James Griffin

Stat 536

Solar Panel Report

**Introduction**

Solar Panel advertisements and door to door salesman have been popping up all over the country. Understanding the output of solar panels is vital to knowing if the panels are themselves a good investment and a variable green power source for the future.

The data comes from a home in Utah County, tracking the dates from January 1st, 2017, until December 31st, 2019. For this analysis we will use the daily observations of this time series data set to understand the output in kilowatt hours from the solar panels.

Chart, line chart

Description automatically generated

Above we show the output of the solar panels across time. We can see that through these three years we have a very seasonal trend, with lower outputs in the winter or in the summer. We first examined two different seasonality lengths: 365 days and 91 days. We wanted to examine the overall trend was different for a season of a year vs the four seasons. Having divided the season into four seasons.

Here we see the trend with a seasonality defined as 91 days

Chart

Description automatically generated

Here we see the trend with a seasonality of 365 days

Chart

Description automatically generated with low confidence

In the seasonality of 91 days, we see a definite sin wave with decreasing amplitude. When we define the seasonality as a year, we see a linear decreasing function. In both cases we see that the solar panels are in fact losing their output power. We notice that the winters observed, have rather similar levels at the troughs of the sign wave, however their peaks are steadily decreasing, as the yearly trend shows.

**Methodology**

Time series data presents a few challenges, primarily that the observations are correlated, while we do have three years of data, we don’t have unique information from those days. Because the previous days heavily influence the next. To illustrate this point we created a plot with a 30 day rolling average and a 95% confidence interval.Chart

Description automatically generated

This highlights the above point, the lack of independence of our observations means we cannot use traditional regression techniques. We also cannot do perform the traditional train test split, as our data is very clearly ordered. Taking a random sample of the data would reorganize and the data would lose its sequential nature.

**The Model and Assumptions**

Having established the decreasing output of the solar panels, and the time series nature of the data, we first approached this problem from the Autoregressive Integrated Moving Average api where we can define a range for the seasonality, trend and noise, and perform a grid search to find the optimal parameters by finding the lowest AIC value. This model was effective but struggled when we had seasons of longer than 4 weeks. So, we then took the examined the Prophet model designed by Facebook in 2017.

As Facebook deals with lots of time series data, this open-source API for both python and R is an additive regression model where non-linear trends fit daily, weekly, monthly and yearly trends, and even accounts for American holidays by using a piecewise linear or logistic growth curve trend. This robust estimator automatically detects differences in trends by selecting the change points from the data. The seasonal components are found by using Fourier[[1]](#footnote-1) transforms and series.

This is clearly a more robust estimator than what my limited grid search would have produced, and we felt even better about using this, because the algorithm uses whichever is more optimal, a linear or logistic growth for the seasonality, and looking at a histogram of the distribution of kilowatt hours is essentially normal[[2]](#footnote-2).

The Prophet uses a decomposable time series model with three main components, trend, seasonality and holidays. They are combined into the following equation:

Where we define: peicewisse linear or logistic growth curve for modeling non periodic changes in the data. period changes, so in our case, seasonally with 91 days or yearly with 365 days, : effects of holidays, and error term which accounts for any usual changes not already explained by the other three variables.

Using time as a regressor, Prophet takes a very unique approach, attempting to fit linear and non-linear functions of time as components, it is framing the forecasting problem as curve fitting exercise rather than looking explicitly at the time base dependence of each observation within a time series. This is how the model approximates

The seasonality effects are approximated by the following equation:

Where P is the period, again, either 91 or 365 days, and are the parameters needed to be estimated for a given N to model seasonality.

Holidays are a list, provided by the user, which mark specific dates as holidays and are dummy variables.

**Results**

Here is a chart of our monthly predicted average, using a seasonality of 91 days, with a 95% confidence interval, and projecting 2020, the data that we are interested in forecasting.

**Chart, scatter chart

Description automatically generated**

Our root mean squared error from the training data we had was 4.72. We notice that most of the deviations are in the summer and spring, and more consistent in the winter, which is a slight comfort because those months have comparatively higher outputs. We also notice the forecasts for 2020 go along with the decreasing in amplitude sin wave we discussed earlier. We were interested in seeing how long it would take for the solar panels to lose half of their power generating capability. We decided to do this as yearly averages. For the first year of the data, 2017, the panels had a yearly average of 33.27 kwh, for the projected 2020 forecasts, we project a mean of 27.01 kwh, and it would take another five years until 2025 for the output to be 16.03 kwh, which is less than half of their original output.

Overall, we have examined the Facebook Prophet api’s ability to examine time series data for solar panels on a house in the Utah County area. This area has very distinct seasons throughout the year, and we see that in terms of the amount of power generated by the solar panels. We believe that this was the best season length to use, rather than that of a year, so that we could capture that understand that part of the data rather than the linear yearly trends. We also believe that it will take until 2025, eight years after instillation for the panels to lose half of their power generating capabilities in terms of a yearly average.

**Appendix**

Chart, histogram

Description automatically generated

1. For intuition on Fourier transforms and series, please look up Fourier Transform 3Blue 1Brown on YouTube. [↑](#footnote-ref-1)
2. See Appendix [↑](#footnote-ref-2)