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 from the h2o package in R. To train our model, we use leave-one-out cross validation (LOO-CV) by leaving a year out of our training data. We applied the model on 14 different validation data, one for each year that was left

out. The database used 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 used the Annual Demand database and we interpolated the residential electricity demand to obtain the hourly residential electricity demand.

To do so, we started by calculating a ratio for each year of the annual residential demand over the total annual demand using the Annual Demand database.Then, we had to consider the variability within the years. Within the annual residential demand, we then calculated a ratio of the electricity demand for each year over the total annual residential demand from all 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.

RESULTS :

With a MAE of approximately 212 on our out-of-sample data, the predictions are quite accurate. The Year variable was meaningful, 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 13 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 mean of the demand for the weekday in question and the hour of the day. 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.