Predicting Solar Power Generation with Python

S. Valenzuela, NMSU Carlsbad; N. Mora, Southwestern Indian Polytechnic Institute; Mentor Dr. Huiping Cao, NMSU

Introduction

Renewable Energy is energy from naturally replenishing sources. Solar photovoltaic (PV) devices change sunlight into energy. Although inexhaustible, the amount of solar energy that is available per unit of time is limited. PV panels can produce between 250 and 400 Watts of power. Actual output depends on factors like shading, orientation, and sun hours.

Clear skies at NMSU parking lot solar array. Photo credit: NMSU Facilities

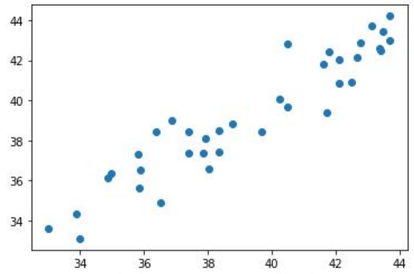
Solar Energy Potential in New Mexico is 500-629 watt hours/sq ft/day. The use of PV in a Smart Grid requires control of the production and distribution of electricity. Power generation and weather data history can be used to predict future power generation. This allows for efficient use of the Smart Grid.

Procedure

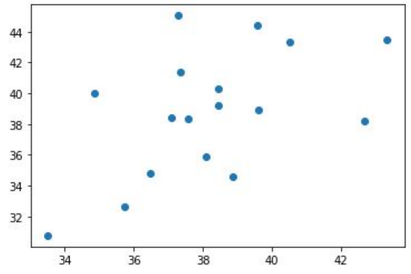
The team was provided two data sets: Solar Power Generation and Weather. All data was collected from January 1st, 2020 to May 31st, 2020.

- Pre-process raw data to have only the needed data points
- Train portion of processed data using Scikit-Learn's linear regression estimator, random forest regressor, and fit method
- Use predict method on trained model
- Test performance of model using Mean Absolute Error and Root Mean Squared Error

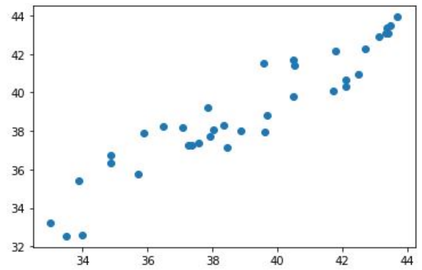
How much energy should we expect?



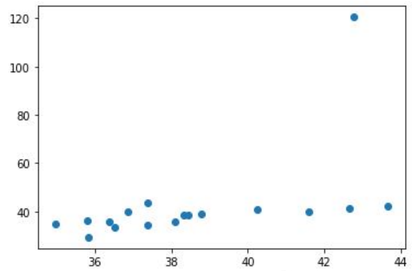
Training set A prediction.
MAE = 0.858197 RMSE = 1.182916



Testing set A prediction.
MAE = 2.856372 RMSE = 3.465006



Training set B prediction.
MAE = 0.856046 RMSE = 1.193732



Training set B prediction.
MAE = 6.399787 RMSE = 19.057185

Results

Machine learning results are only as good as the data. We spent a lot of time pre-processing the data to ensure accurate results. We encountered many problems along the way, specifically with our code. These problems allowed us to become better programmers and researchers. Dr. Cao supplied us with a clean data set and we were able to run the linear regression estimator.

References

- T. Ahmad and H. Chen, "Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment" in Energy, Volume 160, 1 October 2018, Pages 1008-1020
- U.S. Energy Information Administration - EIA - Independent Statistics and Analysis. (2019, December 4). Retrieved July 23, 2020, from https://www.eia.gov/energyexplained
- Dr. Huiping Cao, NMSU, Lecture 17

Conclusion

We use python to do a series of machine learning algorithms to understand the correlation between weather attributions that were observed in Las Cruces New Mexico and the energy data that was collected at NMSU. In doing so a linear regression algorithm was used to make predictions for Mean Absolute Error (MAE) and coefficient of determination(RMSE). Having the team run two different programs for the prediction they concluded that Linear regression is not the best algorithm to use for the targeted range of MSE and RMSE which was 0-1. Set A was the closed in obtaining this goal.

Future Plans

- Exploring more time series data on weather attributes and solar energy generation
- Learning more about machine learning and how historical data can be used to predict how much power is being generated from the New Mexico SMART Grid Center.

Acknowledgements

We would like to thank Dr. Huiping Cao for helping us along this journey and ESPCoR for giving students the opportunity to work with the New Mexico SMART Grid Center. This work is funded by the National Science Foundation (NSF) award #OIA-1757207. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation