Electricity Data for COVID-19 Analysis

Akash Singh
singas@student.unimelb.edu.au

Haoyu Man
haoyum@student.unimelb.edu.au

Introduction

This report analyses the impact of COVID-19 pandemic on the electricity consumption pattern of the state of California, U.S.A. The COVID-19 pandemic has spread rapidly across California. The first confirmed case of COVID-19 in California was found on January 26,2020. And the state of emergency was declared on March 4,2020. During this pandemic period, the usual way businesses and households used to run has seen a dramatic change in electricity consumption pattern. And this report tries to uncover that shift using Machine Learning algorithms.

Dataset

We queried the hourly electricity demand data of California from U.S. Energy Information Administration[1], and hourly temperature data of California from the weather station at Los Angeles International Airport[2]. The combined dataset contains 5 features: hour, day of week, day of month, month, temperature. In the below table the demand column represents the electricity demand in MWh.
Training dataset: 70% data randomly chosen from 2019.
Development dataset: The other 30% data from 2019.
Testing dataset: data from January to September, 2020.

	hour	dayofweek	dayofmonth	month	temperature	demand
30721	1	1	1	1	52.0	26876.0
30722	2	1	1	1	49.0	25989.0
30723	3	1	1	1	43.0	25295.0
30724	4	1	1	1	46.0	24884.0
30725	5	1	1	1	47.0	24922.0

Figure 1: Dataset

Model

We employed Gradient Boosting regression model to predict electricity demand.

Gradient Boosting regression

Gradient Boosting produces a prediction model in the form of an ensemble of weak prediction models (decision trees). Gradient Boosting build trees one at a time, where each new tree helps to correct errors made by previously trained tree. With each tree added, the model becomes even more expressive.

Parameter tuning

In our model we have tuned these three parameters: $n\_estimators$, max_depth and $learning\_rate$. Using a high learning rate results in over-fitting of the model to the training dataset, a $learning\_rate$ = 0.1 is optimal for our data, $n\_estimators$ represents the number of trees used in the model. Usually the higher the number of trees the better the model learns the data. However, adding a lot of trees can slow down the training process considerably. We did a parameter search and found $n\_estimators$ = 500 to be the optimal value for our dataset. max_depth indicates how deep the built tree can be. The deeper the tree, the more splits it has and the more it can capture the information about the data. We fit multiple models with varying depths, and max_depth = 4 gives the highest prediction accuracy.

Analysis

The figure 2 below is representing the predictions for demand on the development dataset. From the figure it can be seen that the predictions are highly accurate with an accuracy score of 0.96 or 96%. This shows that our developed model is able to predict the demand with high accuracy.

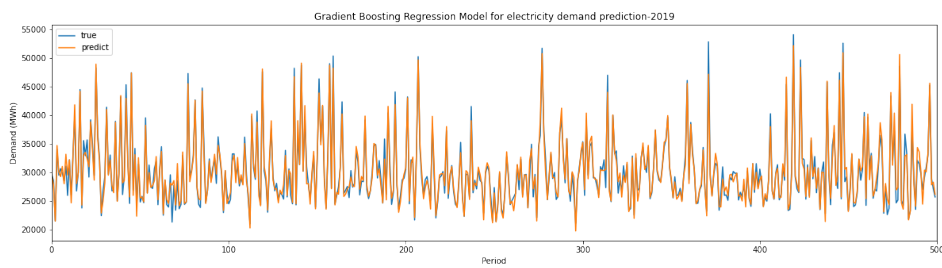

Figure 2: Prediction on development set-2019

The figure 3 below represents the predictions for demand on the test dataset. The accuracy achieved for test dataset is around 0.79 or 79%. From the figure it can be seen that the pattern of trend predicted by the model is almost the same but there is a shift in electricity demand in 2020. Which needs to be analyzed further with other graphs.

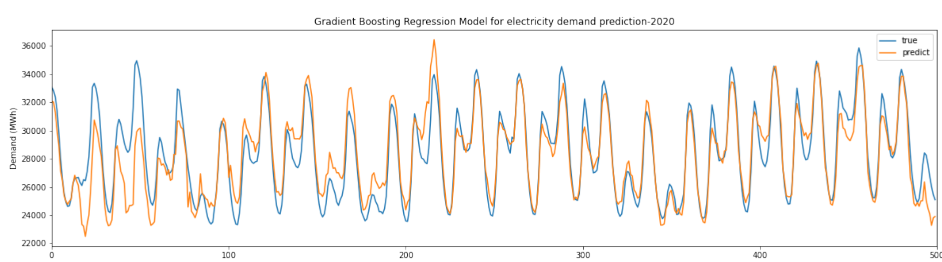

Figure 3: Prediction on test set-2020

The figure 4 below represents the comparison for month of March hourly electricity demand between the true values of 2019, 2020 and predicted value of 2020. From the figure it is quite evident that our model is making very good prediction for demand because the trend pattern for true value of 2019 electricity value matches with out prediction values for 2020. But the true value for electricity demand is increased by a 9.7% from the predicted value for electricity demand for 2020. This big shift in demand is due to increase in activity of the people and the businesses in preparation for the pandemic breakout.

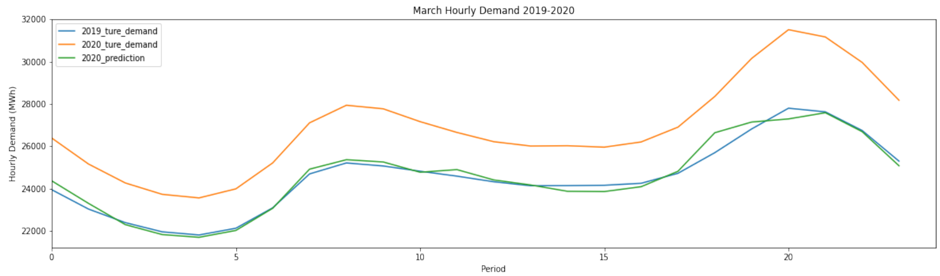

Figure 4: March hourly demand 2019-2020

The figure 5 below represents the comparison for month of April hourly electricity demand between the true values of 2019, 2020 and predicted values of 2020. From the figure it is quite evident that our model is making very good prediction for demand because the trend pattern for true value of 2019 electricity value matches very closely with out prediction values for 2020. But the true value for electricity demand in 2020 is reduced by a 5.1% from the predicted value for 2020 and this shift in demand is due the COVID-19 state of emergency applied in state of California. As people start working more from houses so the offices and business hubs went empty which leads to overall reduced demand.

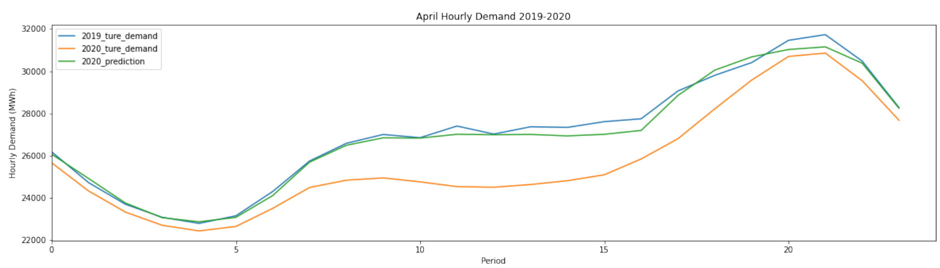

Figure 5: April hourly demand 2019-2020

In figure-6 also this shift in demand can be seen, where month wise % change in electricity demand for 2019 & 2020 is represented. For the month of March and April the shift in demand is very high and it is due to the COVID-19 state of emergency declared in the state. These results shows that our model is capable of making correct predictions. And with the help of our model we can detect these pandemic breakouts.

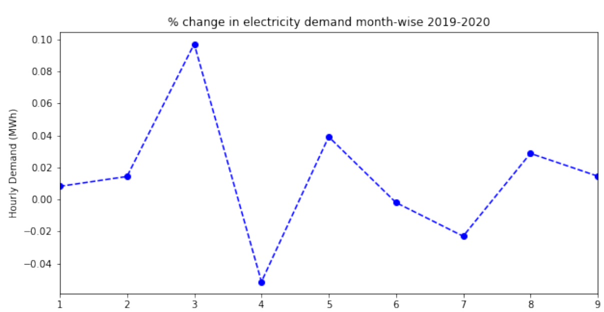

Figure 6: % change in demand month-wise 2019-2020

Result

Our model which is based on Gradient boosting algorithm gives highly accurate electricity demand prediction and can be used to detect shift in demand pattern due to pandemic breakouts. The below figure-7 is the month-wise correlation of electricity demand for 2019-2020, it measures the similarity in electricity demand pattern between months of year 2019 and 2020. For the month of March & April there is a change in the pattern of demand which can be seen by low correlation value for these two months, and it is due to state of emergency declared on March 19, 2020 in California. So the demand trend changed from normal months. For the rest of the months the correlation value is very high around 0.99, which shows that the disruption due to pandemic only lasted for these two months. And there is no change in demand pattern after May, 2020. All these results shows that our designed model is capable of detecting any change in the pattern of electricity demand and this can be helpful in detecting pandemic breakouts.

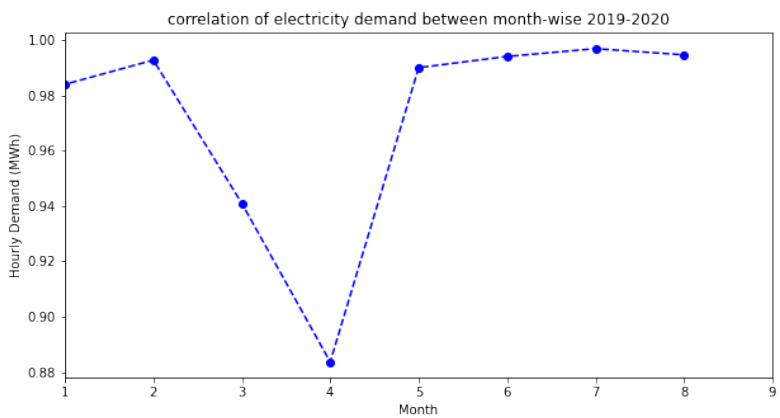

Figure 7: Month-wise correlation of electricity demand between 2019 and 2020

Appendix

[1] The electricity data set is taken from the following site:
`https://www.ncdc.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:23174/detail`

[2] The weather data is taken from the following site:
`https://www.eia.gov/beta/electricity/gridmonitor/dashboard/electric_overview/US48/US48`