**Algorithms used**

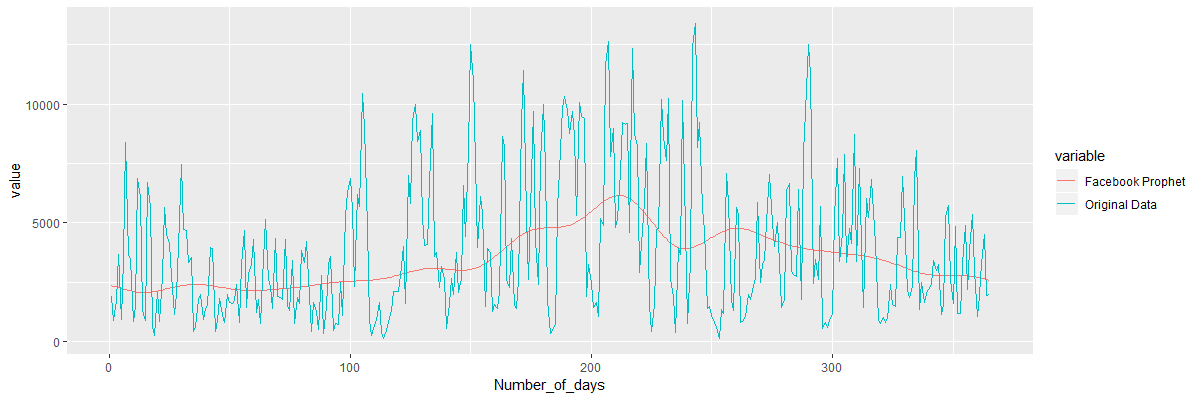
The fact that time series has only a single feature given against a time line makes the time series analysis different from normal regression. For regression the number of values to be predicted never becomes a problem. In regression algorithms the focus is to learn the variations of target feature with respect to the available training features rather than the result of last point. This makes regression algorithms better for shorter predictions. We have used three types of algorithms and techniques. Standard time series forecasting algorithms ARIMA, Holt winter’s and SSA. Regression algorithms GLM net (Lasso and Ridge regression), Random Forest, XGboost and SVM. Neural networks RNN (LSTMs). After applying all these algorithms we have taken the ensamble (weighted mean) of all the trained models. The weights has been calculated using given equation

. ( 4 )

Here w is the weight calculated for ensamble, e is error obtained for that algorithm individually and i denotes the number of algorithm. The denominator contains the sum of all the individual errors. The formula gives more weight to algorithms which have less RMSE and less weight to those which have high RMSE. We have used R for implementing the machine learning algorithms except for RNN which we have implemented using keras in python.

**Preprocessing**

Before training the model the data is pre processed. The data is checked for missing values. Then the data set is checked for outliers. We have imputed the least 10 percent and the highest 10 percent values with mean. The reason for mean imputation is to avoid the effect of too high values and too low values in the given series. Such values might mislead the learning algorithms. Such values might be a result of measurement errors. The data set is visualized. That is, the energy values in MW are plotted against time. The energy production shows annual seasonality for the given data set. The seasonality can be exploited. In R all algorithms ARIMA, Holt winter’s and SSA accept input as time series. Thus the data is mapped with respective time stamps and passed to the algorithms. For regression, features are required to be generated. We have engineered features for regressing the time series. To account for the variations in the time series we have generated the difference between the values of the time series over a fixed time span. For instance in one of the feature we have stored the difference of the current value and the value 7 steps before in another feature we have a difference of 30 steps and so on. We have taken a few such windows to include the effect of variations. Then we have used a number of commonly used regressors to get a time series forecast with engineered features. The time series problem has been converted to a regression problem.

Forecast analysis

Results for ARIMA model is reasonably accurate considering the scale which varies from 10.29 MW to 13411.29 MW of output energy and the duration of prediction which is 365 days. Standard predictions are made for 24 to 72 hours. Moreover they are calculated using multiple features such as wind speed in the region, density of air and various other physical factors such as height of wind mill, location of wind farm, weather changes and many more.

­­

