Prediction of Electricity Demand in Ontario Province

**Tolulope Olugbenga**

University of New Brunswick  
3 Bailey Drive

**t.boy@unb.ca**

**Md. Mahbubur Rahman**

University of New Brunswick

3 Bailey Drive

**mrahma15@unb.ca**

**ABSTRACT**

Proper electricity demand forecasting is an important area for power supply companies because it promotes better scheduling and efficiency management. In terms of power supply and demand, For the stable supply of electricity, the reserve power must be prepared [1]. For this reason, power companies have to find better models that can forecast and help with the planning of energy usage. It is essential to forecast power demand because electricity is difficult to store.

In this paper, we got the previous hourly power demand from 2003 until early 2019 from an Independent Electricity System Operator in Ontario, and the hourly weather conditions from a weather station in Ontario’s most populated city Toronto [2], [3]. We classified the dataset to 85% for training our model and 15% for testing. The results were compared with the actual demand. We compared the results on charts using the hourly, daily and monthly averages, we got the Correlation Coefficient, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) for each case. We ran the dataset on three algorithms: Decision Tree, GaussianNB, and Random Forest. Based on the MAPE, the random forest algorithm performed better with a MAPE ranging from 1.77 to 2.23 for both Market and Ontario Demand.

**KEYWORDS:** power demand forecasting, decision tree, gaussiannb, random forest, artificial intelligence

# INTRODUCTION

Electricity is one of the driving forces of economic development and is essential to our daily life and wellbeing. Power demand forecasting is a difficult task due to the number of the different random variables that needs to be taken into consideration in order to predict human behavior. People often use electricity at any time that suits their lifestyle, but for the most part we all happen to use electricity at the same time. Most people share a similar lifestyle pattern, from when we wake up, to having a shower, making some breakfast, leaving for work, coming back at night, going to bed, doing our laundry on weekends and so on.

In this paper, we got the previous hourly power demand from 2003 until early 2019 from an Independent Electricity System Operator in Ontario, and the hourly weather conditions from a weather station in Ontario’s most populated city Toronto [2], [3]. Weather plays a huge role in electricity usage because, in warmer climate people would use more electricity for air conditioning while during the colder times people would use more electricity for heating. We were able to predict the power demand in the province of Ontario with a lot of uncertainties and missing data values. We implemented and compared three algorithms: Decision Tree, GaussianNB, and Random Forest. Based on the MAPE, the random forest algorithm performed better with a MAPE ranging from 1.77 to 2.23 for both Market and Ontario Demand.

# RELATED WORKS

Machine learning is a subfield of Artificial Intelligence [4]. Machine learning has been used to solve different problems in our society today, it has a lot of daily use utilizations like in google searches, text editors, iPhone Siri, etc. These machine learning algorithms are often used for classification problems and pattern detection, they work best with a large dataset. There are some reports that used machine learning to predict power usage and other similar works are highlighted here.

Williams and Gomez performed a study to predict energy consumption for the next month, they implemented three algorithms: Linear Regression, Regression Trees and Multi-variate Adaptive Regression Spline (MARS). The authors used data from homes in Bexar County, Texas containing four years of monthly consumption. After evaluation, the authors achieved a residual mean squared error (RMSE) of 99.803 ± 3.057, 100.435 ± 3.441 and 94.286 ± 3.238 kBtu/day for the linear regression, regression trees and MARS, respectively. The RMSE is calculated by comparing the original average daily energy (kBtu/day) with the predicted average daily energy (kBtu/day). The authors also did an aggregation on daily consumption predictions, to predict the monthly consumption for groups of homes; in this case, the RMSE turned out to be 25.743±1.097, 18.277±1.156 and 19.831±1.187 kBtu/day for the linear regression, regression trees and MARS, respectively [5][6].

Rodrigues and his team wanted to forecast energy consumption of households using an Artificial Neural Network (ANN). The authors collected data from Lisbon, Portugal containing data of 93 households, the data contained hourly energy consumption during 6 to 8 weeks resulting in a total of 93,744 records. Two methods were proposed by the authors: one is used to forecast daily consumption and the other is used to forecast hourly consumption, they both used an ANN with one hidden layer and 20 neurons. After evaluation, the Mean Absolute Percent Error (MAPE) of the average consumption was 4.2% and the maximum consumption was 18.1%, the results were reported in regards to the daily energy consumption [6][7].

# PROBLEM STATEMENT

Electricity consumption is growing at very rapid rate worldwide. This growth presents the need for better planning of energy usage, this includes planning of future electricity demand by electricity distribution companies. Demand response management (DRM) has been one of the major features in smart grid that helps with balancing the electricity demand with the supply available to outages. DRM controls the electricity consumption at the customer’s end, based on the customer’s preset parameters. To have a stable supply of electricity the reserve power has to be prepared. This means electricity companies need to be able to predict the future electricity demand. Power consumption prediction is an important part of the planning and operation of an electricity distribution company. Accurate demand forecasting is important for policy makers to formulate electricity supply policies.

# OUR APPROACH

A typical workflow for demand forecasting is: a) getting the dataset from disparate sources, such as databases or spreadsheets; b) cleaning the data, removing outliers, noise and combining datasets; c) developing an accurate predictive model based on the aggregated data using forecasting techniques; d) deploying the model as an application in a production environment [8]. Our approach was we got the dataset, filtered and combined them, separated some for training and testing of our models, then analyze the results.

## Datasets

This case study consists of two datasets: the past power usage for Ontario and the weather data from one of their major cities Toronto. The power data was gotten from an independent electricity system operator (ieso) [2] and the weather data was gotten from the government of Canada’s website [3]. The power data consists of hourly data from the 1st of May, 2002 and contained 3 headers namely: Date, Hour, Market Demand and Ontario Demand.

Market Demand represents the total energy that was supplied from the IESOAdministered Market; it is calculated by adding all output from generators registered in the market and all scheduled imports to the province; it is also equal to the sum of all load supplied from the market, exports from the province and all lines losses incurred on the IESO-controlled grid [9]. Ontario demand represents the total energy that was supplied from the IESO-Administered Market for the purpose of supplying load within Ontario; It is also equal to the sum of all loads within Ontario which is supplied from the Market, plus all line losses incurred on the IESO-controlled grid [9]. The Market Demand and Ontario Demand are both in Megawatts (MW).

The weather data consists of various weather information, the most important for this case study are Temperature (℃), Dew Point and Relative Humidity. These were the categories that had the most complete information.

## Combining datasets

There was no weather data until the 4th of June, 2002 at 16:00, for this reason we considered data starting from 2003 still early 2019. All missing hourly weather data were replaced with the data from the cell in the previous hour until we had a fully complete dataset; we felt the temperature wouldn’t change as much within two hours. We combined the columns of the power and weather data together. The final dataset was from the 1st of January, 2003 still the 2nd of April, 2019; and it contains 10 headers: - Date/Time, Year, Month, Day, Hours, Market Demand, Ontario Demand, Temperature, Dew Point and Relative Humidity, as seen in the figure below.



Figure 1: The figure shows the top part of the final dataset used.

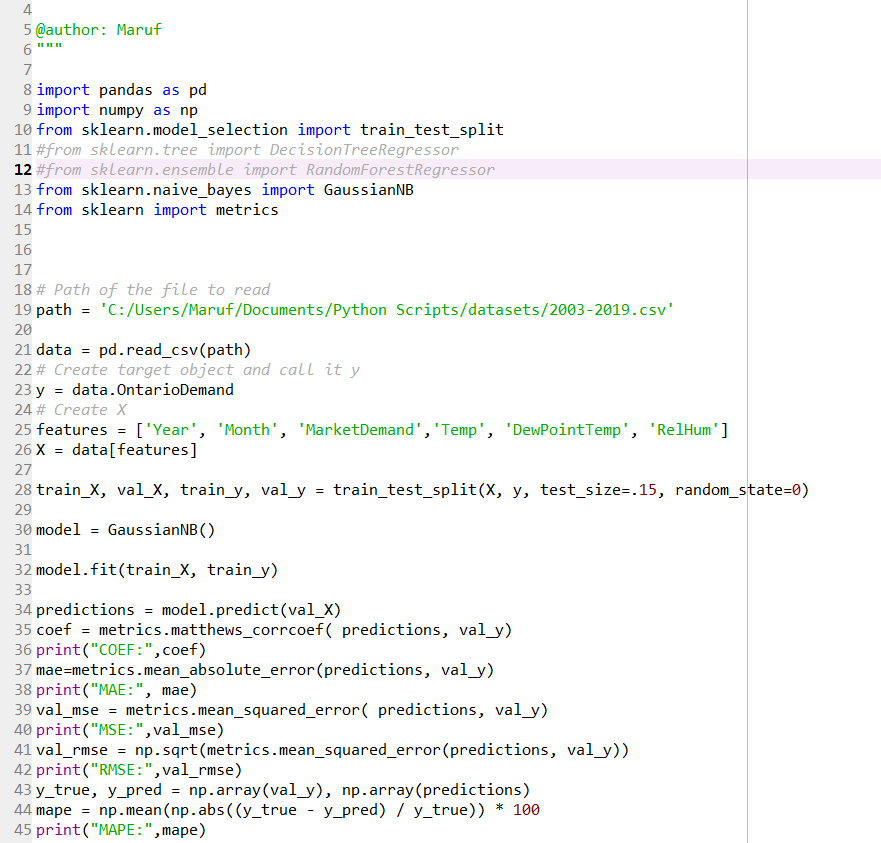
# IMPLEMENTATION

## Design

We have implemented the algorithms with python and the ide we used is Sypder3. We did not develop any system or software. We implemented those algorithms to evaluate the result and compare the results to come to a conclusion that which algorithm works better and provides the most accurate predictions. There are several built-in libraries that we used to implement the algorithms.

## Description of the code/script

We have implemented the algorithms using python and built in libraries of python. The first step was to explore and manipulate the data and for that we have used Pandas. The most important part of Pandas library is the Data Frame which holds the data as a table and Pandas has powerful methods to work with this type of data. NumPy is another important library we used for our implementation. NumPy provides support for large, multi-dimensional arrays and metrics along with a large collection of high-level mathematical functions to operate on these arrays. After exploring data, we selected our prediction target with the features we need to predict that target. Then we split our dataset for training and testing purpose. After the split, we created a model and fit that data into our model. Finally, we predicted and evaluated the predictions with the original dataset.



# Figure 2: Sample Code of Implementation

# 

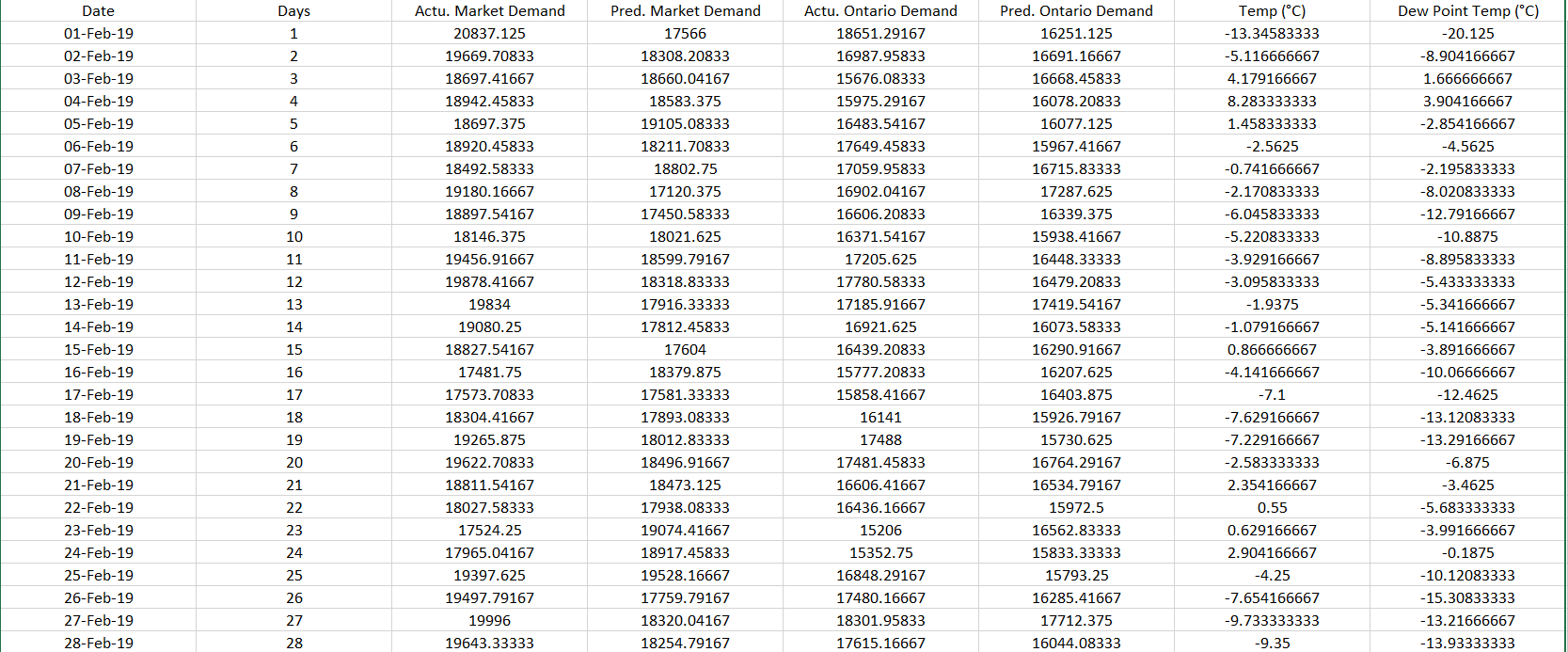
# EVALUATION

## Experimental setup

We have used Spyder3 as IDE and the specification of our device was core i7-4700HQ quad core processor, 8GB DDR3 ram, and 1TB SATA drive. The Spyder3 IDE has its own kernel to display the results. We divided the dataset into two part, one is for training purpose and other one for testing. We used 85% of the data for training the algorithms and 15% data to test. Moreover, we tried to use as many features as possible so that the prediction result becomes more accurate. We represented our predictive data and validation data using Excel sheet. We displayed our validation data in hourly, daily and monthly basis. We have represented the comparison results in graphical view for a better understanding.

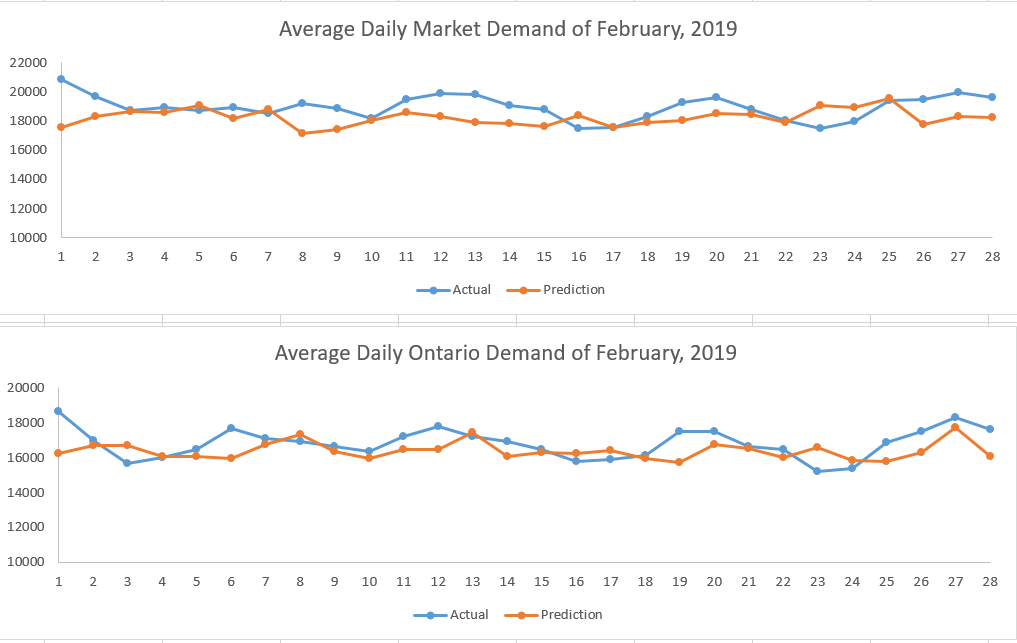
## Experimental results

We have tried to compare the results we got by implementing all the algorithms in tabular form as well as in graphical view. Figure 3 displays the daily predictions of Market Demand and Ontario Demand using the Decision Tree Regressor of the month February 2019.



**Figure 3: Market Demand and Ontario Demand Prediction for Month February 2019**

Figure 4 displays the comparison graph of average daily original Market Demand with predicted Market Demand as well as average daily original Ontario Demand with predicted Ontario Demand of the month of February 2019 of the results displayed above.

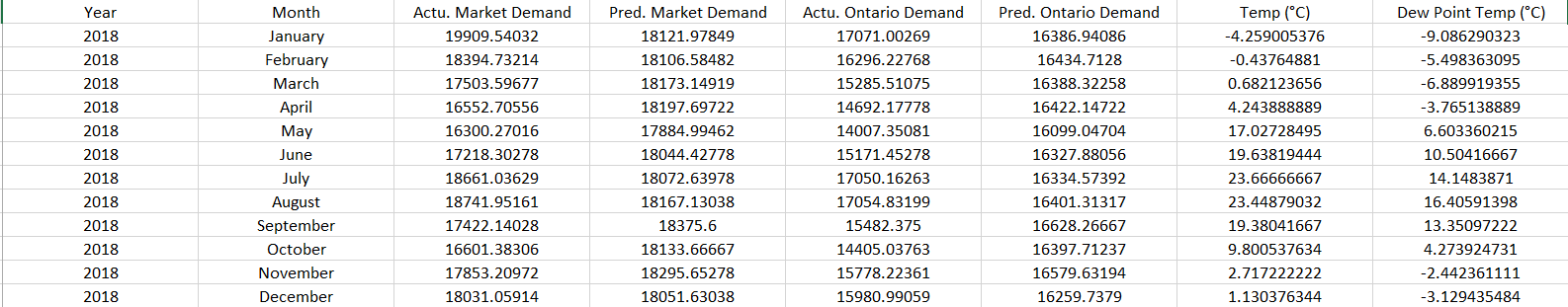


**Figure 4: Average Daily Market Demand and Ontario Demand Comparison**

Figure 5 displays a comparison graph of average daily temperature and dew point of the month February 2019.

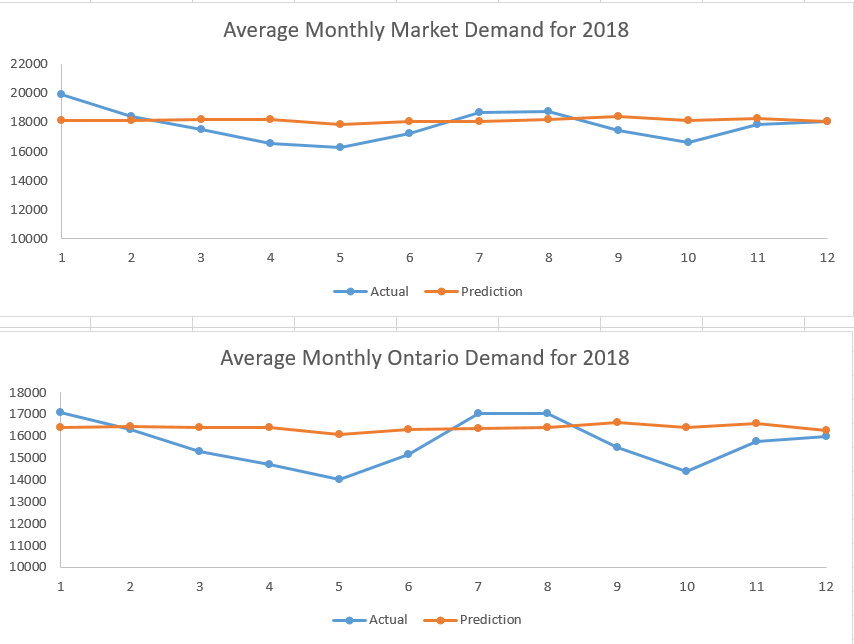
# Figure 5: Average Daily Temperature Vs Dew Point

The Figure 6 displays the monthly predictions of Market Demand and Ontario Demand using the GaussianNB (Gaussian Naïve Bayes) of the year 2018.



**Figure 6: Market Demand and Ontario Demand Prediction for 2018**

Figure 7 displays the comparison graph of average monthly original Market Demand with predicted Market Demand as well as average monthly original Ontario Demand with predicted Ontario Demand of the year of 2018 of the results displayed above.

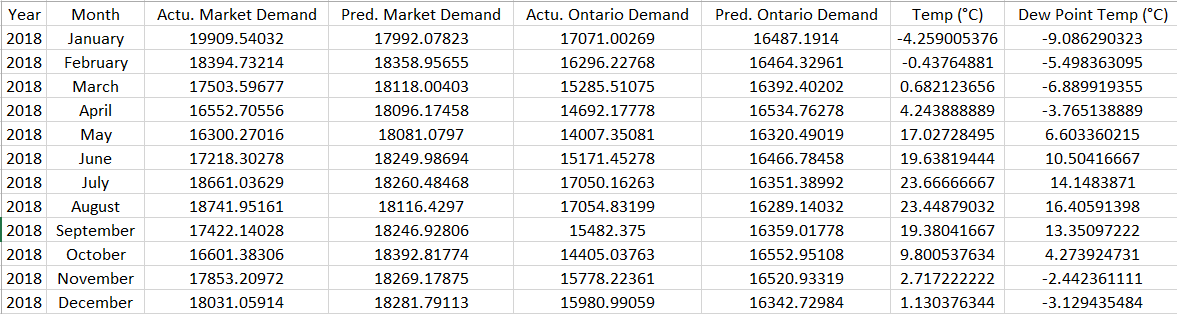


**Figure 7: Average Monthly Market Demand and Ontario Demand Comparison**

Figure 7 displays a comparison graph of average daily temperature and dew point of the month February 2019.

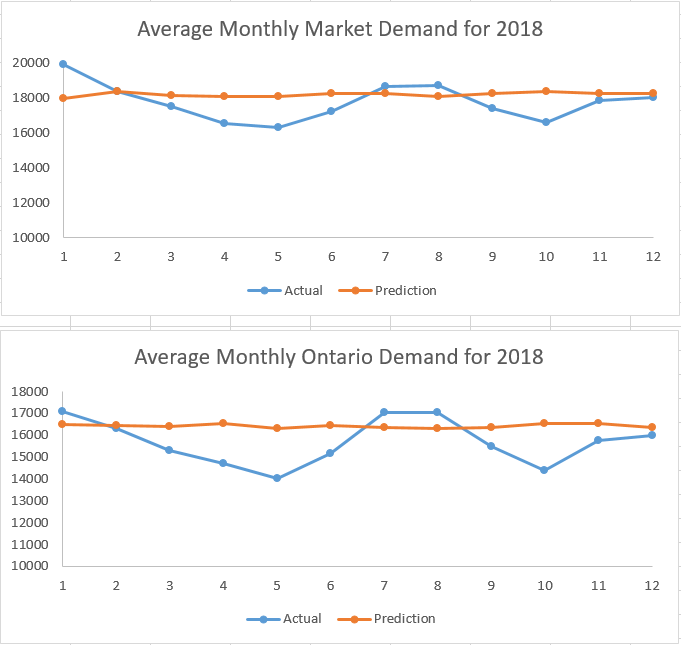
# Figure 7: Average Monthly Temperature Vs Dew Point

The Figure 8 displays the monthly predictions of Market Demand and Ontario Demand using the Random Forest Regressor of the year 2018.



**Figure 8: Market Demand and Ontario Demand Prediction for 2018**

Figure 9 displays the comparison graph of average monthly original Market Demand with predicted Market Demand as well as average monthly original Ontario Demand with predicted Ontario Demand of the year of 2018 of the results displayed above.



**Figure 9: Average Monthly Market Demand and Ontario Demand Comparison**

Figure 10 displays the comparison results of all the validation matrices (Correlation Coefficient, Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and Mean Absolute Percentage Error) on hourly, daily, and monthly basis.



**Figure 10: Comparison of Validation Matrices**

From the validation matrices we can come to a conclusion that overall Random Forest Regressor performed better that other algorithms. But in certain case like for Mean Squared Error and Root Mean Squared Error GaussianNB provided better result sometimes close to Random Forest Regressor. Decision Tree Regressor preformed worst among all the three algorithms. But for Correlation Coefficient Decision Tree Regressor performed close to the Random Forest Regressor.

# CHALLENGES

During the whole project we have encountered some challenges. The first challenge was to collect the dataset. We needed a consistent large dataset with sufficient number of features. At first our plan was to implement our algorithms with UNB’s (University of New Brunswick) dataset but the dataset we collected did not meet our requirements and that’s why we experimented with Ontario dataset. Another challenge was to find out proper algorithms to compare which can provide us continuous values. The paper we followed has compared different algorithms than ours because at first, we wanted to just follow the steps of the paper but later on we wanted to experiment with some new algorithms. So, we had to spent time finding out the proper algorithms with our requirements. The final challenge was to manage the time and complete the project within the time frame. In the beginning we started individual project and only about two weeks ago we built the team and started working together. So, managing the time was one of the crucial challenges for us.

# INDIVIDUAL CONTRIBUTION

We both contributed equally to this project. Tolulope searched and analyzed different datasets, he found and imported the dataset from ieso and the weather station [2], [3]; he filtered and cleaned the datasets, catered for empty or missing data; he classified and compared the results on charts and tables among the different algorithms that were implemented. So, he was mainly responsible for all the datasets handling and results evaluation.

Rahman experimented the possible algorithms for the project and chose three algorithms for implementation. He implemented all the algorithms with Python and extracted all the predicted results in the excel file. As well as he implemented all the validation matrices to compare the efficiency of all the algorithms. So, he mainly responsible for all the implementation required for the project.

We wrote the project report together.

# CONCLUSIONS

In this project we mainly compared the results of different machine learning algorithms to evaluate their performance. Though we planned to implement algorithm to predict the future demand, for the time constraint we could not implement that. So, we tried to find out which algorithm would perform better to predict within a range of time and compare the predictions with the original data. From the validation matrices result we can conclude that Random Forest Regressor performed better but, in some cases, GaussianNB performed close or better than Random Forest Regressor. For future work, we plan on developing a model that can predict the power demand for the future using deep learning approaches that contains one or more hidden layers e.g. Convolutional Neural Networks (CNNs).

# REFERENCES

[1] T. Kim and S. Cho, *Intelligent Data Engineering and Automated Learning – IDEAL 2013*, vol. 8206. Springer International Publishing, 2013.

[2] Independent Electricity System Operator, “Hourly Demand Report.” [Online]. Available: http://reports.ieso.ca/public/Demand/. [Accessed: 23-Apr-2019].

[3] “Station Results - Historical Data - Climate - Environment and Climate Change Canada.” [Online]. Available: http://climate.weather.gc.ca/historical\_data/search\_historic\_data\_stations\_e.html?searchType=stnProv&timeframe=1&lstProvince=ON&optLimit=yearRange&StartYear=2002&EndYear=2019&Year=2019&Month=4&Day=22&selRowPerPage=25&txtCentralLatMin=0&txtCentralLatSec=0&. [Accessed: 23-Apr-2019].

[4] Martin Hansen, “Prediction of Electricity Usage Using Convolutional Neural Networks,” 2017.

[5] K. T. Williams and J. D. Gomez, “Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach,” *Energy Build.*, vol. 128, pp. 1–11, Sep. 2016.

[6] R. F. Berriel, A. T. Lopes, A. Rodrigues, F. M. Varejao, and T. Oliveira-Santos, “Monthly energy consumption forecast: A deep learning approach,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2017–May, pp. 4283–4290, 2017.

[7] F. Rodrigues, C. Cardeira, and J. M. F. Calado, “The Daily and Hourly Energy Consumption and Load Forecasting Using Artificial Neural Network Method: A Case Study Using a Set of 93 Households in Portugal,” *Energy Procedia*, vol. 62, pp. 220–229, Jan. 2014.

[8] “Data-Driven Insights with MATLAB Analytics: An Energy Load Forecasting Case Study - MATLAB &amp; Simulink.” [Online]. Available: https://www.mathworks.com/company/newsletters/articles/data-driven-insights-with-matlab-analytics-an-energy-load-forecasting-case-study.html. [Accessed: 23-Apr-2019].

[9] J. Van Der Merwe, “Carrots Monthly Market Report,” *Amt*.