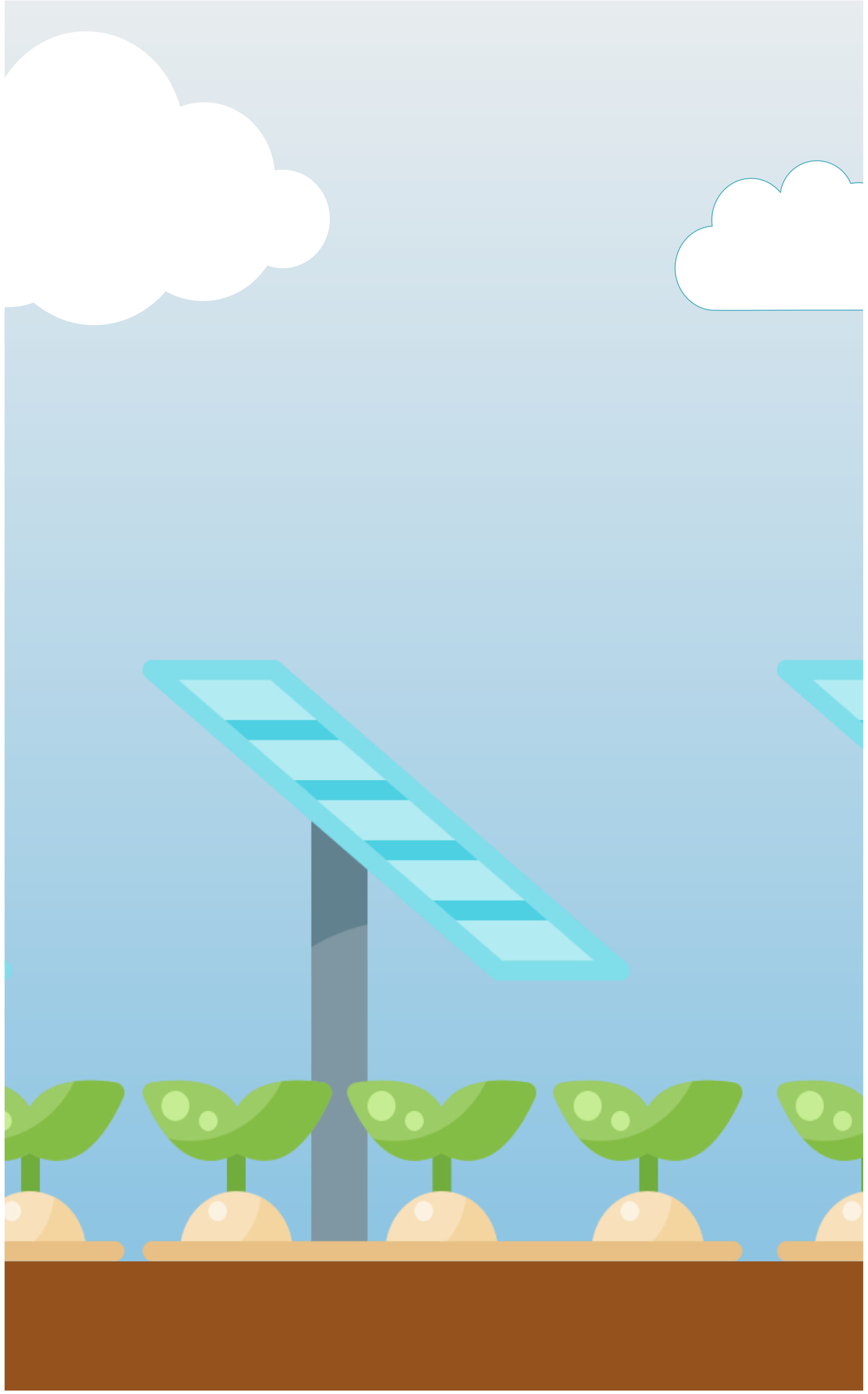


Predicting Photovoltaic Energy Production Performance

This project uses environmental and operational data to predict photovoltaic energy performance, helping optimize solar production, storage, and grid management through machine learning analysis.

Summary



Who are we?

This project was created by Marcus Serginio Inossaint and Alberto Sylvéus, two data scientists who decided to apply their expertise to the challenging field of grid energy. It was developed as part of their capstone project for the 6-month Data Science Bootcamp organized by Akademi Inc. in partnership with Flatiron School.

https://www.linkedin.com/in/marcus-inossaint/

https://www.linkedin.com/in/albertosylveus/

April 2025, DataScience bootcamp powered by : Akademi partnered with FLATIRON SCHOOL

Overview

What is this project for?

The goal of this Data Science project is to develop a **predictive model** capable of estimating PV energy production with high accuracy. By doing so, the organization will improve resource allocation, and power delivery reliability.

What can the models from this project predict?

The models are capable of predicting whether the power production for the next day will be : **Low**, **Medium**, or **High**, according to given input parameters.

What makes it important?

The project benefits renewable energy firms improving PV efficiency, government bodies (like EDH) shaping energy policy, local communities seeking stable clean power, and data science and academic teams handling data, model development, research, and system integration.

Application fields

Grid Management (EDH): Use predictions to optimize daily energy distribution between solar, diesel, and hydro sources.

Energy Storage Planning: Anticipate production peaks or drops to schedule battery charging and discharging cycles.

Maintenance Scheduling: Detect abnormal performance trends that may indicate panel degradation or dust accumulation.

Disaster Preparedness: Provide data-driven insights on energy resilience before storms or extended cloudy periods.

Data understanding

Where did the data come from?

The dataset, collected by Tiechui Yao et al. (2021) and hosted on the **SciDB** platform, contains over **270,000** records from **10 solar stations**.

It combines Numerical Weather Prediction and Local Meteorological Data, featuring 7 key variables: **global irradiance**, **diffuse irradiance**, **temperature**, **pressure**, **wind direction**, **wind speed**, and **photovoltaic output**.

How can we use it in local context?

Though not from Haiti, the dataset reflects similar tropical conditions.

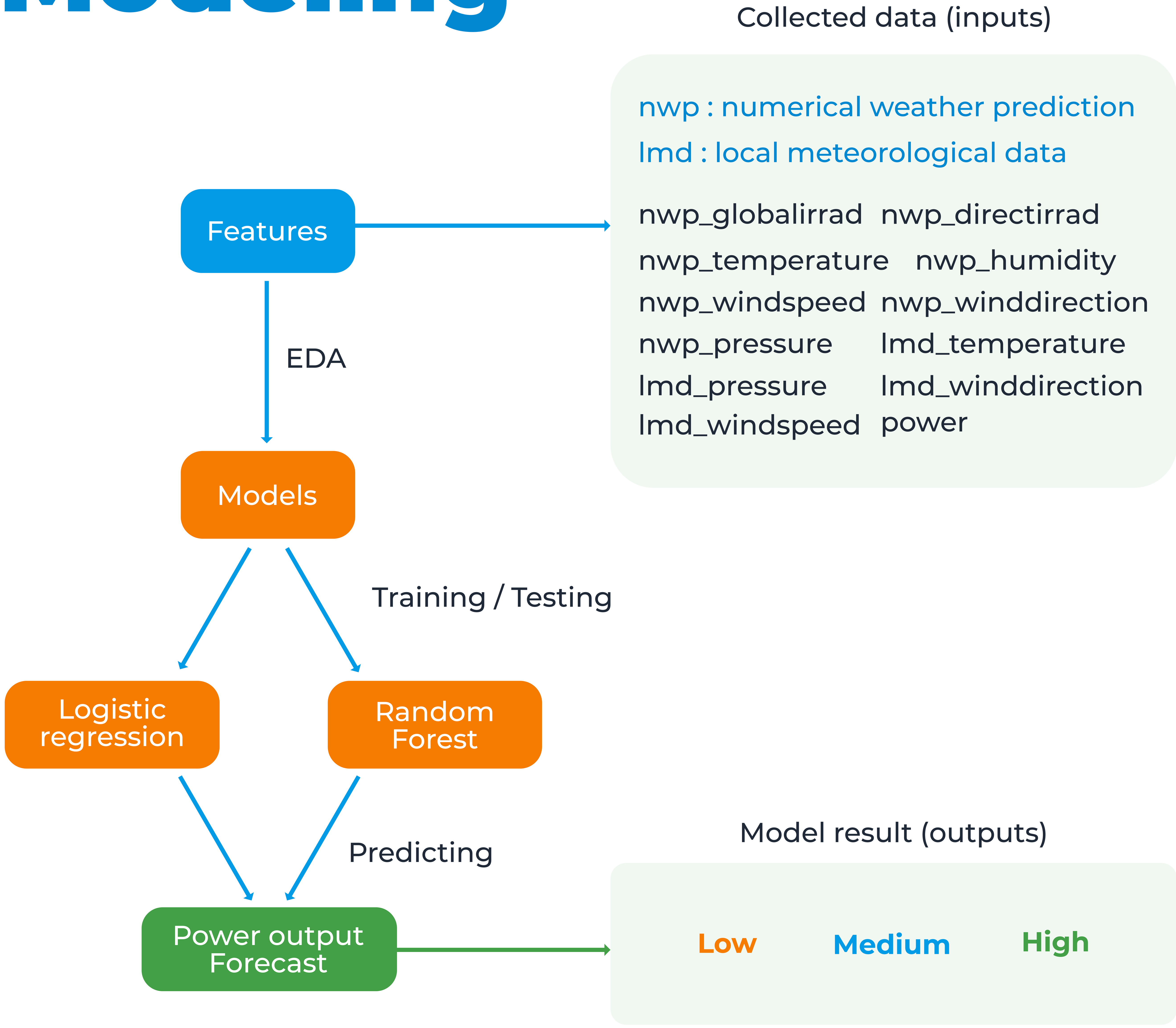
The model focuses on key transferable factors like: irradiance, humidity, wind—making it useful for proof-of-concept.

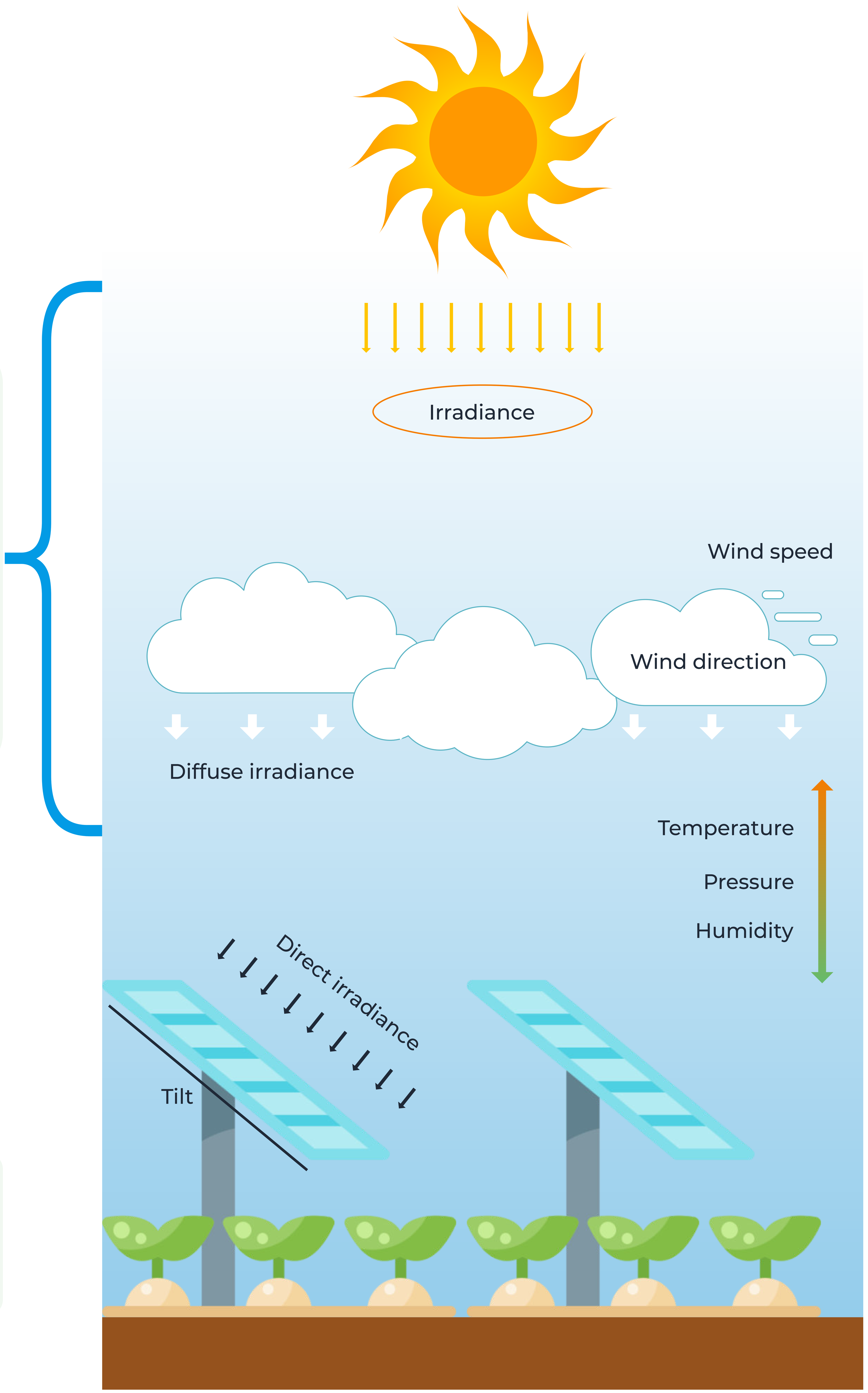
It forms a scalable base ready for adaptation with Haitian or Caribbean data.

Dataset Link: https://www.scidb.cn/en/detail?dataSetId=f8f3d7af144f441795c5781497e56b62

Method

Modeling





Method

This project tackles a **classification** task to predict **PV energy levels** (Low, Medium, High) from environmental data.

The target variable **power_class** was derived from the continious feature **power**.

Models tested: Logistic Regression and Random Forest.
Evaluation metrics: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

Model 1 observation

Logistic Regression Accuracy: 0.8990881347207412

	precision	recall	f1-score	support
High	0.87	0.91	0.89	13065
Low	0.95	0.97	0.96	35152
Medium	0.62	0.45	0.52	6177
accuracy			0.90	54394
macro avg	0.81	0.78	0.79	54394
weighted avg	0.89	0.90	0.89	54394

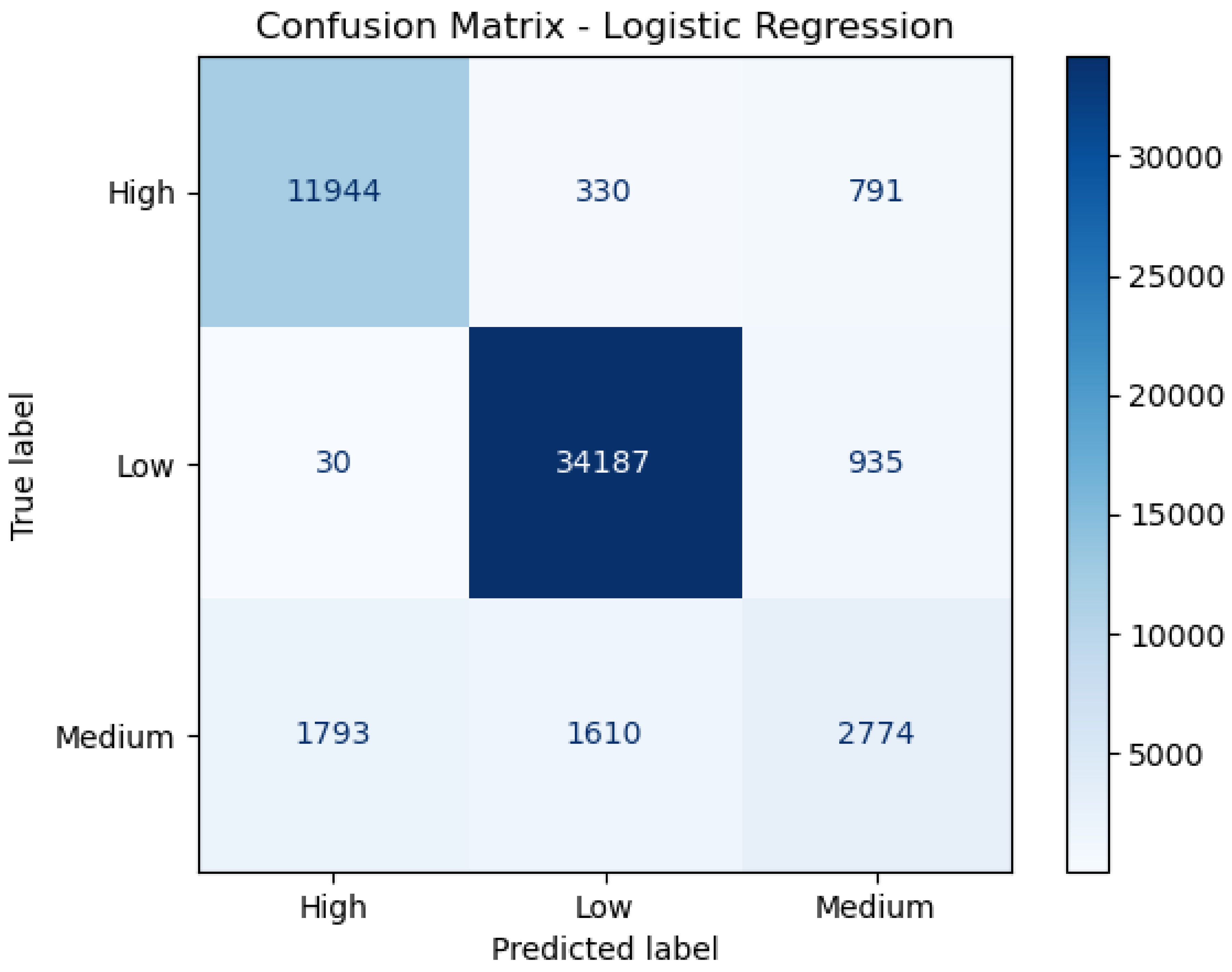
Model 2 observation

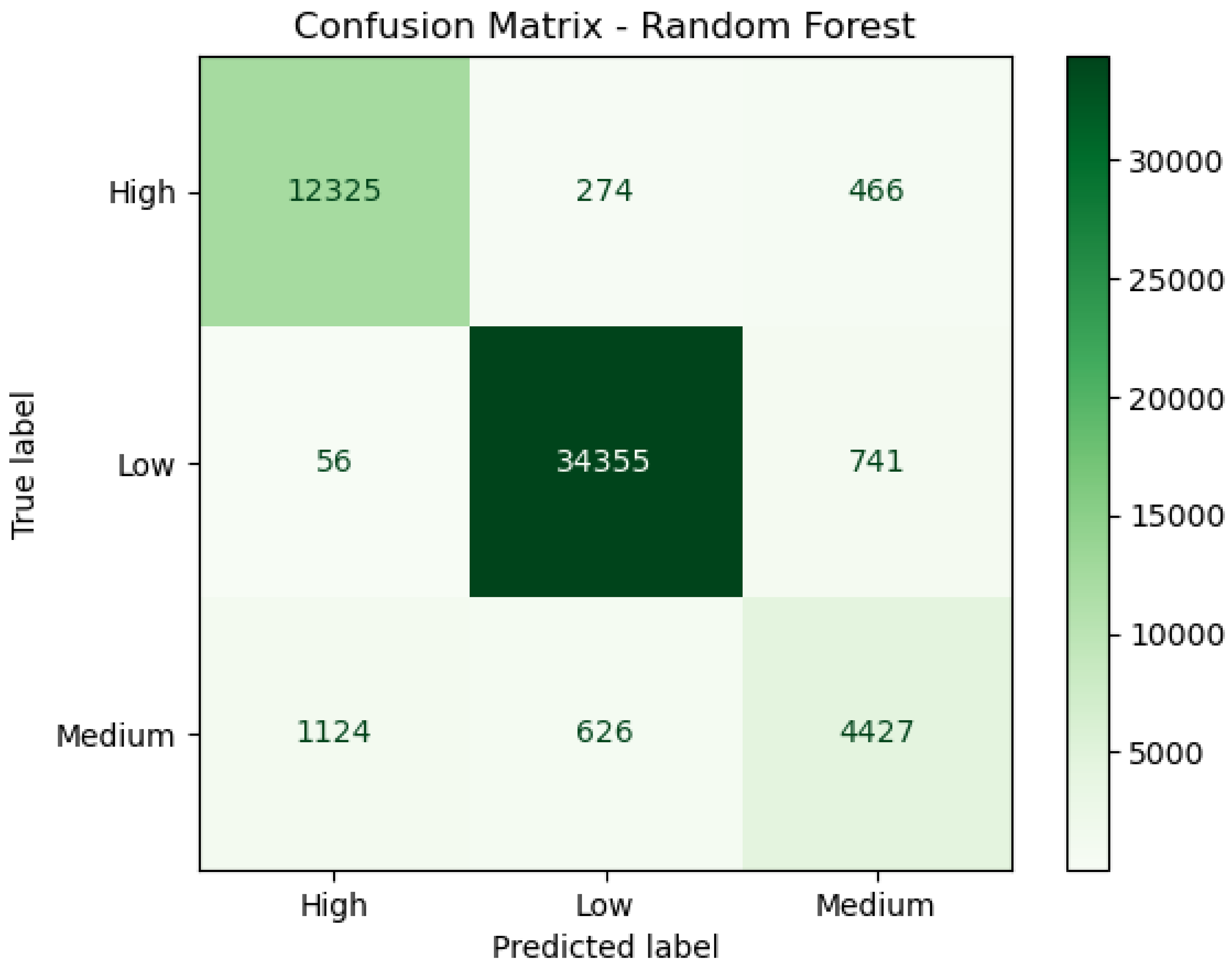
Random Forest Accuracy: 0.9395705408684781

	precision	recall	f1-score	support
High	0.91	0.94	0.93	13065
Low	0.97	0.98	0.98	35152
Medium	0.79	0.72	0.75	6177
accuracy			0.94	54394
macro avg	0.89	0.88	0.88	54394
weighted avg	0.94	0.94	0.94	54394

Observation

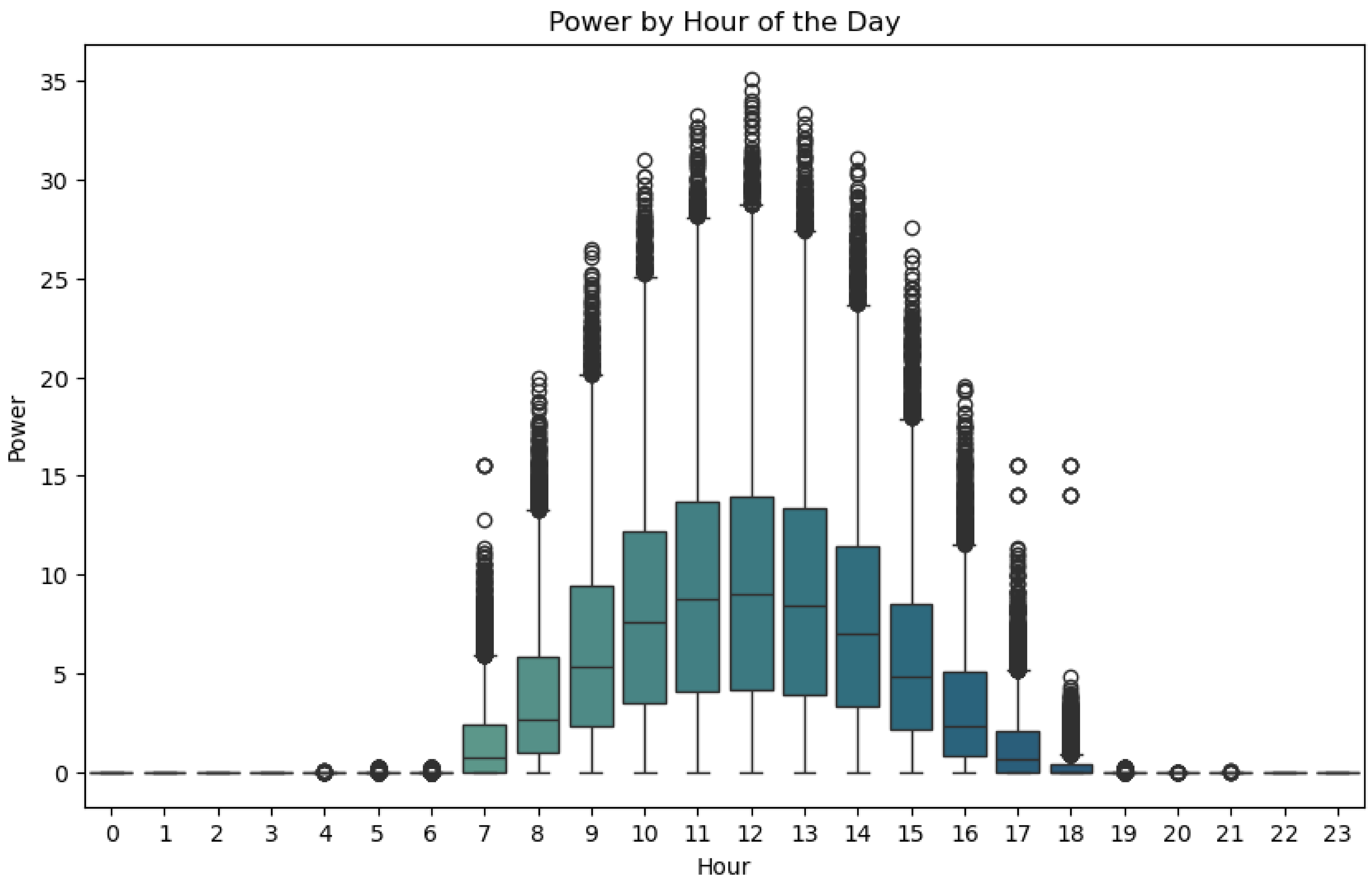
Both models shows a high score of accuracy after training and testing process which indicates great potentiel of predicting future solar energy production performance with precision.



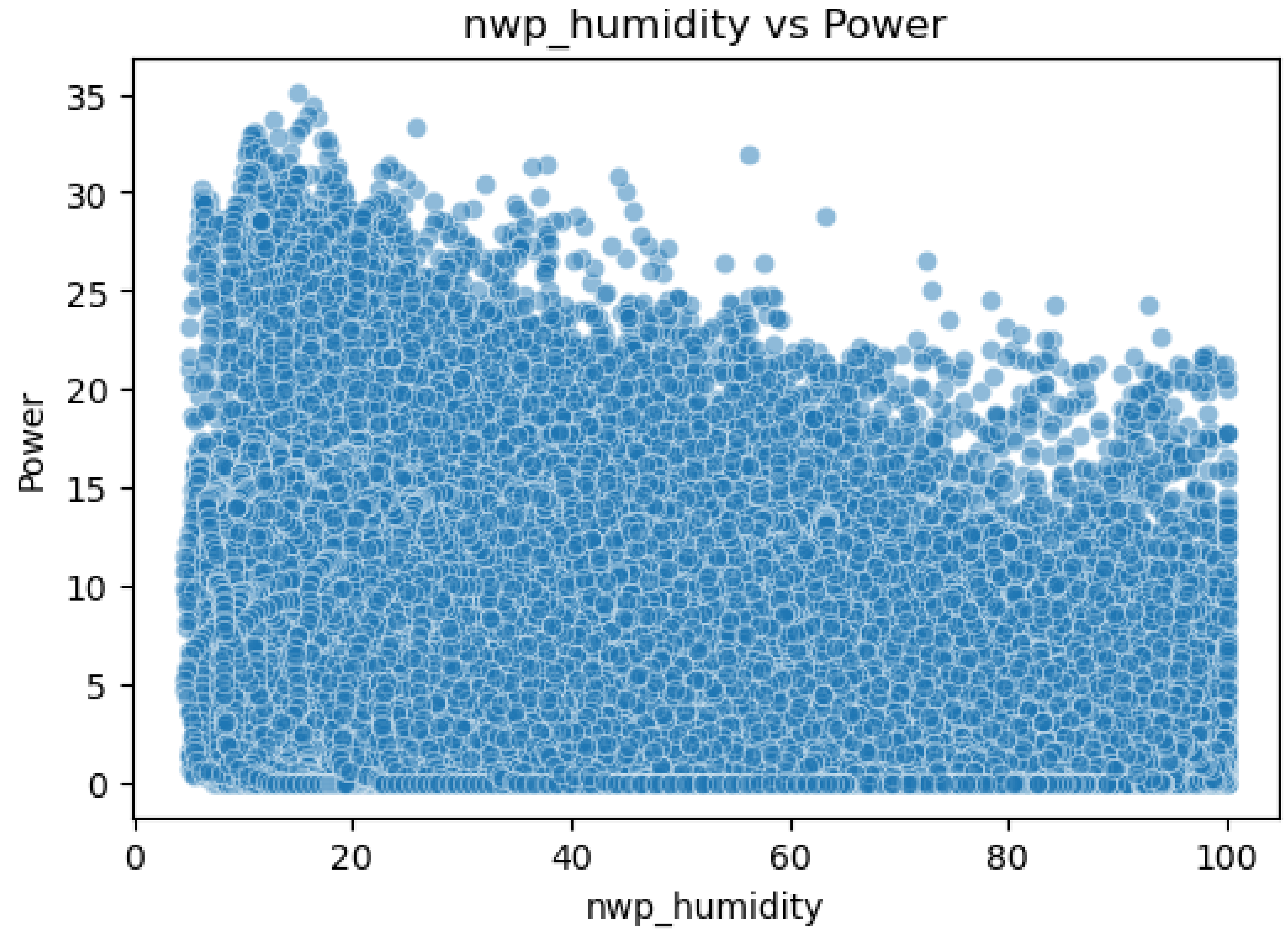


Observation

Here are a few observations we noted through our analysis starting from **trivial** to **subtil**.

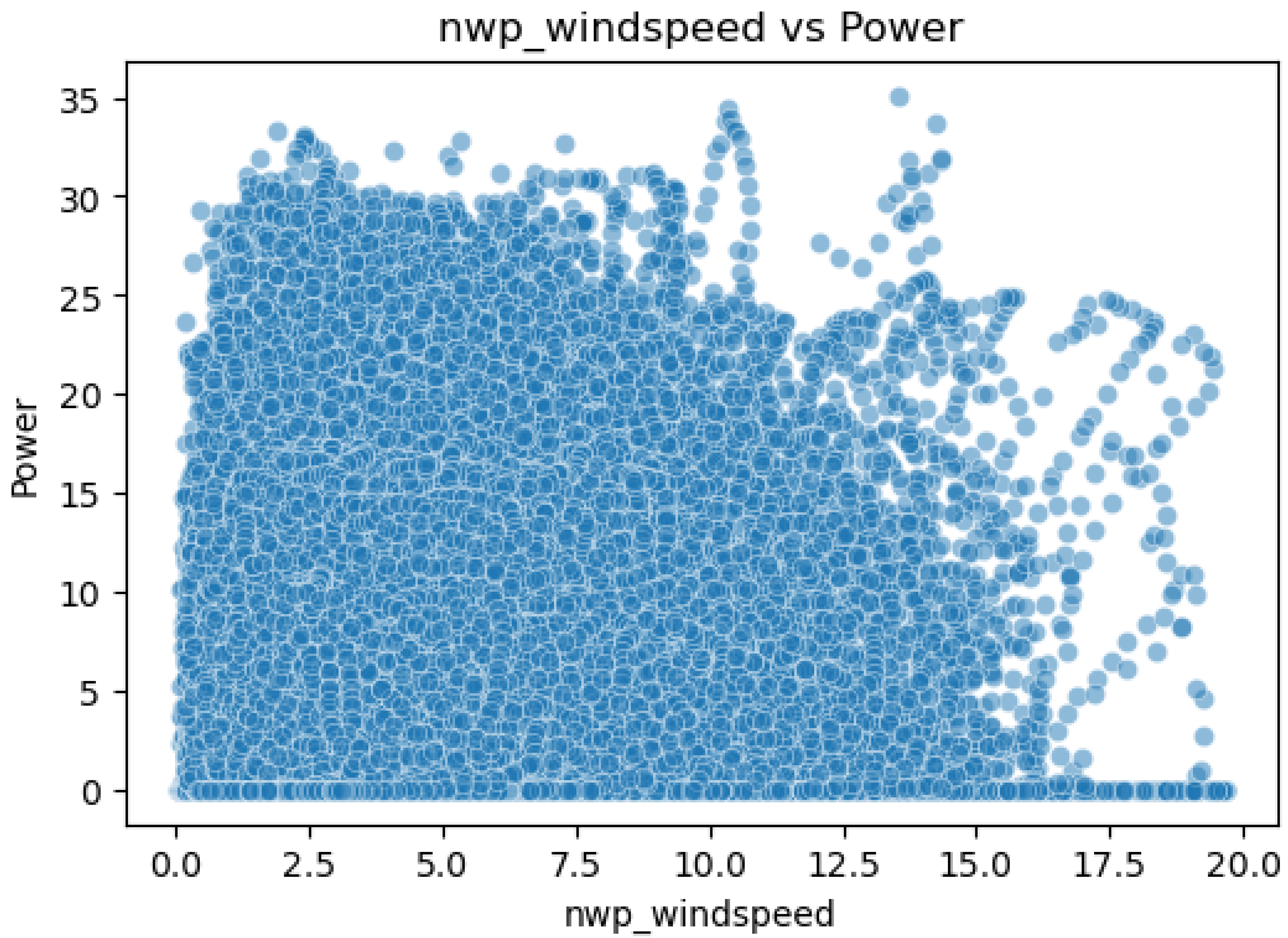


In normal condition, solar power peak often occurs between 11h - 13h.



High humidity in weather tend to negatively affect solar power production

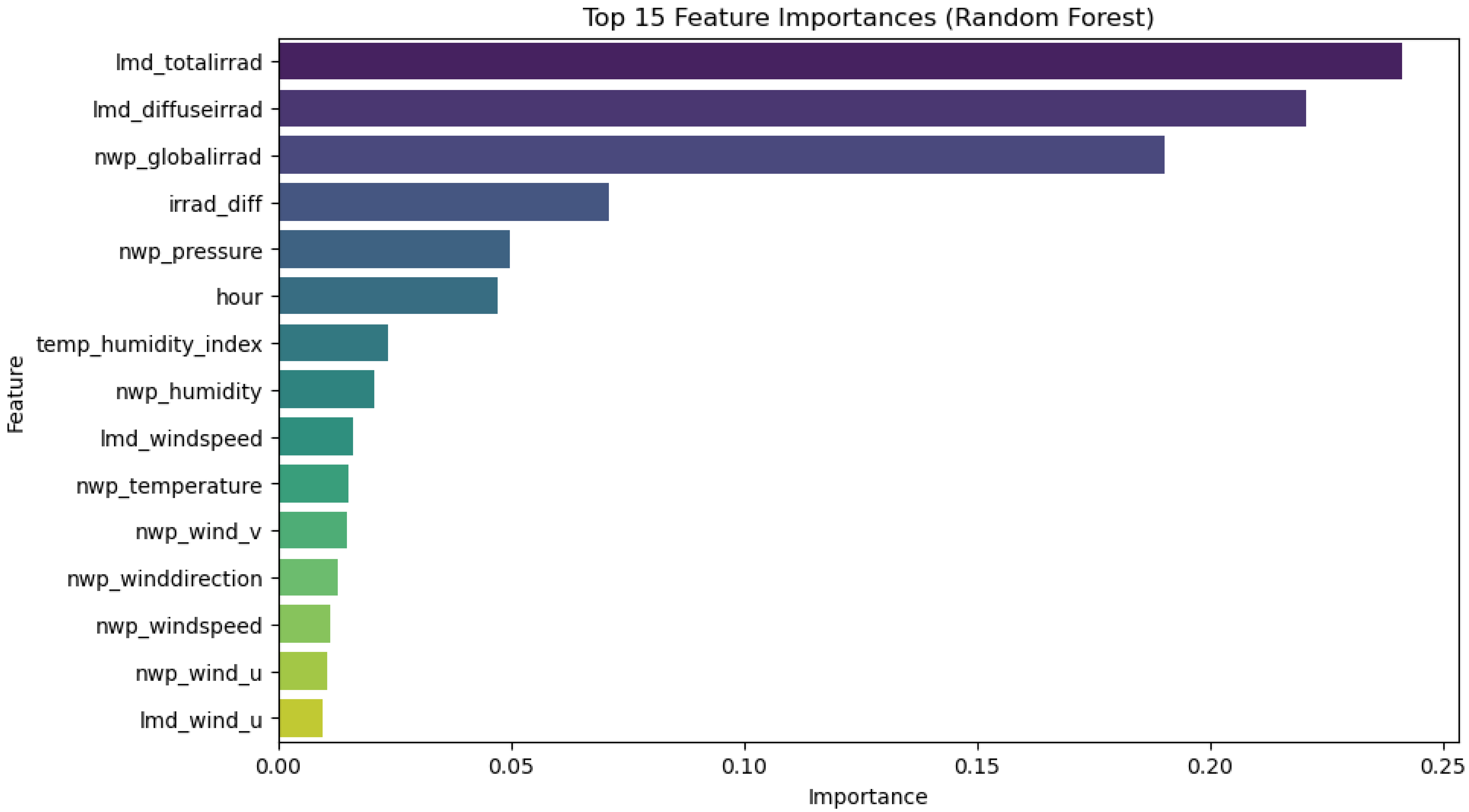
Observation



At a certain speed, wind may affect solar production since it can move cloud mass in the sky and which reduce the amount of direct irradiance that reach panels. It can also break solar plants or cables in case of a bad wheater climate (storm, hurricane...)

Observation

Here are the important features which impact solar energy production :



Conclusion

This work establishes the methodological foundation of data preprocessing, temporal modeling, and interpretability analysis while recognizing that local calibration will be an essential next step for real world deployment in Haiti or other comparable regions.

The developed model successfully provides accurate classifications with interpretable and physically meaningful drivers. It demonstrates that key environmental variables—particularly **irradiance**, **pressure** and **temperature** — are strong predictors of photovoltaic performance.

Thank you