The Future is bright with Solar Forecast

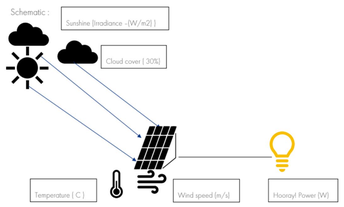
Team Members : Sephora Yameogo, Karin Farajnejadi, Licela Martinez, Du Nguyen and Berkan Ozturk.
Faculty advisor : Dr. Mikyoung Jun
Industry Mentors: Dr. Aniruddha Panda, and Dr. Pandu Devarakota

This project is supported by a grant from the National Science Foundation (NSF IIS-2123247).

Introduction

Clouds can rapidly envelop solar farms, leading to a substantial decline in power generation within minutes. To address this challenge, the team has been invited to creatively forecast the percentage of cloud coverage for a 2 hours intervals by leveraging existing weather data and sky camera information. The team is encouraged to devise their own methodologies for predicting cloud coverage in the sky. The accuracy of these predictions will serve as the basis for evaluating the submitted algorithm solutions.

Figure 1
Figure 1



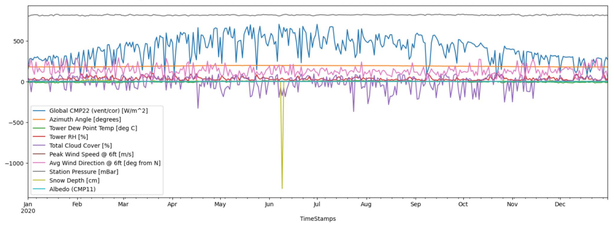


Figure 2: Predictors Vs. Time in months

Data Preprocessing

- **Data was aquired from Shell Hackathon data provided by Industry Mentors**
- **Quality control:**
  - Remove outliers by studying dataset distribution
  - Erroneous values such as negative irradiance
- **Feature Selection**
  - Using a heatmap, we keep the variable whose correlation fall in the Interval [-0.7 0.7]
- **Dataset Partitioning**
  - Sample observation every 10 mn (by dropping every 9 observations)
  - Keep continuous sequences of 4 hrs and more

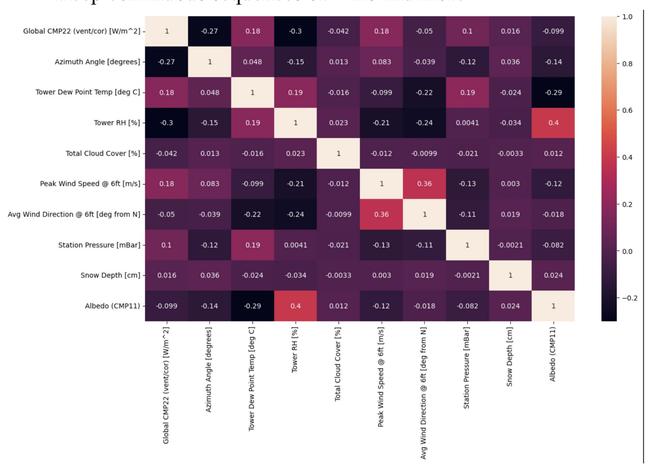


Figure 3: Heatmap

Methods

Training and Testing code written in Python
All models were tuned using RandomizedSearchCV
Metrics: R^2 and Graphical Comparison

Models

Model 1: Multiple Linear Regression :

Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable.

R^2
0.12

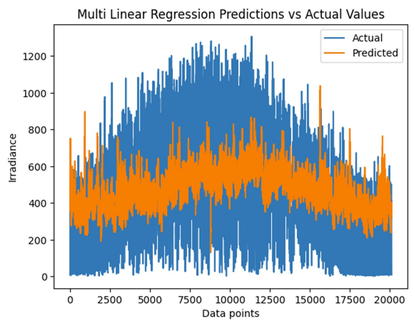


Figure 4

Model 2: Extreme Gradient Boosting

Extreme Gradient Boosting is an efficient open-source implementation of the stochastic gradient boosting ensemble algorithm.

R^2
0.89

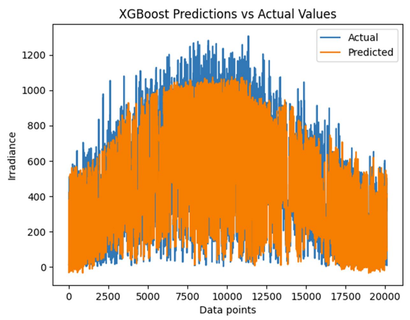


Figure 5

Model 3: Support Vector Regression

The SVR algorithm's goal is to minimize the error by identifying a function that puts more of the original points inside the tube while at the same time reducing the "slack."

R^2
0.54

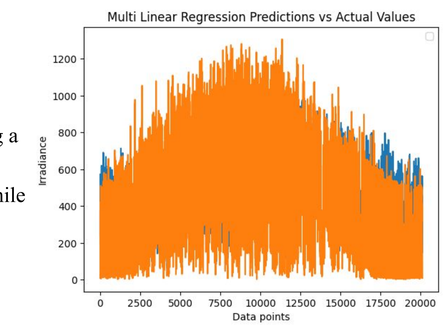


Figure 6

Testing/Results

Out of 300 sample days, test sets number 1, 4, and 45 were used from 7:30 am to 9:30 am. They respectively belong to the year 2018, 2019, 2021.

Test set 1	R^2
MLR	- 0.84094
XGB	0.78773
SVR	- 7.56706

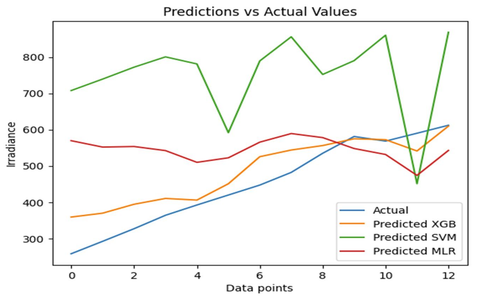


Figure 7

Test set 4	R^2
MLR	- 0.84094
XGB	- 0.09323
SVR	- 0.35767

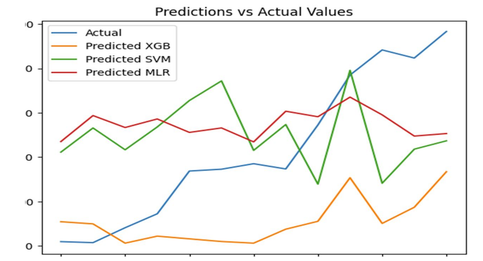


Figure 8

Test set 45	R^2
MLR	0.43149
XGB	0.61059
SVR	- 1.62115

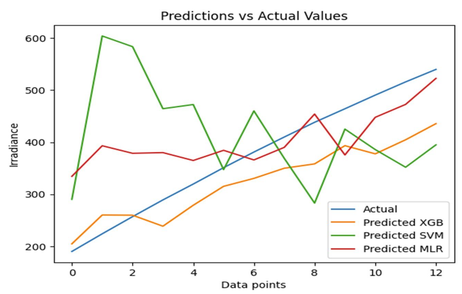


Figure 9

Results and Conclusion

Extreme Gradient Boost showed the best results for all three test sets, particularly with the test set 1, with an R-squared of 0.78773, showing a promising fit with like datasets. It is notable that the testing sets are not at the same time/season as the training. Hence, the model may not be generalized to yearly data, showing it overfitted to the original dataset. In future works, it is reasonable to explore the usage of neural networks, particularly RNN for a time series and the usage of training data over an entire year encompassing yearly changes. In all, solar forecasting has shown promising use for the energy industry and better evaluate the usage of solar energy for client use.