**CHMLTech- Modeling Results**

**Dataset Description**

|  |  |  |
| --- | --- | --- |
| **Features** | **Description** | **Range [Min – Median – Max]** |
| **DateTime** | **Timestamp for each observation in hourly intervals.** | **-** |
| **Load Demand (kWh)** | **Total energy demand of the system, including building loads and auxiliary components.** | **15000.1 < 17536.7 < 19999.7** |
| **PV Production (kWh)** | **Energy generated by the photovoltaic panels.** | **0.0 < 9.9 < 1500.0** |
| **Energy Exported to Grid (kWh)** | **Surplus energy sent back to the grid.** | **10.0 < 54.7 < 100.0** |
| **Energy Imported from Grid (kWh)** | **Energy drawn from the grid to meet demand.** | **10.0 < 104.4 < 199.9** |
| **Hydrogen Production(Kg)** | **Amount of hydrogen produced by the electrolyzer.** | **20.0 < 34.9 < 50.0** |
| **Fuel Cell Output(kWh)** | **Energy generated by the fuel cell using hydrogen.** | **10.0 < 25.1 < 40.0** |
| **Ambient Temperature (°C)** | **Temperature at the system location, impacting efficiencies..** | **15.0 < 23.6 < 45.0** |
| **Electrolyser Power (kW)** | **Power consumed by the electrolyzer.** | **100.0 < 400.0 < 1499.7** |
| **Electrolyser Efficiency (%):** | **Efficiency of the electrolyzer in converting electricity to hydrogen.** | **60.0 < 62.5 < 65.0** |
| **PV Efficiency (%)** | **Efficiency of the photovoltaic system in converting sunlight to electricity.** | **13.0 < 14.0 < 15.0** |
| **Fuel Cell Electrical Efficiency (%)** | **Efficiency of the fuel cell in converting hydrogen to electricity.** | **40.0 < 42.5 < 45.0** |
| **Converter Efficiency (%)** | **Efficiency of the converter in transforming energy.** | **94.0 < 95.0 < 96.0** |
| **Solar Irradiance (W/m²)** | **Solar power per unit area received by PV panels.** | **0.0 < 105.1 < 1000.0** |
| **Hydrogen Supply to Fuel Cell (kg)** | **Hydrogen supplied from storage to the fuel cell.** | **0.7 < 1.8 < 2.8** |
| **Hydrogen Consumption Rate (kg/kWh)** | **Rate at which the fuel cell consumes hydrogen to generate electricity.** | **1. 0.07** |
| **Hydrogen Discharge to Storage (kg)** | **Hydrogen stored in the storage system.** | **17.3 < 33.1 < 49.2** |
| **Hydrogen Flow from Tank to FC (kg)** | **Hydrogen delivered from the storage tank to the fuel cell.** | **0.7 < 1.8 < 2.8** |
| **Building Load Demand (kWh)** | **Portion of the total load demand specific to building operations.** | **6000.3 < 10638.5 < 17996.7** |
| **Energy Supplied to Electrolyzer (E\_PV2EZ) (kWh)** | **Energy from PV used to run the electrolyzer.** | **0.0 < 7.5 < 1192.5** |
| **Energy Supplied to Fuel Cell (E\_NT2FC) (kWh)** | **Energy supplied from hydrogen storage to the fuel cell.** | **10.0 < 25.1 < 40.0** |
| **Levelized Cost of Hydrogen (LCOH) ($/kg)** | **Cost of hydrogen production.** | **2.5 < 3.5 < 4.5** |
| **Levelized Cost of Storage (LCOS) ($/kWh)** | **Cost of storing energy.** | **0.1 < 0.2 < 0.3** |
| **Electrolyzer Losses (kWh)** | **Energy lost during the operation of the electrolyzer.** | **0.0 < 1.0 < 170.7** |
| **Converter Losses (kWh)** | **Energy lost during conversion.** | **756.7 < 1303.8 < 1988.2** |
| **Storage Losses (kg)** | **Hydrogen lost during storage.** | **0.4 < 1.2 < 2.5** |
| **Balance of Plant (BoP) (kWh)** | **Energy used by auxiliary systems within the microgrid.** | **304.5 < 609.9 < 996.7** |

**Data Exploration**

**Correlation**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**Observations**

* The strongest correlations are between PV Production, Building Load Demand, and Electrolyser Power/Losses.
* Hydrogen Production and Storage Losses have strong positive correlation.
* Variables like Energy Exported/Imported from Grid and Solar Irradiance show very weak or no correlation with other variables.

**A collage of blue squares

AI-generated content may be incorrect.Scatter Plot**

The above graph suggests a consistent pattern of efficiency measurements across a wide range of negative values. No relationship can be extracted as most of the data show constant distribution along the range

**Histogram**

A graph showing the value of building low and low

AI-generated content may be incorrect.

The load demand profile shows a single peak in August, however, it shows constant distribution along each month, which tell us about the synthetic nature of data.

**Train-Test Split**

We have considered last two months as testing dataset and starting 10 months as training dataset as shown below:

A blue graph with white text

AI-generated content may be incorrect.

**Model: To predict the Hydrogen production**

**Input Features**: *Solar Irradiance, Ambient Temperature, day, month, year*

**Target**: *Hydrogen Production*

A screen shot of a black screen

AI-generated content may be incorrect.

**Observations:**

* There are high error rates and negative R² values in the above table.
* All the r2 values are negative which indicates that the model is not fitting the data well at all.
* Linear Regression, Lasso Regression, and Ridge Regression are performing similarly and slightly better than the others, but still not well.

**Model: To predict Storage utilization**

**Input Features**: *Load Demand (kWh), PV Production (kWh), Energy Exported to Grid (kWh), Energy Imported from Grid (kWh), Fuel Cell Output (kWh), Ambient Temperature (°C), Electrolyser Power (kW), Electrolyser Efficiency (%), PV Efficiency (%), Fuel Cell Electrical Efficiency (%), Converter Efficiency (%), Solar Irradiance (W/m^2), Hydrogen Supply to Fuel Cell (kg), Hydrogen Consumption Rate (kg/kWh), Hydrogen Discharge to Storage (kg), Hydrogen Flow from Tank to FC (kg), Building Load Demand (kWh), E\_PV2EZ (kWh), E\_NT2FC (kWh), LCOH (/kg),LCOS(/kg),LCOS(/kWh), Electrolyzer Losses (kWh), Converter Losses (kWh),Balance of Plant (BoP) (kWh), day, month, year*

***Target*:** *Hydrogen Discharge to Storage (kg)*

A black screen with numbers and a black background

AI-generated content may be incorrect.

**Observations:**

* We removed high correlation features like Hydrogen Production and Storage losses.
* The results show that most models (Linear Regression, Random Forest, XGBoost, Lasso Regression, Ridge Regression) have negative R² values which means they don’t fit the data well.
* Linear Regression performs the best among the models, with the lowest SMAPE (23.606), MAE (7.645), RMSE (8.767), and MAPE (0.265), and the highest R² (-0.010).
* Lasso Regression also shows competitive performance with slightly better R² (0.007) and similar error metrics. In contrast, XGBoost performs the worst, with the highest errors and the lowest R² (-0.218).
* Overall, simpler models like Linear and Lasso Regression outperform more complex models in this scenario.

**Model: To predict System efficiency**

**Input Features**: *Electrolyser Efficiency (%), PV Efficiency (%), Fuel Cell Electrical Efficiency (%), Converter Efficiency (%),day, month, year*

***Target*:** Average System Efficiency

A black screen with white numbers

AI-generated content may be incorrect.

**Observations:**

* We performed backward elimination to get the best features to predict system efficiency.
* **Linear Regression** performed perfectly with all error metrics (sMAPE, MAE, RMSE, MAPE, MSE) at 0.000 and an R² score of 1.000. This means it predicted the data without any errors, which is unusual and might indicate overfitting or a very simple dataset.
* **Random Forest**: This model did well with low error rates and a high R² score of 0.993. It means it made very few mistakes and explained most of the data variance.
* Like Random Forest, XGBoost also performed well with slightly lower errors
* **Lasso Regression**: This model didn't perform as well, with higher errors and a lower R² score of 0.870.
* Like Linear Regression, ridge regression also had perfect scores with all error metrics at 0.000 and an R² score of 1.000. This suggests it also predicted the data perfectly, which might be due to the same reasons as Linear Regression.
* **Support Vector Regressor** performed well with low error rates and a high R² score of 0.983. It was very accurate in its predictions.

**Summary:**

* For hydrogen production, there are large negative r squared values which means the model is not fitting the data well.
* For Storage Utilization, there are large negative r squared values which means the model is not fitting the data well. Simpler models like Linear and Lasso Regression outperform more complex models
* For System Efficiency, Random Forest and XGBoost have performed well, and the results are reasonable. While Linear regression and Ridge models are too good to be true and might mislead.

**Steps for Synthetic Data Generation**

**A diagram of a diagram

AI-generated content may be incorrect.**

1. Data Exploration - completed
2. We explored and checked the dataset provided.
3. The balancing wasn’t correct. The summatory of inputs and outputs is not equal to zero.
4. The load demand profile was constant throughout the year, it should show some peaks and valleys, some variation.
5. Applied modeling with this dataset won’t give us meaningful results.
6. Creating a new dataset. completed
7. Extracted weather data from OpenWeather real data. The dataset is from Last year to present date. First trials, using Canada locations. Calgary.
8. Created synthetic load demand using some variables (meteorological conditions) for the location selected. Temperature, Humidity, Wind Speed.
9. Also, we use a measure of Average Energy Use Intensity (EUI) from Statistics Canada.
10. To complete the calculations, we came up with an area that matches the building from the EUI list. Approximation.
11. Extracted solar irradiance data. NASA POWER real data. Last year to date. In progress
    * Same location as weather.
    * Calculated PV Production
    * Calculated Converter Losses.
    * Balancing Equation.
12. Results from this balancing give us the amount of energy surplus or deficit, this difference is energy delivered to or taken from the Grid. In progress
13. Prepare Database management draft. In progress
14. Prepare Visualization and Dashboard draft. In progress