A Machine Learning Pipeline for the Preventive Maintenance of a Building HVAC System

Elizabeth Chelmecki, Anoja Muthucumaru, Chi Zhang, Shirley Zhang, Sherry Zuo

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Institute for Multidisciplinary Design & Innovation

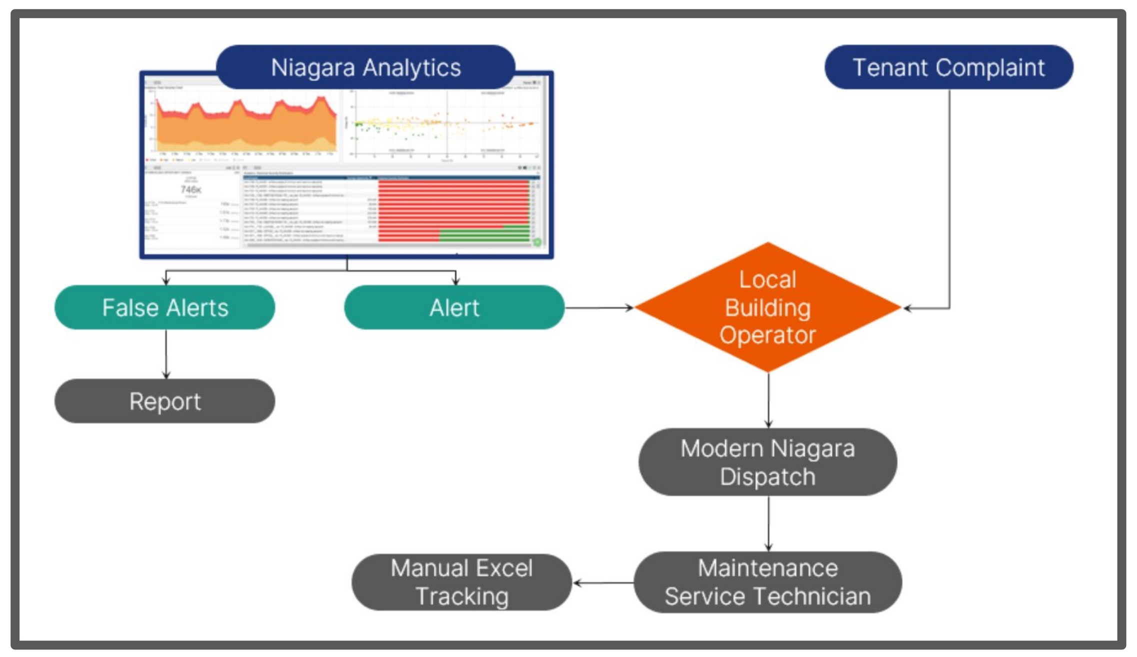
Client: Modern Niagara Group | Supervisors: Professor Seungjae Lee and Professor Markus Bussman

Problem Statement

For maintenance issues such as abrupt faults, Modern Niagara Group (MNG) is reliant on the customer's local building operator to place a call to MNG's 24/7 dispatch for reactive service.

This results in delays in regular building activities which are:

- Costly
- Detrimental to customer satisfaction



Objectives

Create a proof-of-concept that can identify potential building maintenance failures before they occur, without manual intervention.

- **Autonomous:** No human intervention required.
- **Reliable:** Training Accuracy > 90%.
- **Modular:** Can be separated into components.
- **Computationally Efficient:** >16 RAM

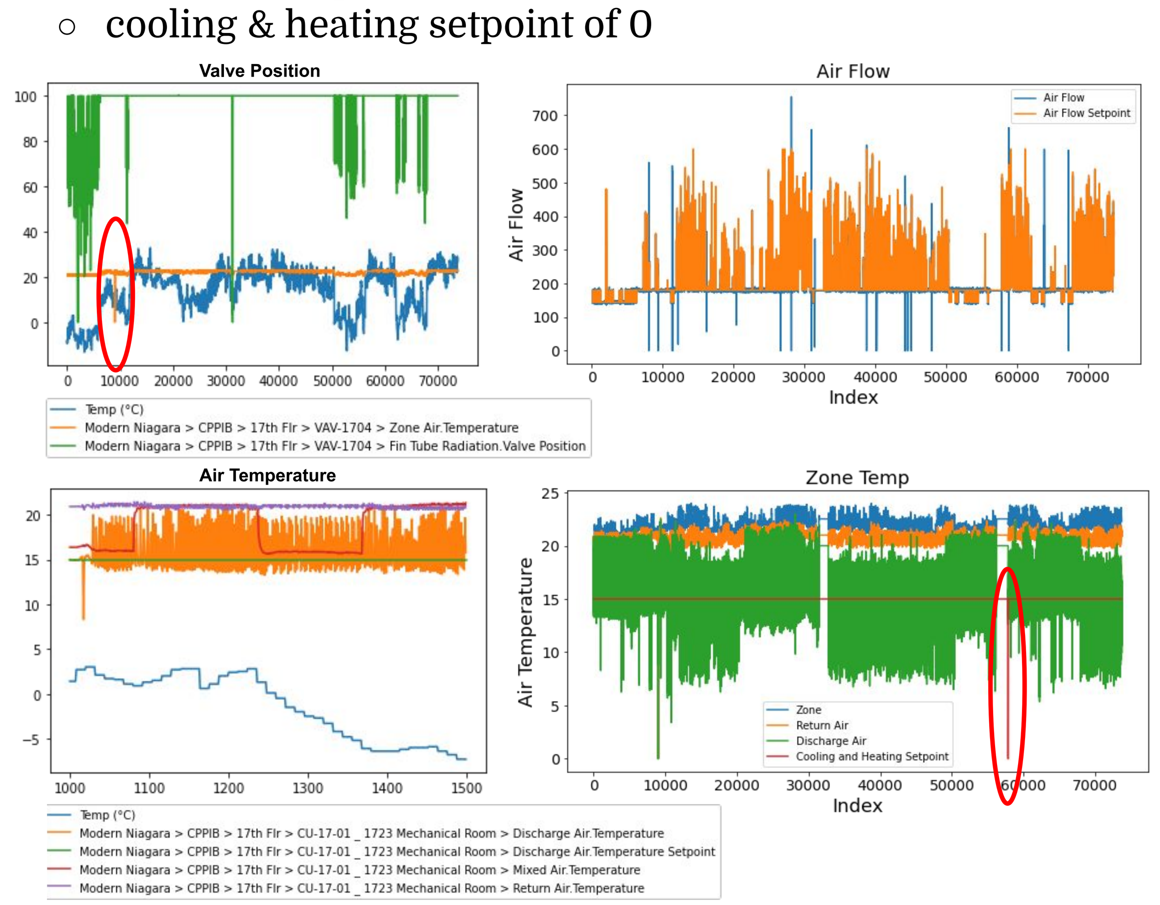
Exploratory Data Analysis

Datasets Used

- HVAC Sensor data from 1 central unit and 42 variable air boxes from the 17th floor of the CPPIB building over a 10-month period
- Weather data for Toronto City Centre from Environment Canada (historical and forecast)
-

Data Cleaning

- identified and removed outliers that are caused by human intervention or sensor error
  - zone-air temperature of 0
  - cooling & heating setpoint of 0



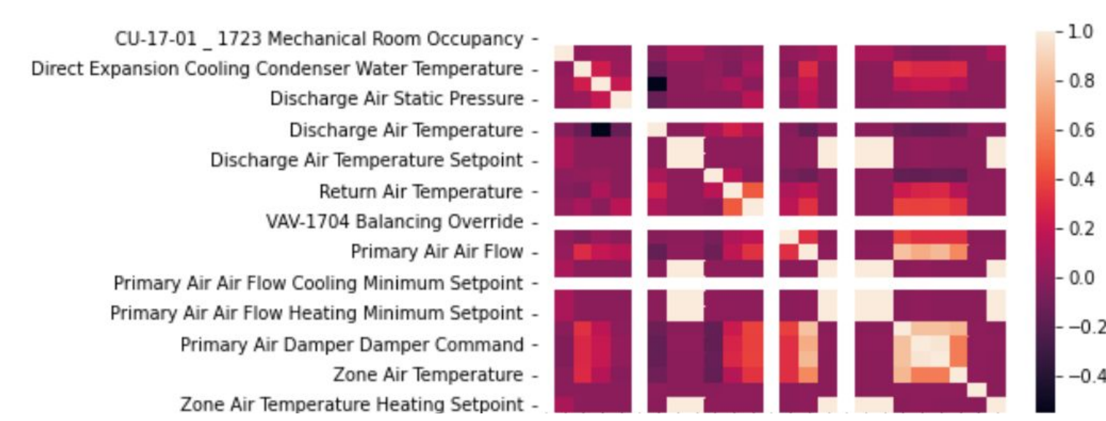
Prediction Selection

Two variables will be predicted: ***Primary Air. Air Flow*** and ***Fin Tube Radiation.Valve Position***. Other variables were eliminated because:

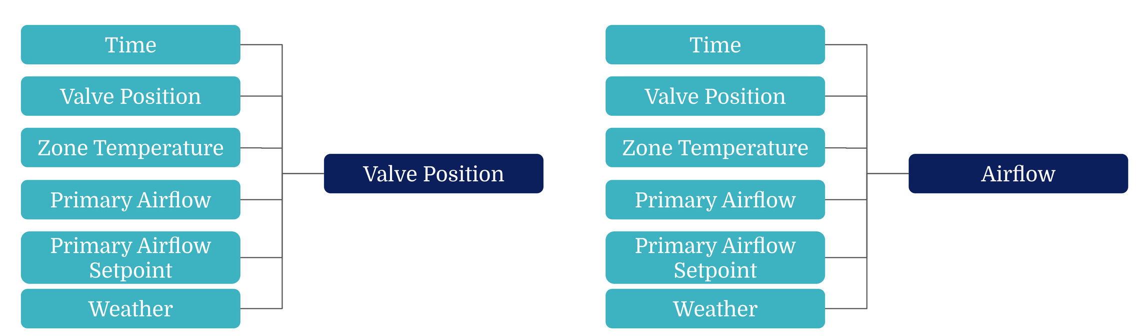
- they were setpoints
- the resolution of the dataset was too large

Feature Selection

Heat map of Spearman Correlation Coefficient:



Considered the correlation coefficient and domain knowledge to determine causation relationships between the variables and develop the inputs and outputs of the valve position model (left) and airflow model (right):

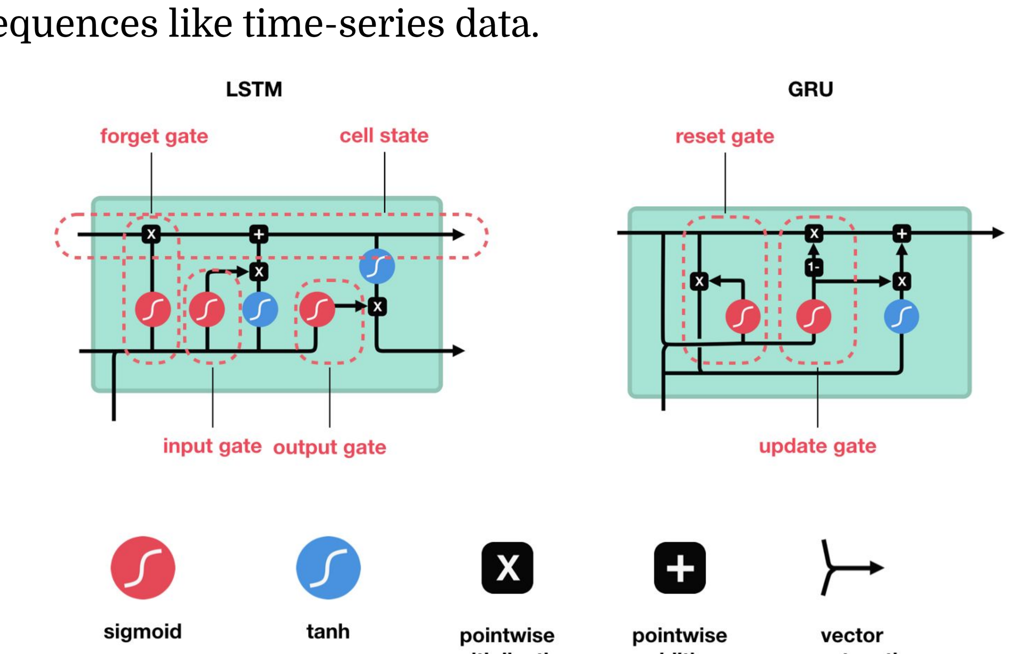


Machine Learning Model Selection

3 models selected:

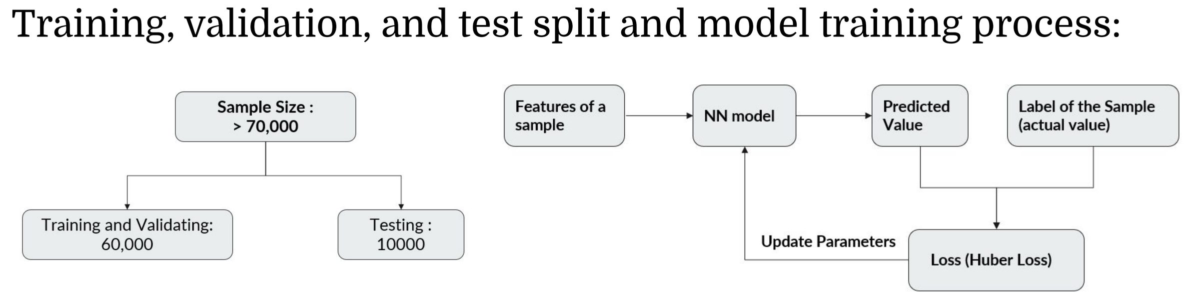
- Baseline: Vector Autoregression (Statistical method)
- LSTM (Recurrent Neural Network)
- GRU (Recurrent Neural Network)

Chain-like nature of RNNs make them suitable for use with sequences like time-series data.



Machine Learning Workflow

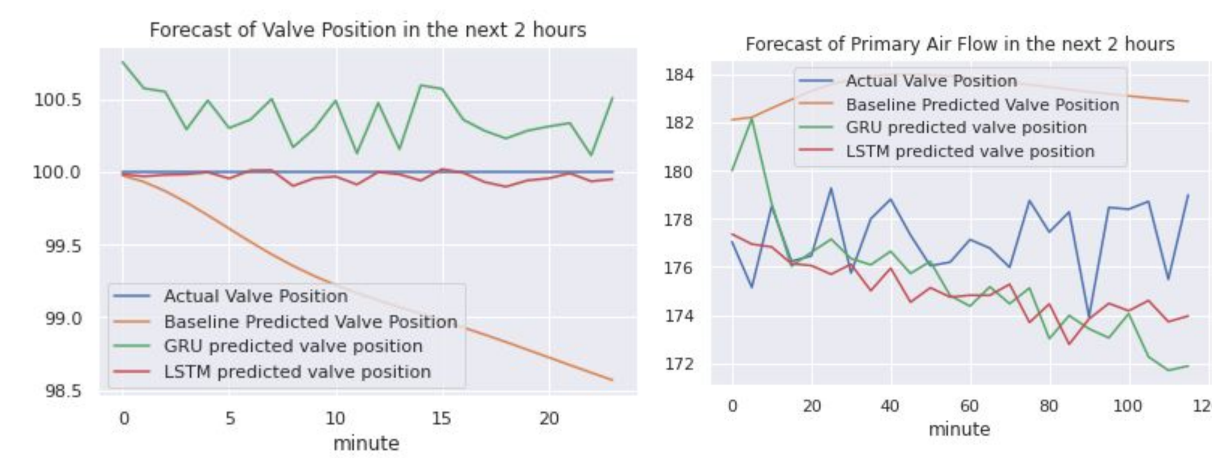
Training, validation, and test split and model training process:



ML Results

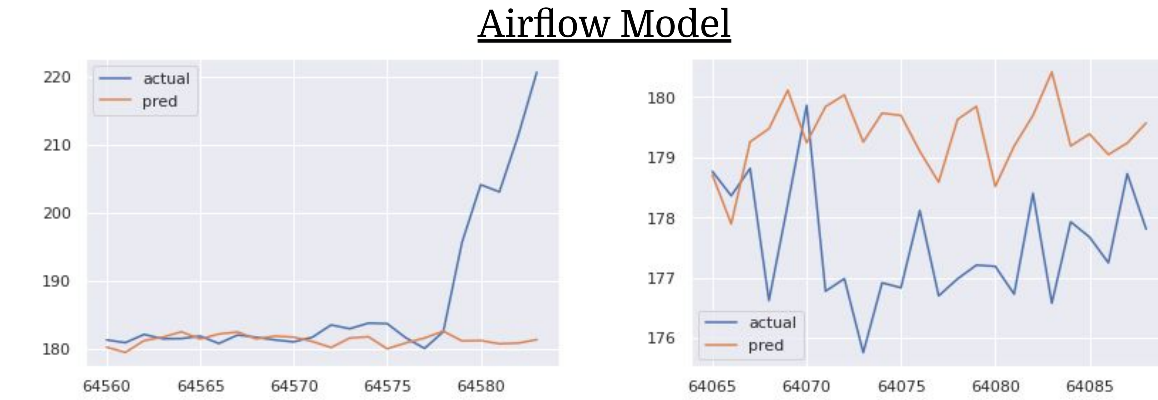
MAE achieved for each model and task:

Model	MAE	
	Airflow	Valve Position
VAR	23.05	1.32
LSTM	17.27	1.14
GRU	17.55	1.31



Error Analysis

Airflow Model



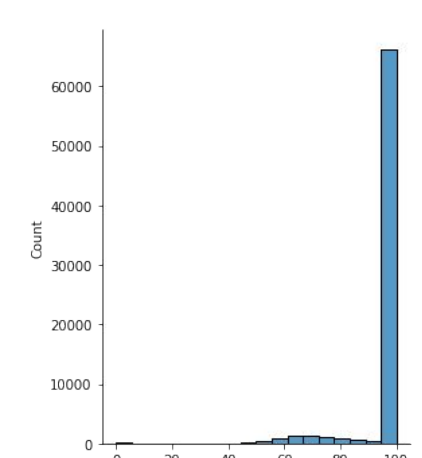
Prediction failing to catch up

Overestimation

Variance

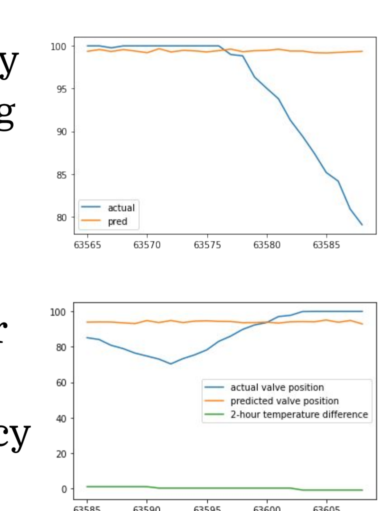
Underestimation

Valve Model



Histogram of Valve Position

- Lower accuracy when deviating from 100
- Future weather data does not improve accuracy



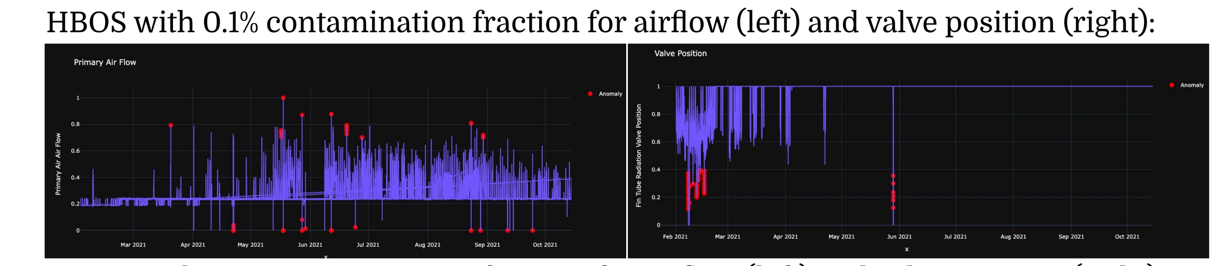
Anomaly Detection

Due to unlabelled data, use two unsupervised anomaly detection algorithms to find global outliers and report to user.

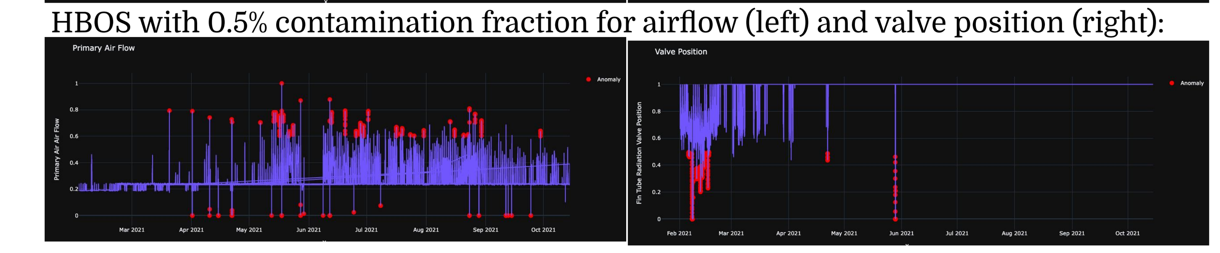
- Histogram-based Outlier Score (HBOS) is a statistical method that models the historical empirical distribution
- Isolation Forest (iForest) is a tree-based methods that isolates outliers based on average distance from the tree root

I. Number of anomalies depends on set contamination fraction

HBOS with 0.1% contamination fraction for airflow (left) and valve position (right):

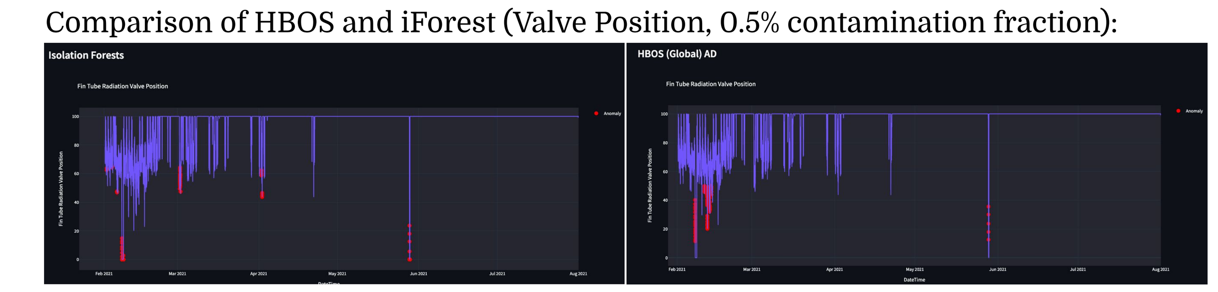


HBOS with 0.5% contamination fraction for airflow (left) and valve position (right):

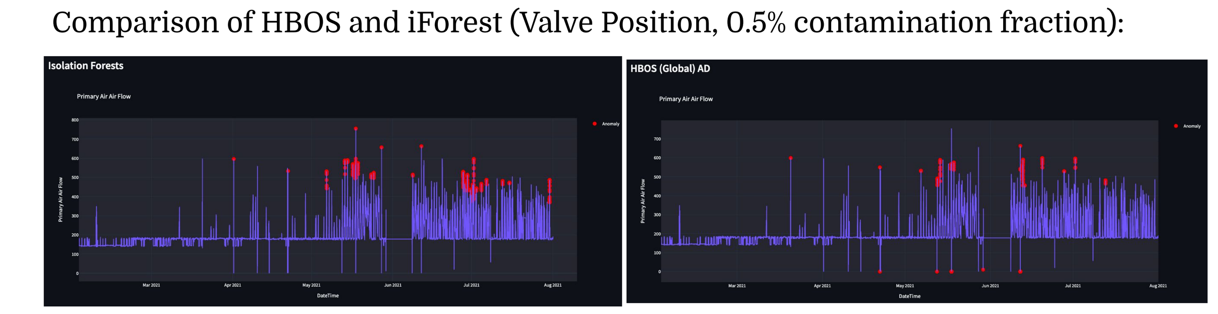


II. HBOS and iForest label different points as anomalous, but without confirmed outliers, cannot make definitive judgement about algorithm performance

Comparison of HBOS and iForest (Valve Position, 0.5% contamination fraction):

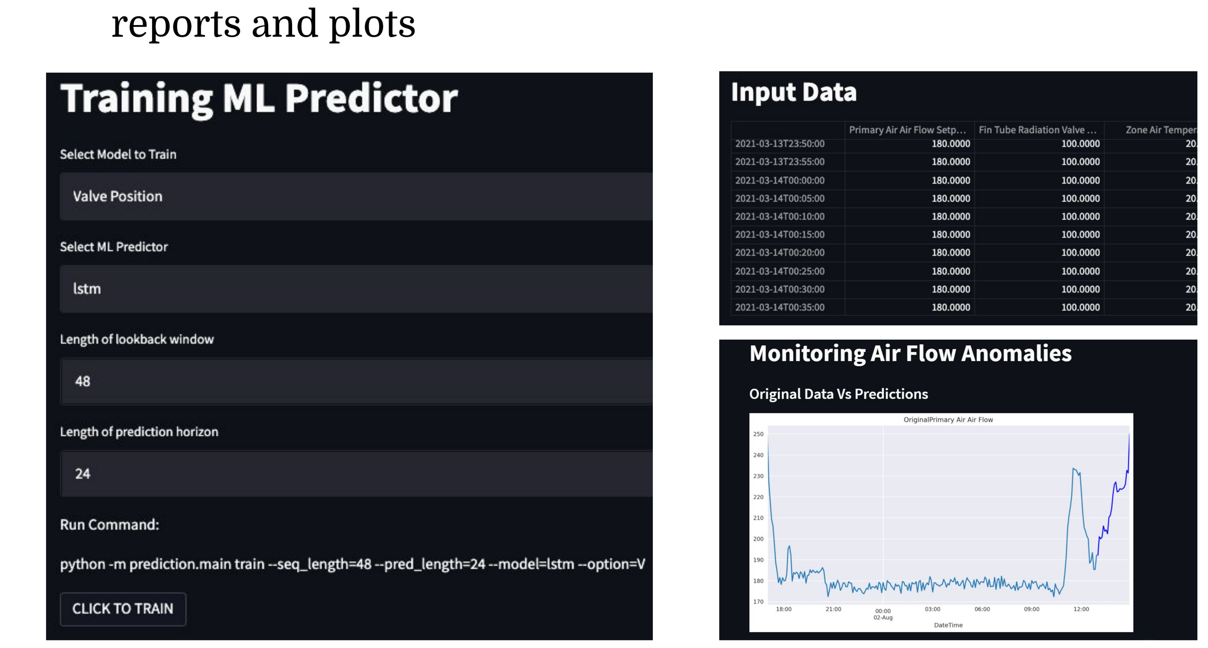


Comparison of HBOS and iForest (Valve Position, 0.5% contamination fraction):



UI Design

- User can select parameters of new GRU or LSTM to train
- User can monitor input data, new predictions and anomaly reports and plots



Business Impact

- Satisfies MNG's goal of automating building service requests and provides centralized information and alerts.
- Reduces costly delays and downtime caused by emergency repairs.
- Makes buildings more comfortable for occupants by preventing the building parameters from entering extreme ranges.

Next Steps

- **Improving the Dataset:** Adding more time-series data and documenting service calls that align with sensor data.
- **Improving the Machine Learning Model:** Further tuning of PoC and modelling for further accuracy and reliability.
- **Expanding the Diversity of Predictions:** Develop more models and identify more inputs & outputs.