Utilizing Machine Learning for Short-Term Water Demand Forecast

Waleed El Damaty

Water Operations Data Analyst

Waleed.eldamaty@ewec.ae

Co-authors: Mohamed Abdallah, Khalid Al Zaabi

Overview

- Importance of demand forecast
- Factors affecting demand forecast
- Machine learning model flow chart
- Model training
- Model performance
- Model evaluation & results
- Conclusion
- Way forward and next steps

Why Accurate Short Term Water Demand Forecast Is Important?

- Effective water supply system management strategies are needed to maintain a balance between supply and demand for water. This balance is attained through operational measures, many of which call for accurate water demand forecast.

1. Adequately plan for transmission outages and identify potential water supply shortage in advance, thus improving the mitigations procedures carried out by stakeholders.
2. Determine the required water reserve precisely.
3. Develop optimum operational plans for pumping stations and water production plants.
4. Can be utilized in Hydraulic models to improve the modelling of the water network and enhance the accuracy of the simulation.

Water Demand Forecast Factors and Methods

- Climatic conditions: it affects the short-term forecast, such as temperature, relative humidity, precipitation, lagged demand, etc.

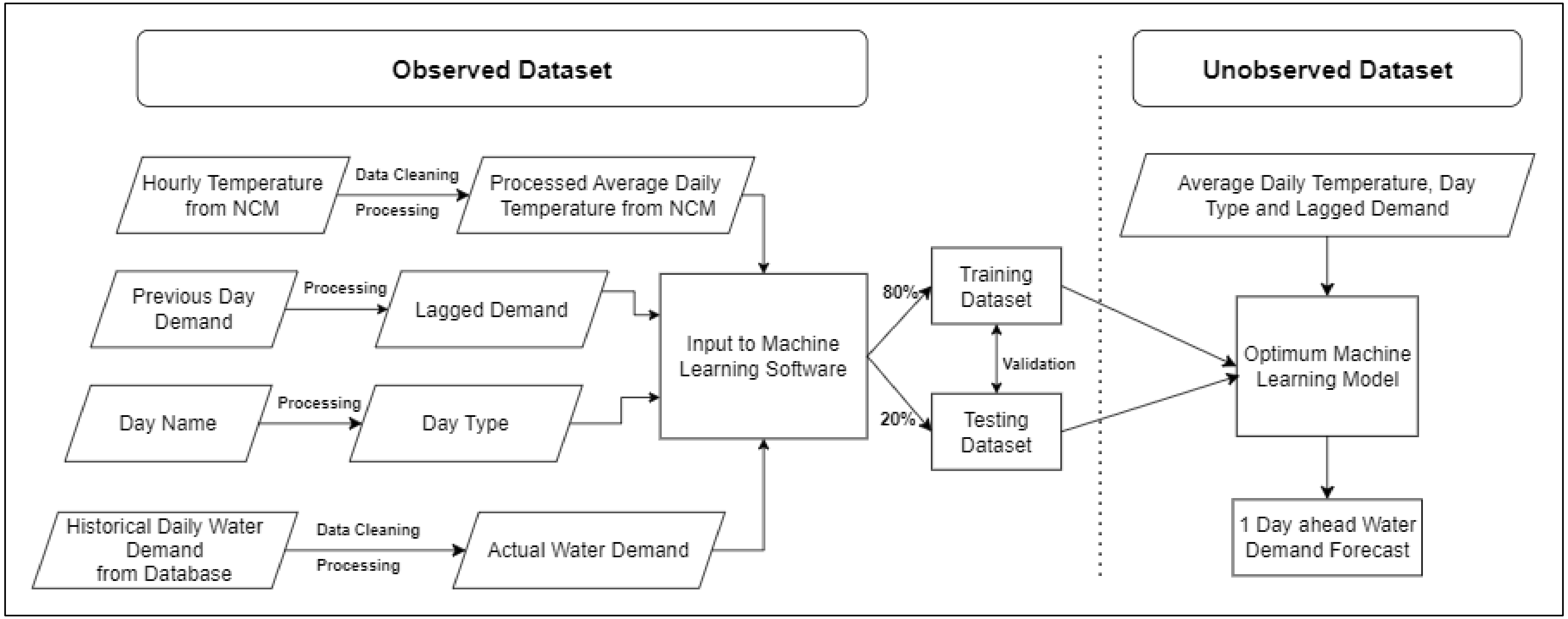
- Socio-economic: it affects the long-term forecast such as population growth, GDP, Tariff rate, etc.

- The consensus is that there isn’t a global single optimum method used to forecast the water demand.

- It is more of a case-by-case approach depending on the network complexity, situation, operational limitations, available data, forecast horizon, availability of tools, accepted level of deviation.

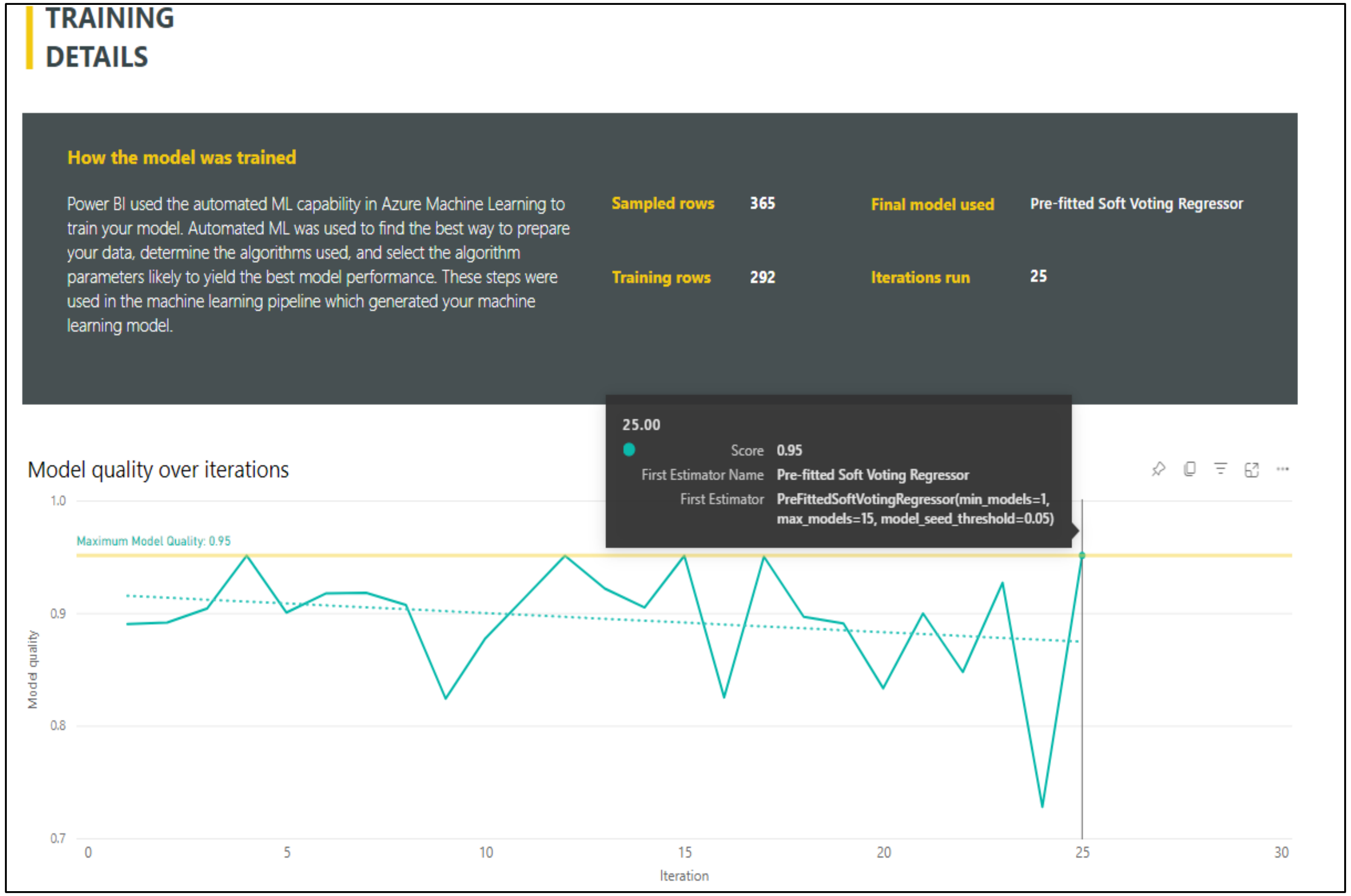
- Machine learning was chosen due to its ability to deal with complex models.

Machine Learning Regression Model Flow Chart



Observed Data Set: Regression Model Training

- Light Gradient Boosting Model (GBM) Regressor
- Elastic Net
- Random Forest Regressor
- Extra Trees Regressor
- Decision Tree Regressor
- Gradient Boosting Regressor



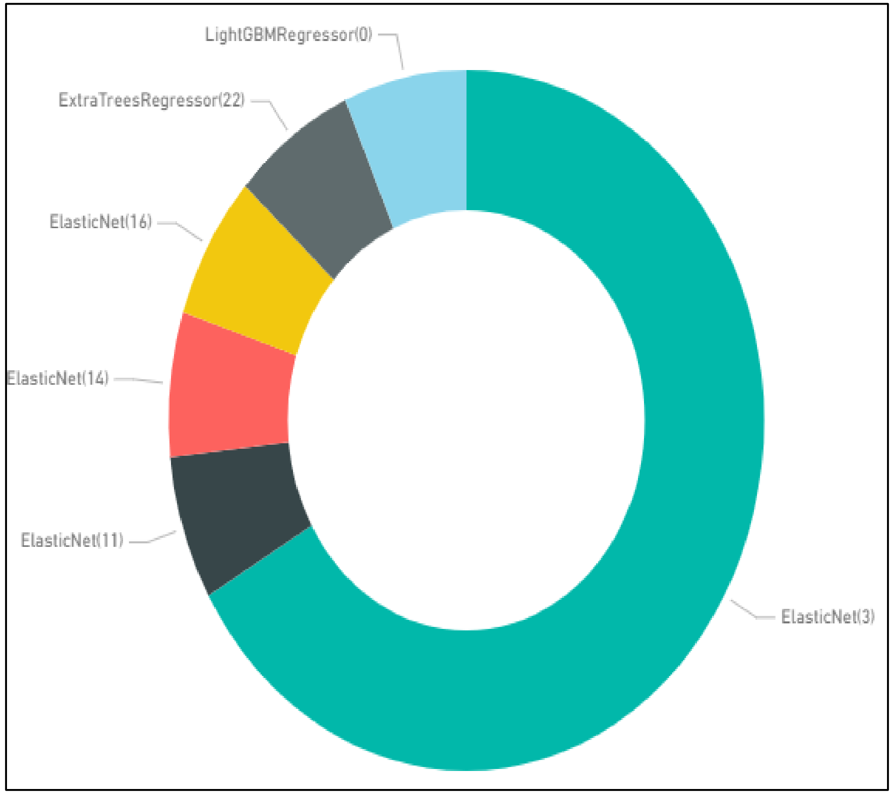
Voting Regressor

- This ensemble method combines the predictions from multiple individual regression models to make a final prediction by leveraging weighted averaging of the predictions of its constituent models, which leads to more accurate and robust predictions compared to individual models.

- The individual models are trained on the training data and during prediction, each model provides its own prediction which are then combined using the weighted averaging to form the final ensemble prediction.

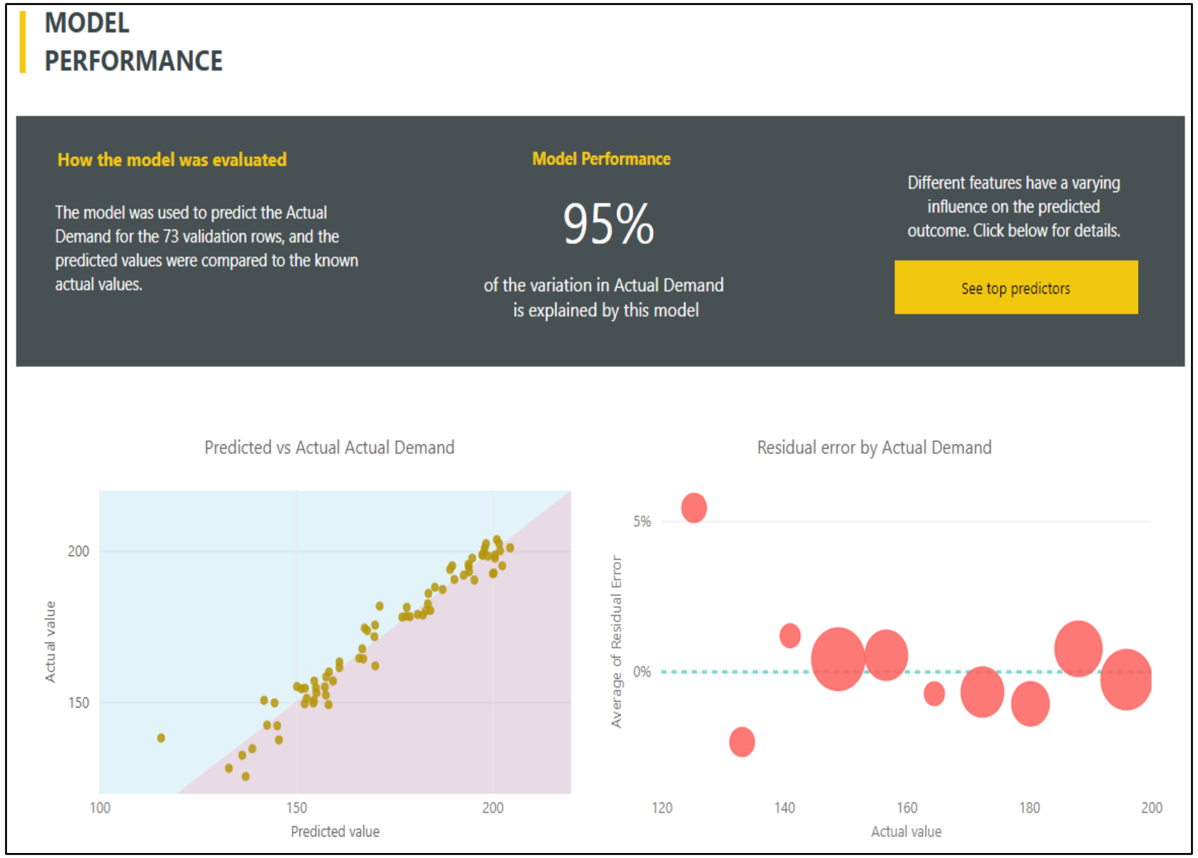
- It is required to include diverse models (traditional, tree-based, etc.). These models can have different strengths and weaknesses which can work as a complement to each other when passed through a Voting regressor.

- The Voting regressor can use simple averaging in which equal weights are assigned for all models or weighted averaging where different weights are assigned to each model's prediction. The weighted averaging allows us to emphasize the predictions of certain models which are expected to perform better on specific subsets of the data.

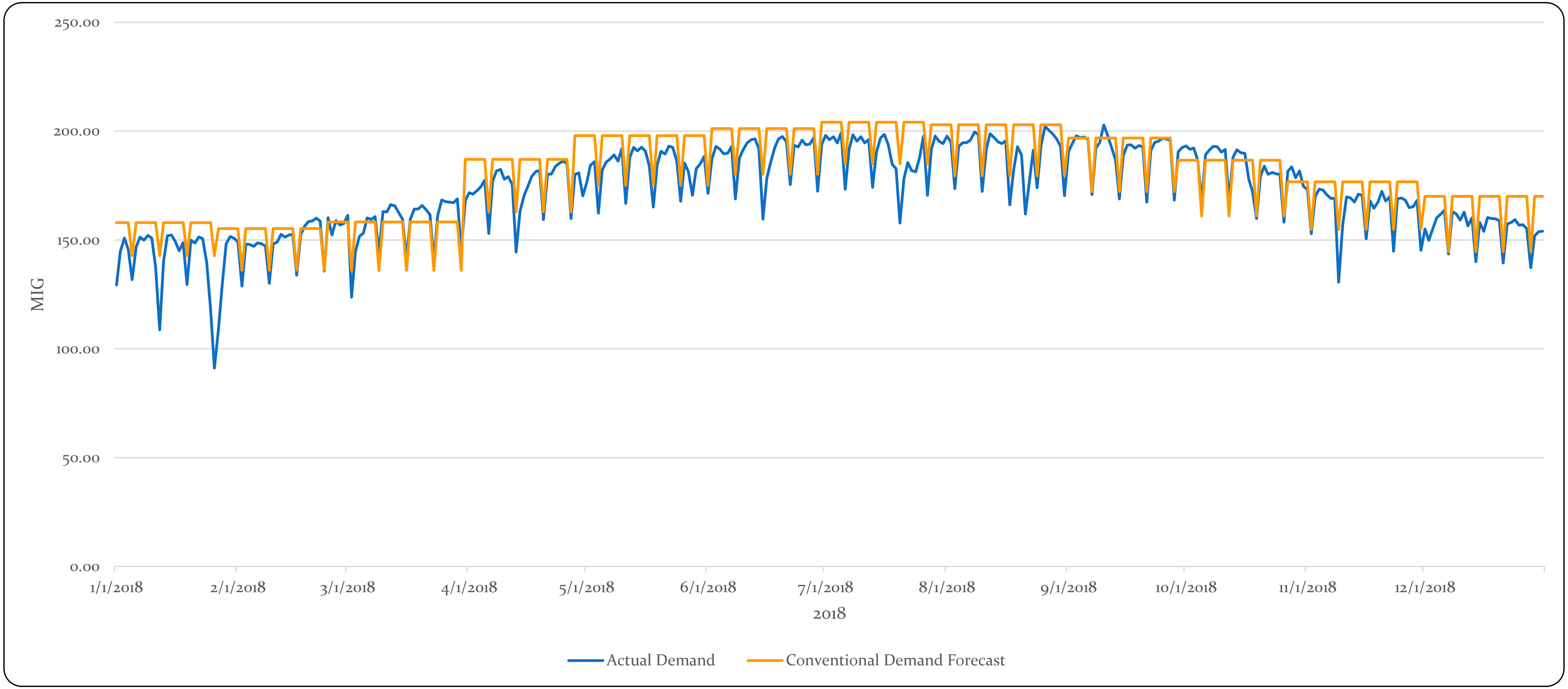


Observed Data Set: Regression Model Performance

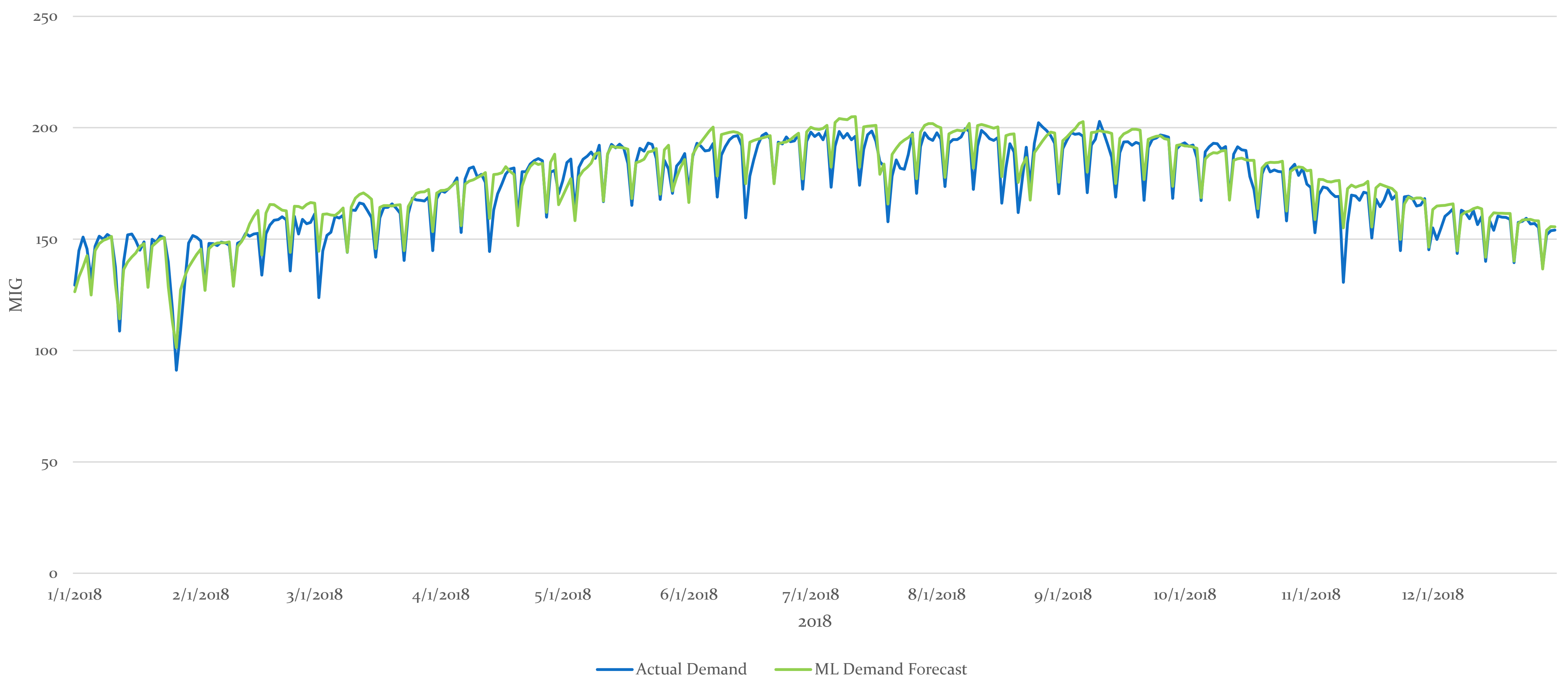
- The values lying in the blue area are underestimated by the model, while those lying in the red area are overestimated by the model.
- The distance from the diagonal is the error in prediction.
- In a good model, most test samples should be clustered near the diagonal.
- The residual error by demand chart represents the distribution of the percent of average error for different values in the sample data set.
- An average residual error of 5% means that the value in that range tends to be overestimated by 5% by the model.
- The size of the bubble indicates the frequency of values in that range in the validation data set.



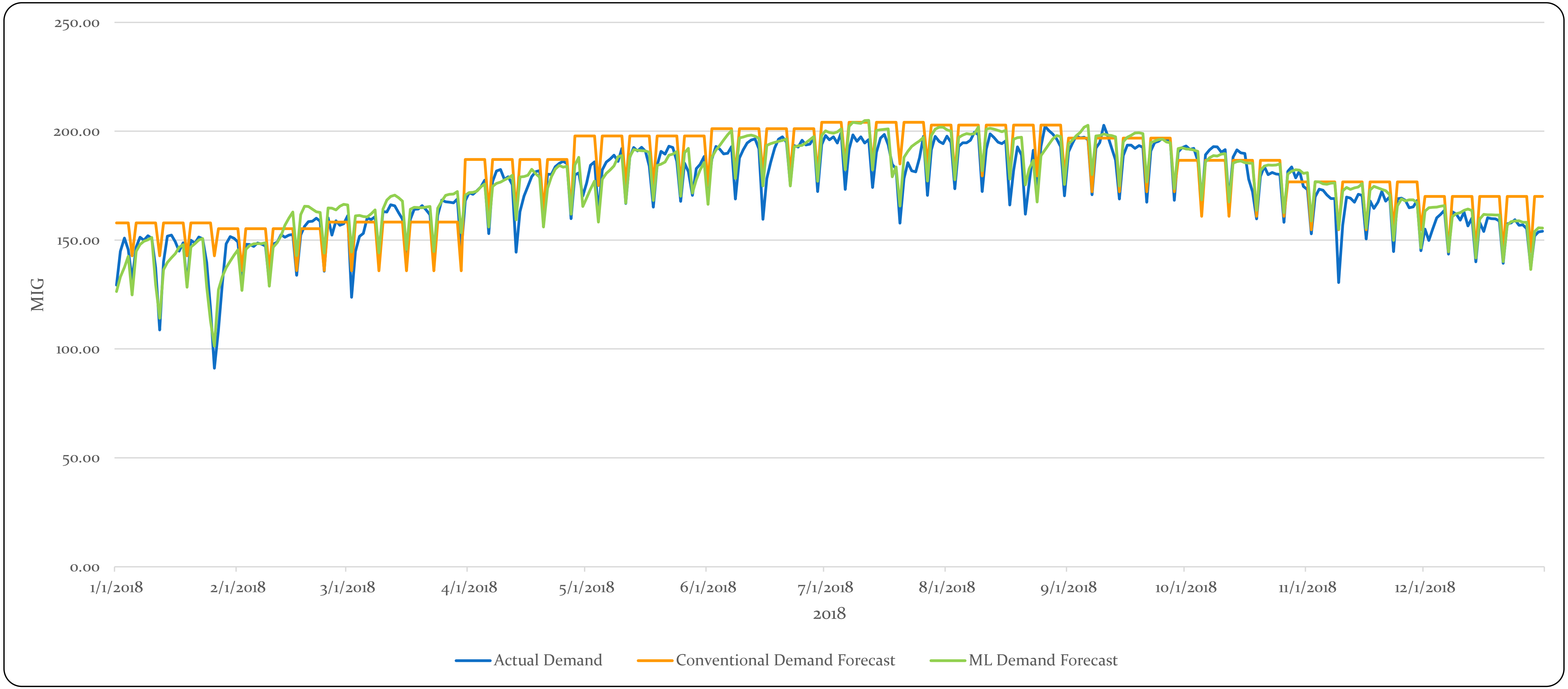
Unobserved Data Set: Actual Vs Conventional Demand Forecast



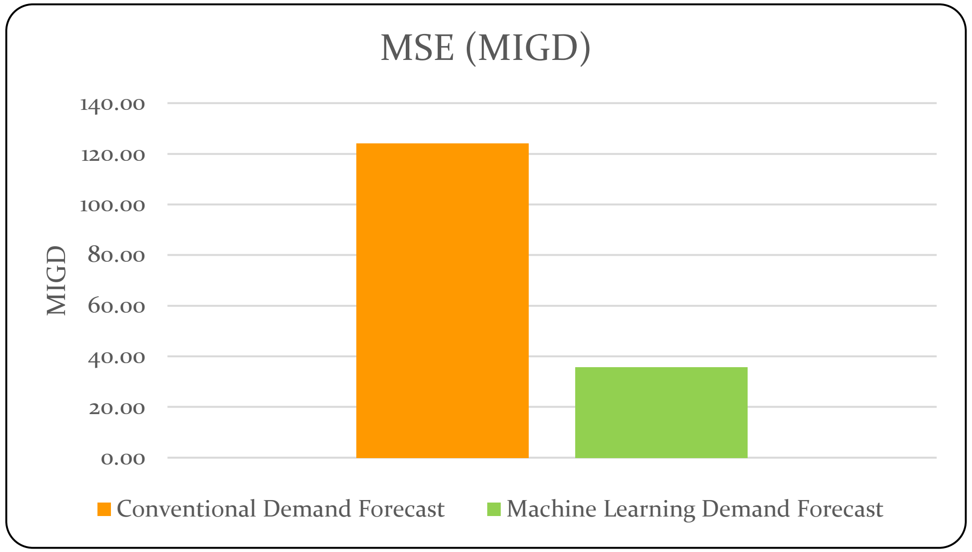
Unobserved Data Set: Actual Vs Machine Learning Demand Forecast

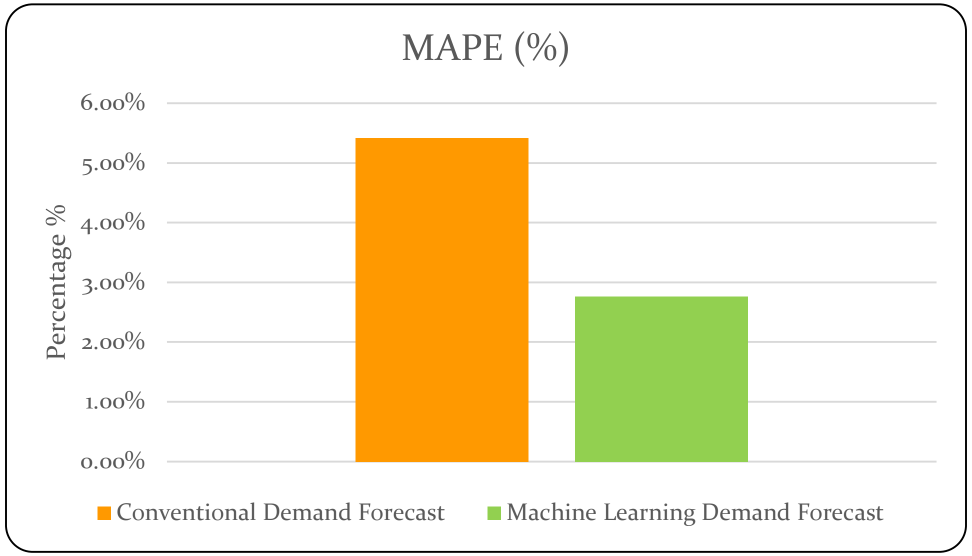


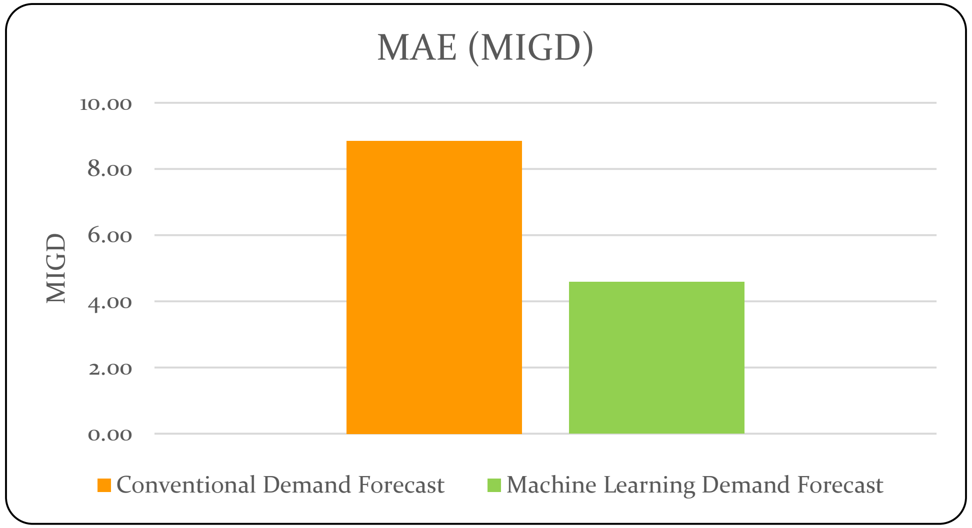
Unobserved Data Set: Actual Vs Conventional Vs Machine Learning Demand Forecast

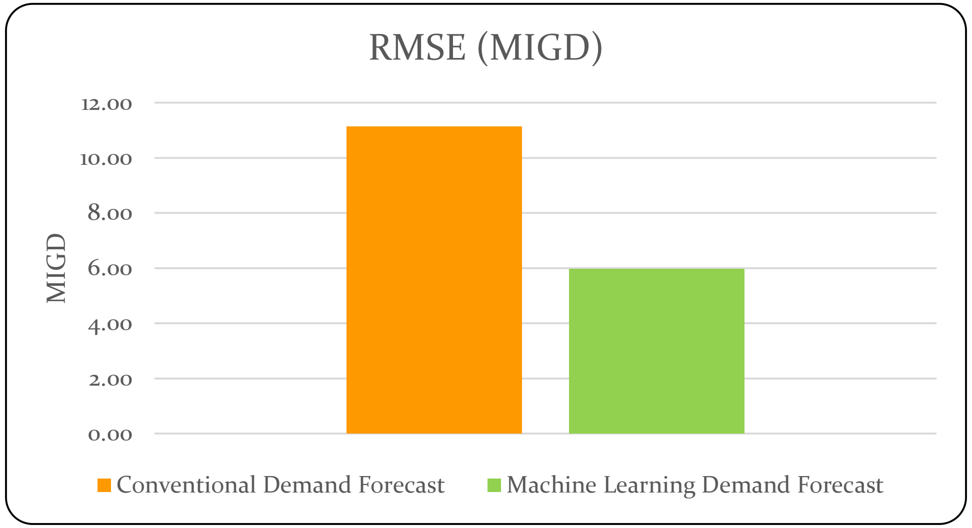


Unobserved Data Set: Model Evaluation and Results

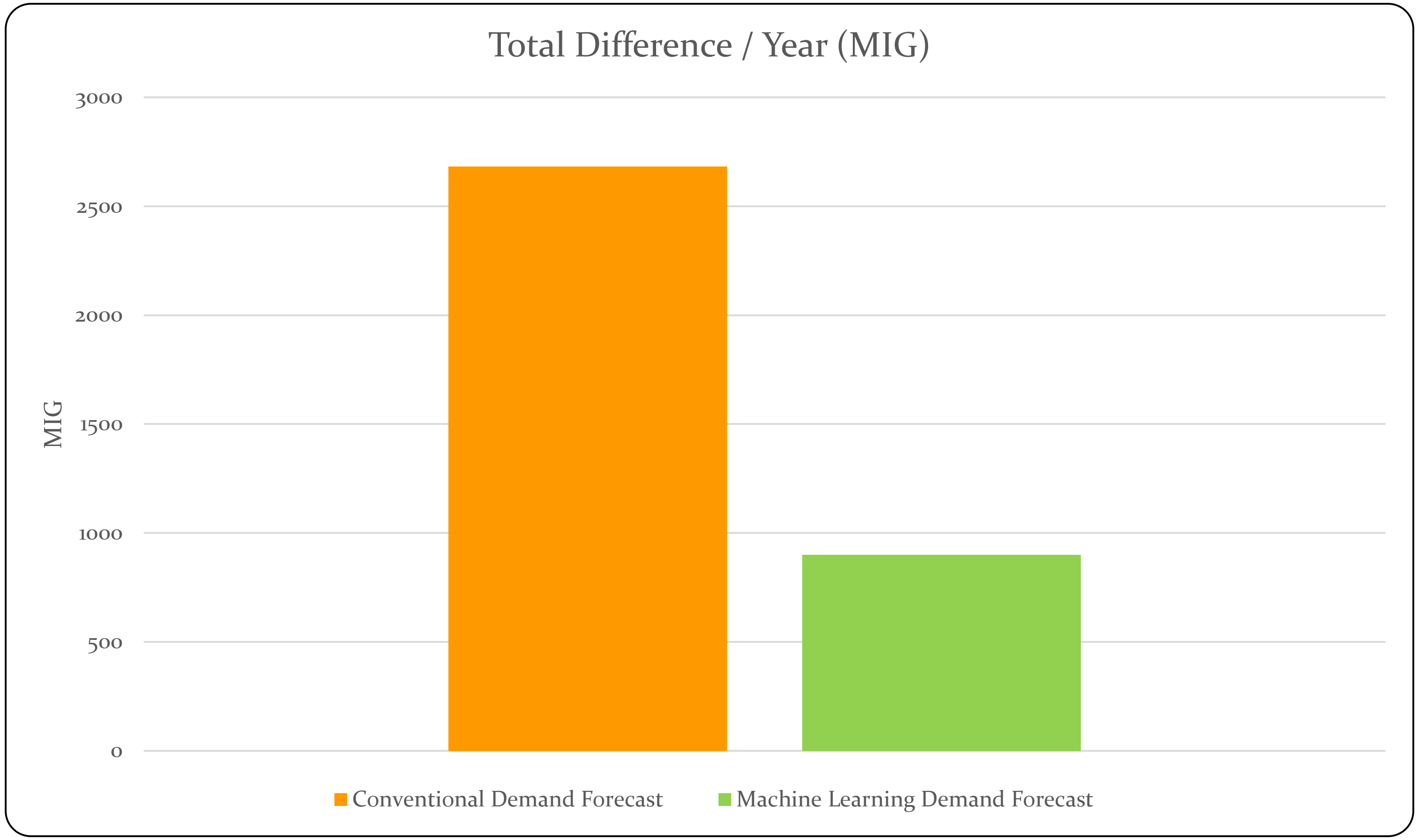








Unobserved Data Set: Model Evaluation and Results



Unobserved Data Set: Model Evaluation and Results

Error Metric	Conventional Demand Forecast	Machine Learning Demand Forecast
MSE	124 MIGD	35.7 MIGD
MAE	8.85 MIGD	4.59 MIGD
MAPE	5.42 %	2.76 %
RMSE	11.14 MIGD	5.97 MIGD
Total Difference / Year	2683 MIG	900 MIG

Conclusion

- The extra saved 1785 Million Imperial Gallons can be utilized to:

1. Adequately plan for water production and transmission outages.
2. Identify potential water supply shortage in advance, thus improving the mitigations procedures carried out by stakeholders.
3. Determine the required water reserve precisely.
4. Develop optimum operational plans for pumping stations and water production plants.

- These endeavours will result in a reduction of the disparity between actual and forecasted demand, thereby improving the overall efficiency and performance of the water management system and contributing to the attainment of global water sustainability goals.

Next Steps and Future Work

- The impact of new different variables (humidity, rain, etc.) on the model performance is being tested.
- Different algorithms are going to be tested to compare their performance in comparison to the currently used model.
- Risk assessment will be conducted to mitigate any potential risks arising from significant difference between actual and forecasted demand as a result of extreme events.
- Cutting-edge technologies and highly intricate algorithms alone cannot address the issue of water scarcity.
- It is imperative to establish appropriate regulations and a solid legal framework to ensure an efficient process

Questions?

EWEC
Emirates Water & Electricity Co.