Data Challenge

Problem Definition:

1. Identify the problem and how to address it using machine learning

The installation of high-efficiency appliances across all rooms on a ship, or even a fleet, is a costly and complex undertaking. The data science team must develop an algorithm that can predict total energy consumption into the future, so that lifetime environmental and financial savings can be included in the decision-making process.

Data has been collected for 5 months from January 11th 2016 to May 27th 2016, in intervals of 10 minutes. The first iteration of the algorithm will predict the total energy consumption values (target variable) of the month of May, up until the 27th , given the predictor and target variables from January 11th 2016 to April 30th 2016.

Proposed Solution:

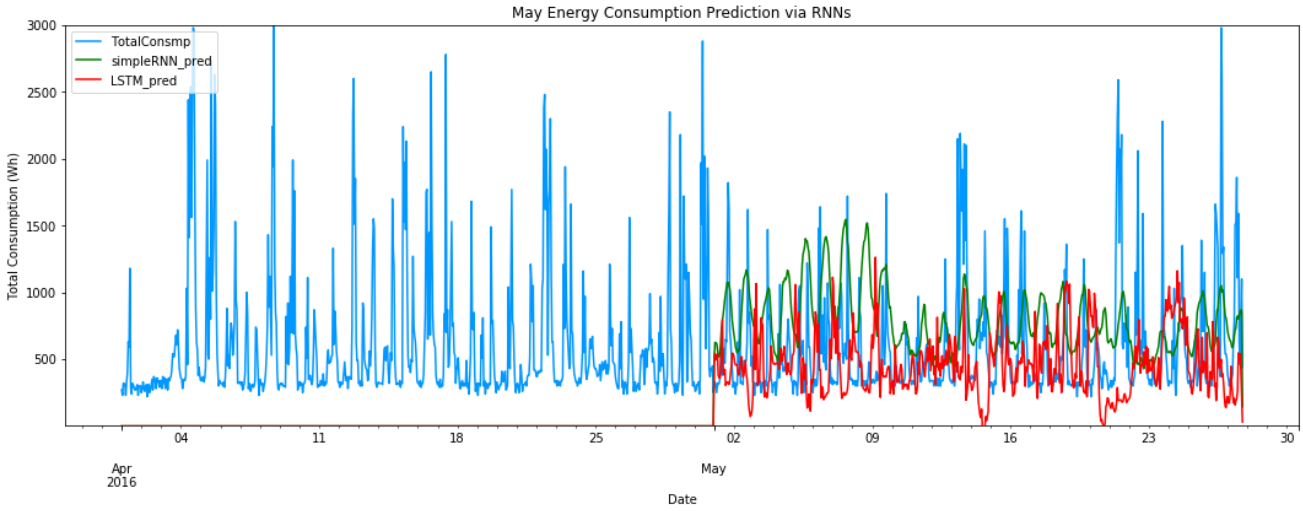
After data cleaning and feature engineering, 5 predictor variables remained: *RTo*, the temperature difference between the mean room temperatures and the outside temperature, *H*, the mean room Humidity, and the outside windspeed, visibility, and pressure, all in their original units.

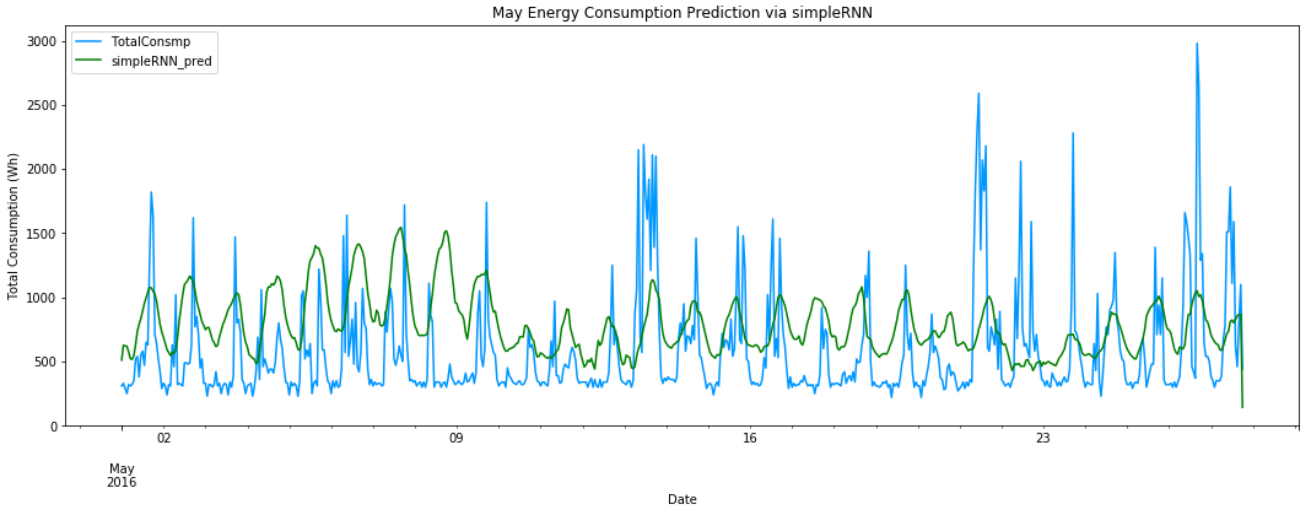
Initially classic regression algorithms were utilized, MLPRegressor, Lasso Regression, SVR. It was soon understood that because the target variable is continuous, an algorithm capable of remembering the previous its previous value was required, it was time for Recurrent Neural Networks (RNN). With little experience in deep learning techniques the author assembled a simlpeRNN and a Long Short Term Memory (LSTM) network. LSTM are a type of RNN that are performing quite well in time-series regression and speech recognition applications.

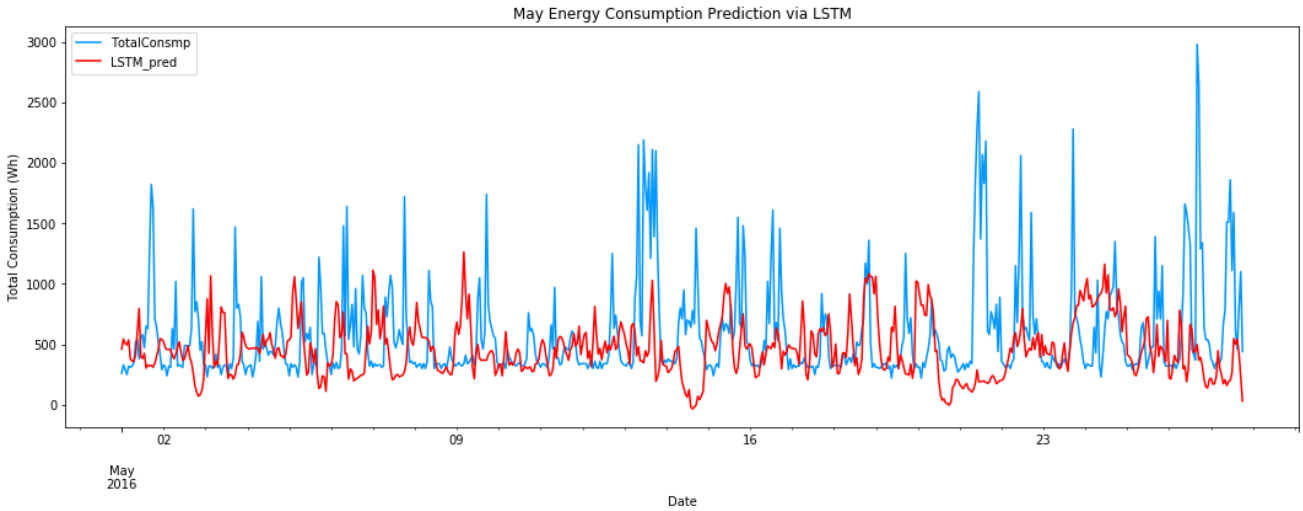
The mean squared error in the testing set for the simpleRNN was of 0.01635 and for the LSTM was respectively 0.00824.

The RNN models can be improved by performing further hyperparameter tuning, (computationally expensive, time-constrained), and by tuning the time frequency of the data.

Please check the jupyter notebooks, included in the folder as .ipynb, html, and pdf. If you have any questions do not hesitate to reach out to alanarvelo@gmail.com. I remain at your disposal.


May Energy Consumption Prediction via RNNs
TotalConsmp
simpleRNN_pred
LSTM_pred
Total Consumption (Wh)
3000
2500
2000
1500
1000
500
04
11
18
25
02
09
16
23
30
Apr
2016
May
Date


May Energy Consumption Prediction via simpleRNN
TotalConsmp
simpleRNN_pred
Total Consumption (Wh)
3000
2500
2000
1500
1000
500
0
02
09
16
23
May
2016
Date


May Energy Consumption Prediction via LSTM
TotalConsmp
LSTM_pred
Total Consumption (Wh)
3000
2500
2000
1500
1000
500
0
02
09
16
23
May
2016
Date