Module 2 Capstone Proposal

Marshall McQuillen

4/11/2018

The Question

How accurately can Direct Normal Irradiance (DNI) be predicted at a granular level (30 minute intervals)?

My thought process is if this can be predicted with a certain level of accuracy, solar power plants can plan scheduled maintenance on the "best" day, that being the day with the lowest DNI, at the "best" time.

The Data

NREL has an API where I can obtain DNI measurements in half hour intervals for a given latitude and longitude, for years dating back to ~2010. In addition, they have data on CSP and PV power plant locations in the United States, as well as other locations around the world (see MVP++).

Most of the latitude and longitude values are populated, however there are a few missing values. Some of this information could be scraped from an accompanying webpage to where the data set is located, however if that doesn't work I plan on using Google's Geolocation API to supplement the existing data.

A preview of the data...

Year	Month	Day	Hour	Temperature	DNI	DHI	GHI
2010	1	1	8	-1.008124	295	25	47
2010	1	1	8	0.154718	505	51	135
2010	1	1	9	1.317529	655	65	228
2010	1	1	9	2.891779	748	74	318
2010	1	1	10	4.466028	809	82	400

MVP

For all solar power plants in the USA...

Use weather data (temp, cloud cover, humidity, etc,), time (30 min. intervals) and location (latitude, longitude) data to predict DNI.

- Compare all complex models (I want to train random forests, boosting and a regression Neural Network) against multiple linear regression

MVP+

Forecast DNI X days out (with confidence intervals) for all locations in US

MVP++

- Do the above for solar power plants outside of the United States.