2019 SSC Case Studies in Data Analysis Poster Competition

|  |  |
| --- | --- |
| Case study # | 1 |
| Team Lead and Members | Mathilde Bourget and **Gabriel** **Morin** |
| Department and University | Department of mathematics, UQAM |
| Supervisor | Mathieu Boudreault |
| Project Title | Predicting Hourly Electricity Demand in Ontario |

INTRODUCTION :

In order to have enough electricity production each hour to fulfill Ontario’s residential demand, it is essential to have the hourly demand data. Unfortunately, this data is not available publicly and so statistical models must be used to estimate the demand.

OBJECTIVE :

Our main objective was to predict the hourly demand of electricity for the residential sector in Ontario with accuracy using time and weather as predictors. Our priority was accuracy, so we wanted a model complex enough to capture hourly variations in electricity demand.

METHODS :

We used a random forest algorithm with 500 trees and 12 randomly selected variables at each split. To train our model, we split our data using 70% for training and 30% for validating. The database used for the model was a merge of the Hourly Demand database and the Weather database. Since the electricity demand in the Hourly database was all sectors aggregated, we had to tweak it in order to consider only residential demand.

*To do so, we started by calculating a ratio of the annual residential demand over the total annual demand using the Annual Demand database from the CER. For each year, we had a ratio representing the proportion of residential electricity demand. Then, we had to consider the variability within the years (we wanted to give weight to each year based on the total annual electricity demand). Within the annual residential demand, we then calculated a ratio of the electricity demand for each year over the total annual residential demand from the 14 years. We did the same with the annual demand for all sectors aggregated. Doing a ratio of those ratios, we ended up with a weight for each year based on their total electricity demand.*

*We multiplied the hourly aggregated electricity demand by those two ratios and ended up with a variable that we called ‘Load\_Mw’, which represented an estimate of the hourly residential electricity demand. We had to calculate this variable to be able to train our model.*

Most variables were kept, except for the following : Date (since it was unique for almost each line), irradiance\_surface (since it was not linked to the electricity demand), Year (since it was too influent and we wanted our out-of-sample data to perform great) and densite\_air (since it was too correlated with the temperature variable). We also added some variables while trying to capture the variations in hourly residential electricity demand.

RESULTS :

With a MAE of approximately 290 on our validation data, and a R-squared of roughly 0.82, the predictions are quite accurate. The Year variable made the MAE even smaller, but since the purpose of this model is to be used for future years, we had to keep it out since it is not reliable enough on a 10 years training data. We could have kept if we had more data.

CONCLUSION :

In conclusion, the variables that predict best the hourly residential electricity demand are the price of electricity and whether the day observed is a holiday. By using our models, electricity producers could predict the hourly demand only by using data accessible to them. This model could also be used to analyse energy consumption and educate the public through government program on how to save energy.