CS 519 Project Report

Randall Woodall
New Mexico State University
hossrw@nmsu.edu

Emrah Sariboz
New Mexico State University
emrah@nmsu.edu

Aya Elsayed
New Mexico State University
aynasser@nmsu.edu

ACM Reference Format:
Randall Woodall, Emrah Sariboz, and Aya Elsayed. 2020. CS 519 Project Report. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 PROBLEM STATEMENT

Prediction of power production for solar panels from weather information.

1.1 Motivation

Since starting up generators is costly, having an idea of how much power a utility would need to produce can help the utility to reduce costs of generation. There exists an abundance of data in relating to solar production, and an even greater abundance relating to weather and forecasting. Combining weather history, weather forecasting, and information about solar production could lead to a schedule of solar power production that would allow, in the long run, utilities to create more optimal generator schedules and power rates. This type of scheduling would help the consumer to save money by allowing the producer to avoid startup costs and idling costs. Because of this, new and more accurate methods will be valuable to all parties involved in electrical power.

1.2 Direct Problem Definition

Given five second solar production data in the form of power, and also given historical weather data including wind, temperature, humidity and pressure, can we make semi-accurate predictions about the solar data at the time of the weather data? To restate this problem in fewer words, can you use current state of the weather to predict current output of a solar panel? If this is the case, then a day-ahead power forecast should be directly correlated to a day-ahead weather forecast. The goal of this project is to show that, given current data, we can predict what a solar panel produces, and by logical extension, if we can predict what it will produce tomorrow.

1.3 Problem Limitations

We acknowledge that we have a hard limit on the accuracy of any day-ahead power predictions, based in the uncertainty of day-ahead weather predictions. Going back to statistics, this limit on accuracy and, therefore, the introduced error, goes back to the statement, "Garbage in, garbage out". Due to this, we will eliminate the space for introduced error by restraining our semester project to include only the historical data, where we have known values and actual measurements. If we can assume a normal distribution around the predicted weather values wherein the actual weather values will fall, then we can incorporate this into our prediction for power produced to give a normal distribution showing what power will probably be.

2 ALGORITHMS TESTED

2.1 Linear Regression

For linear regression analysis, we first used a correlation matrix to find those features that have a high correlation with the power attribute. In other words, we first conducted a feature section. Upon this process, we have left with two features. By applying linear regression, we were able to get an $r^2 score of 0.796$.

2.2 Polynomial Regression

To use polynomial regression, we first applied PCA to reduce down to 3-dimensions. This is to limit the size of the resulting polynomial coefficient size. Using this technique, we were able to get an $r^2 score of 0.819$.

2.3 Random Forest Regression

2.4 Algorithm selection

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn