Designing a Machine Learning Pipeline for Preventive Maintenance of Building HVAC Systems

APS490 | Multidisciplinary Capstone Project

Final Showcase

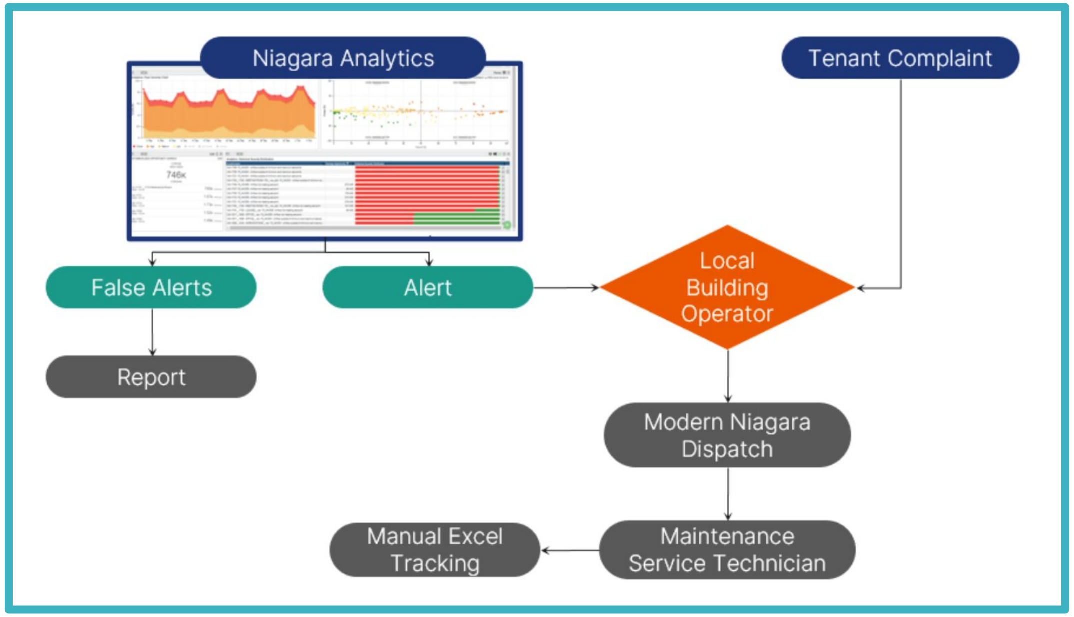
Elizabeth Chelmecki | Anoja Muthucumaru | Chi Zhang | Shirley Zhang | Sherry Zuo

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.

Results in delays in regular building activities which is:

1. Costly
2. 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.

Reliable:
Training Accuracy > 90%.

***Modular*:**
Can be separated into components.

Computationally Efficient:
<16 RAM

Exploratory Data Analysis

MODERN NIAGARA

Sensor data from 1 CU and 42 VAVs from the 17th floor of the CPPIB building over a 10-month period.

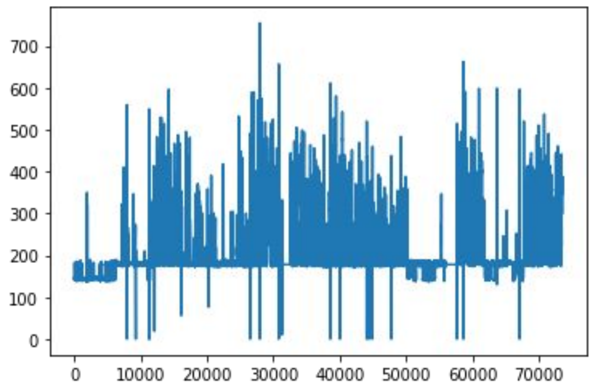
Environment Canada

Hourly temperature data at Toronto City Centre.

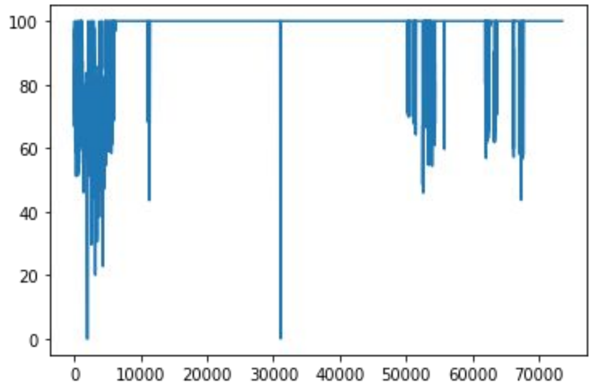
Exploratory Data Analysis - HVAC data

Univariate Visualization

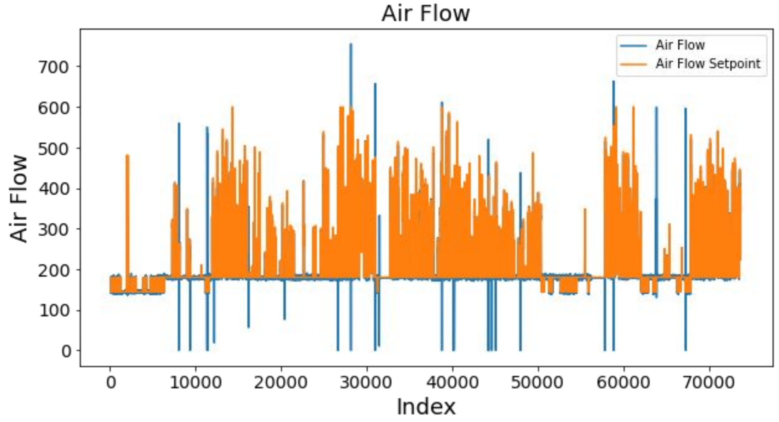
air flow

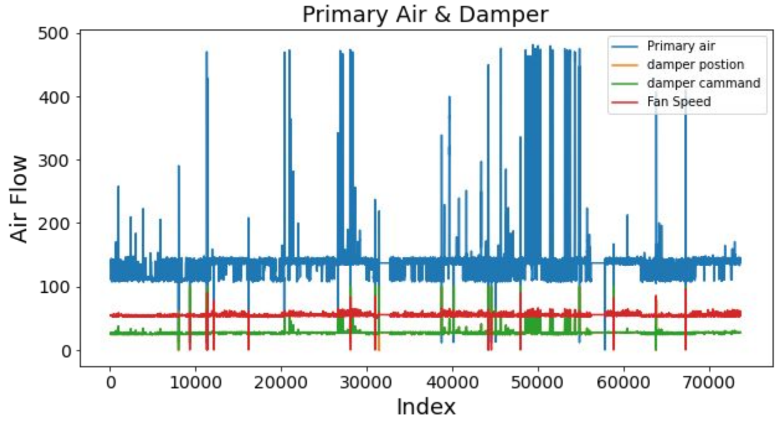


damper

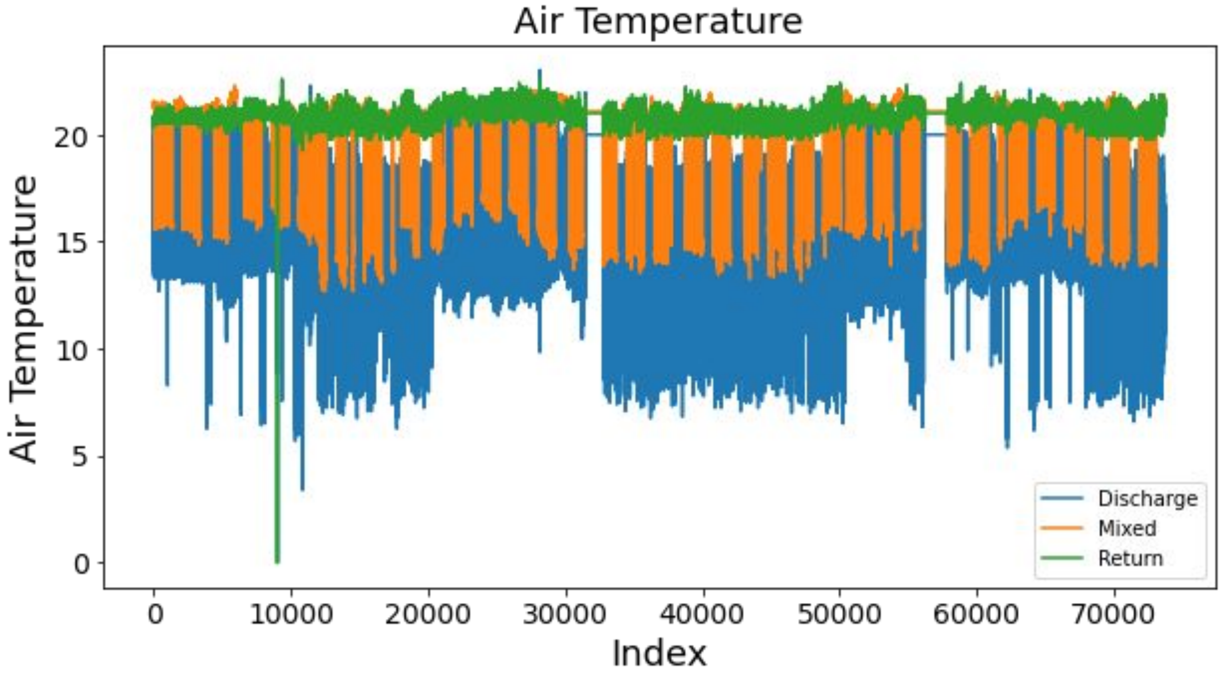


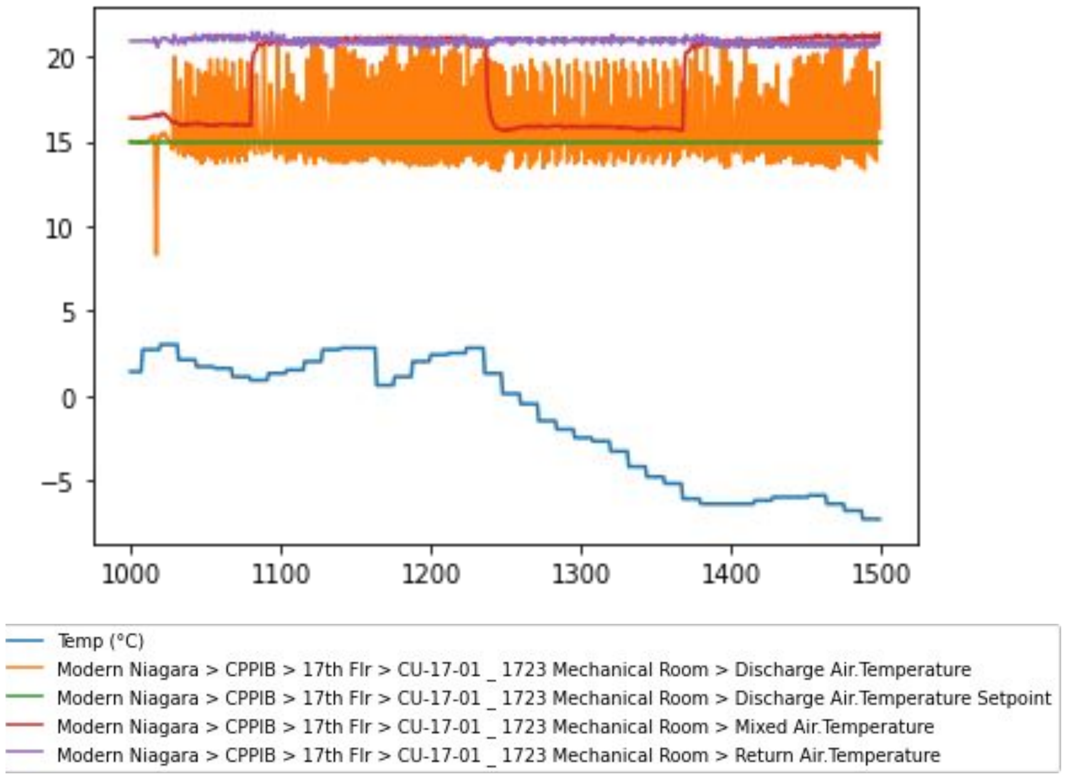
Multivariate Visualization



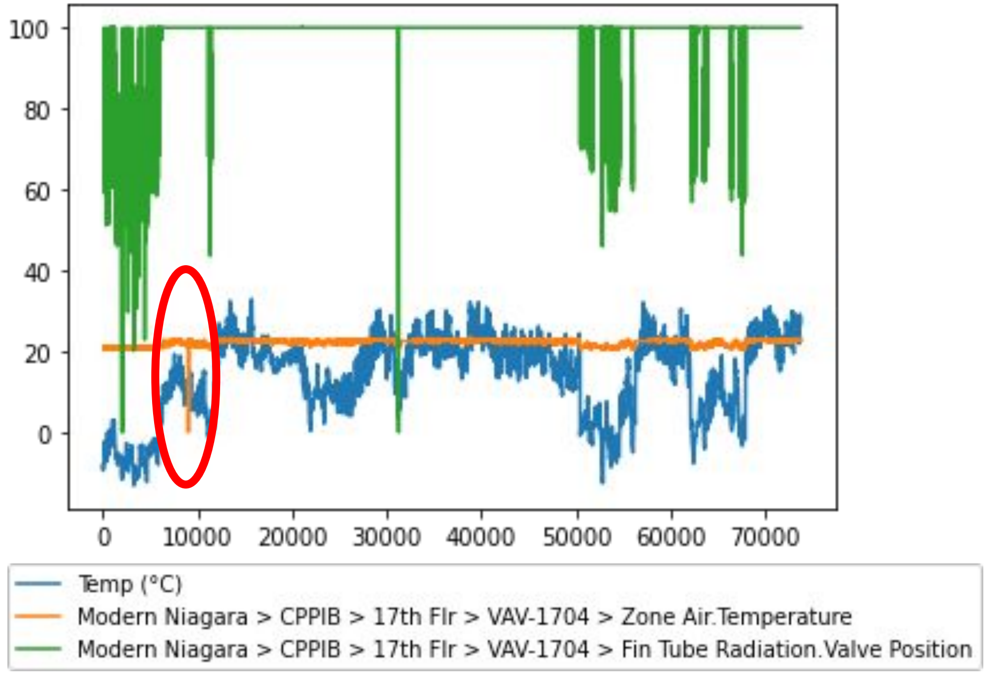


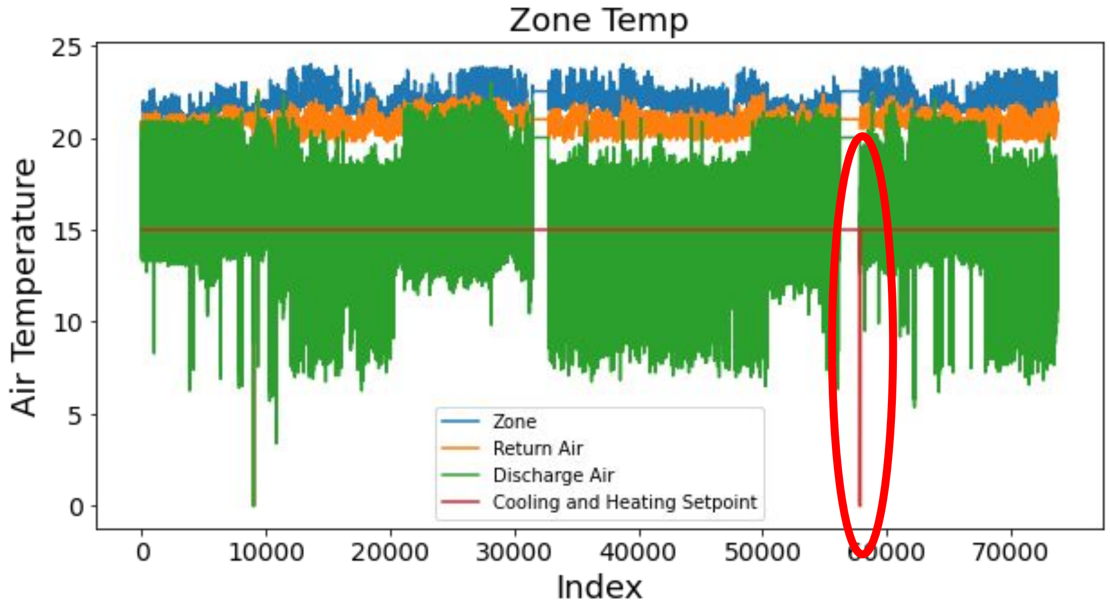
Exploratory Data Analysis - weather data





Data Cleaning

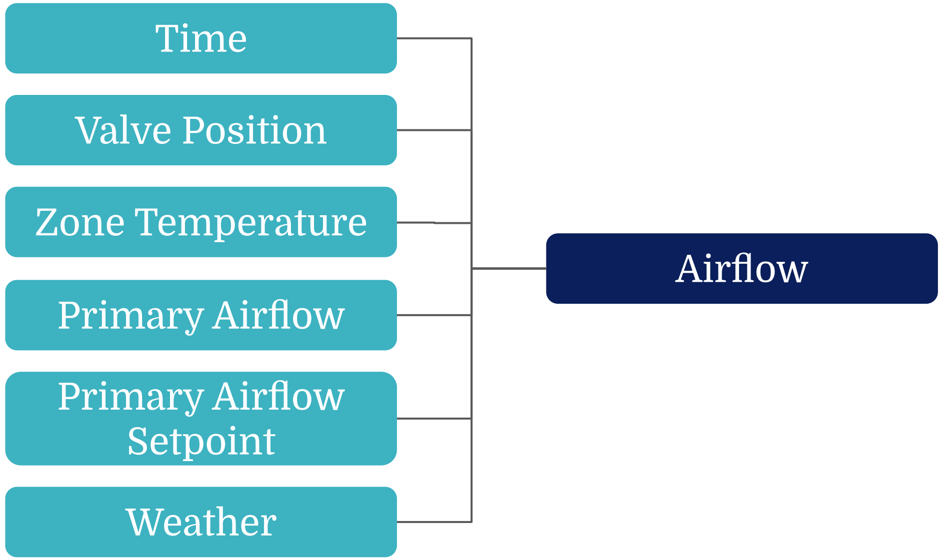




Feature Selection: Airflow Model

	Airflow	Valve Position	Airflow Setpoint	Zone Air Temperature	Outdoor Temperature
Airflow	1.000	0.3260	0.8335	0.6225	0.6190

Spearman Correlation Coefficients

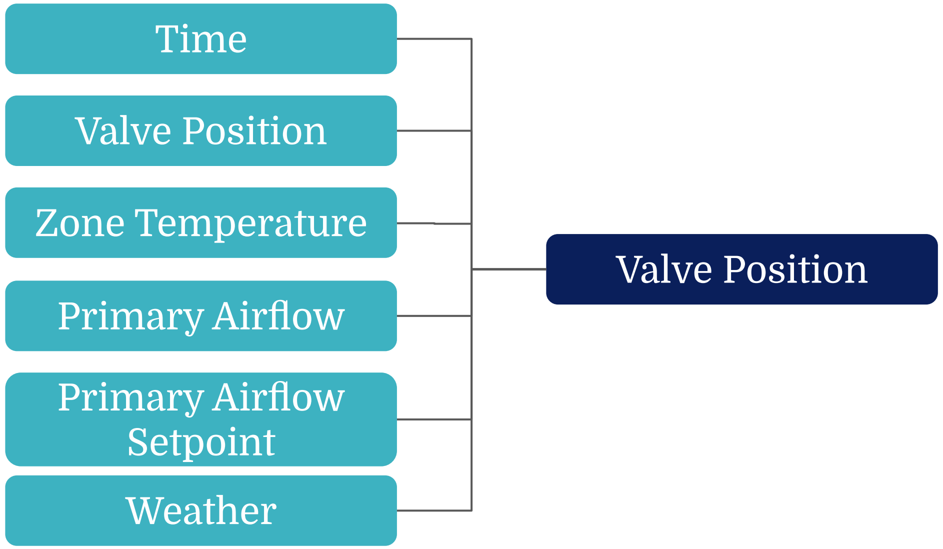


Inputs and Outputs of Airflow Model

Feature Selection: Valve Position Model

	Valve Position	Airflow	Airflow Setpoint	Zone Air Temperature	Outdoor Temperature
Valve Position	1.00	0.277	0.278	0.553	0.597

Spearman Correlation Coefficients



Inputs and Outputs of Valve Model

Model Selection

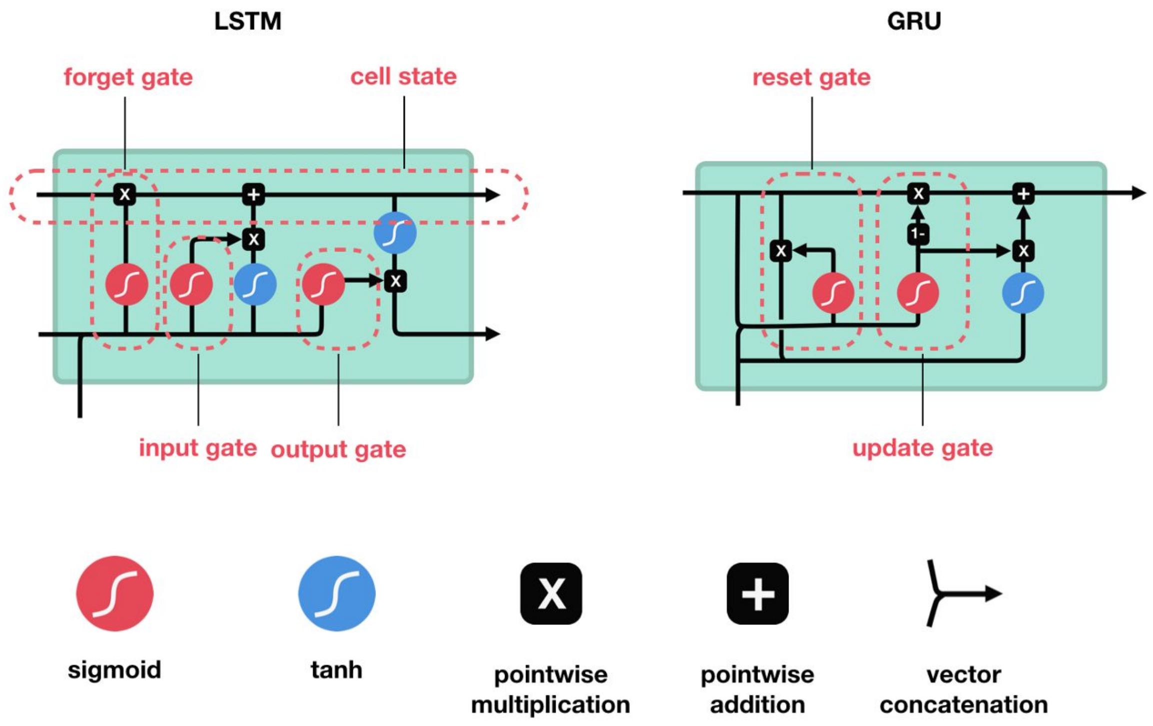
Baseline

Vector Autoregression (VAR)

$$
\begin{aligned}
Y_t &= \alpha + \beta_{11} Y_{t-1} + \beta_{12} Y_{t-2} \\
&\quad + \gamma_{11} X_{t-1} + \gamma_{12} X_{t-2} + \varepsilon_{1t} \\
X_t &= \alpha_2 + \beta_{21} Y_{t-1} + \beta_{22} Y_{t-2} \\
&\quad + \gamma_{21} X_{t-1} + \gamma_{22} X_{t-2} + \varepsilon_{2t}
\end{aligned}
$$

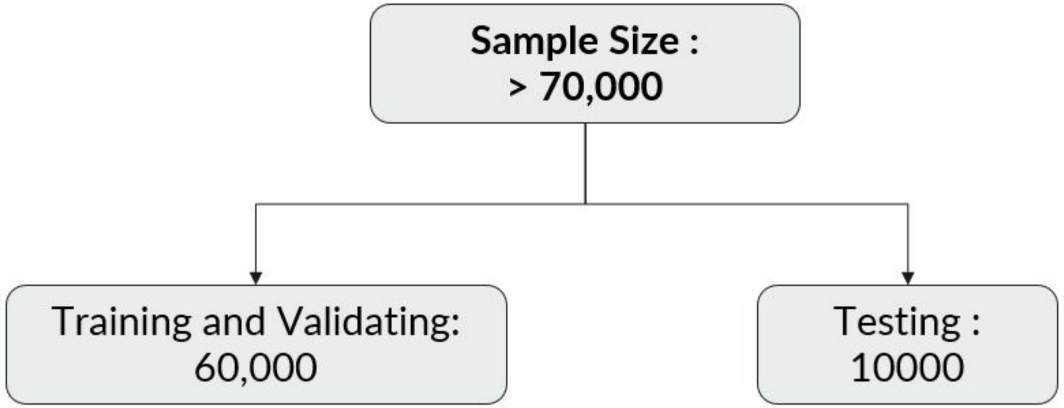
Recurrent Neural Networks (RNNs)

LSTM and GRU

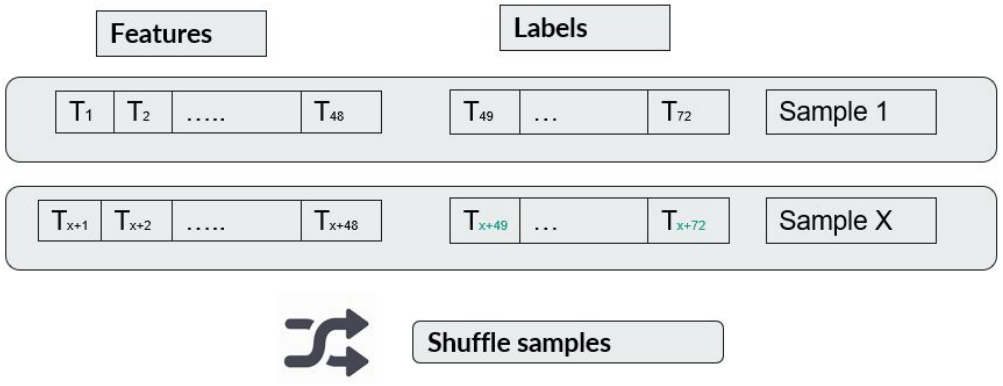


Machine Learning Data Preparation

Training/Validation/Test Split

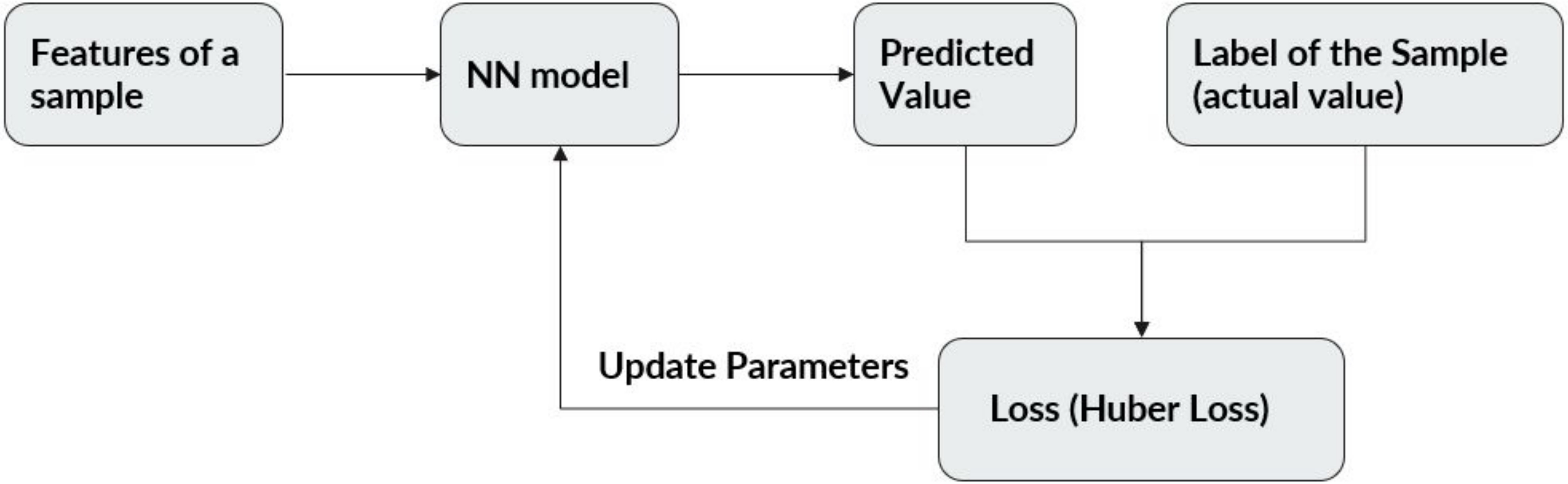


Organization of Data Samples



Machine Learning Training

RNN Model Training Process

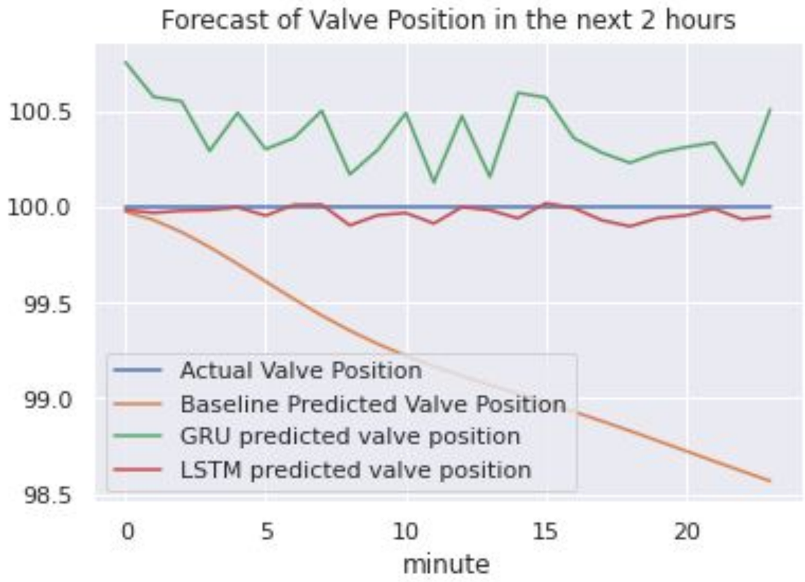


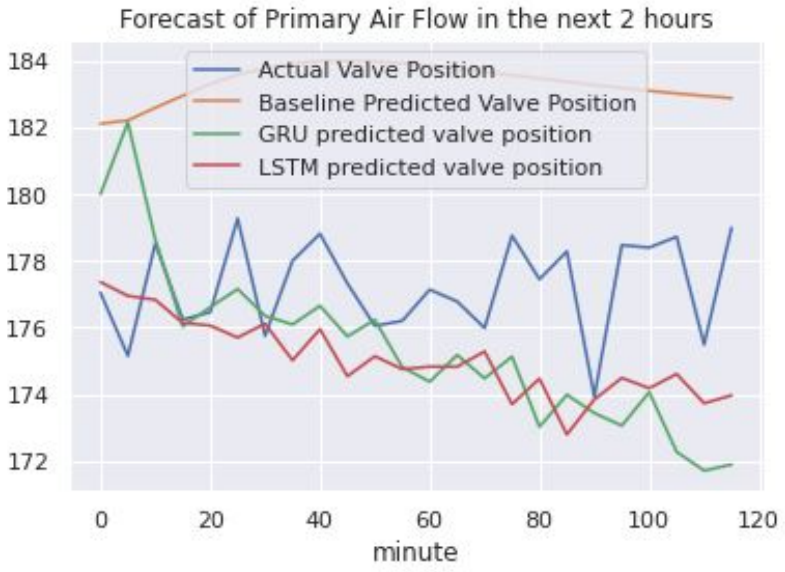
Machine Learning Prediction Results

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

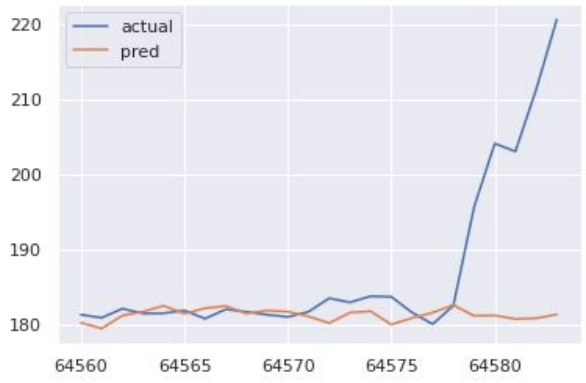
Machine Learning Prediction Results

Sample

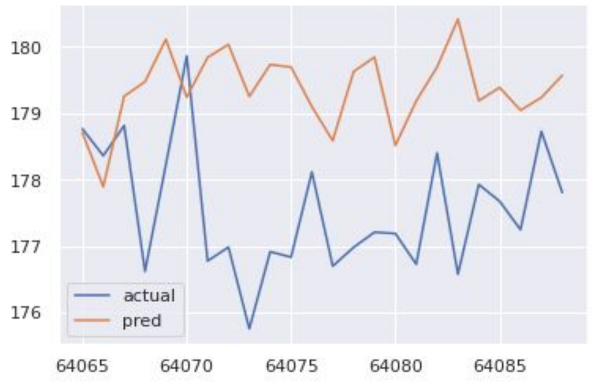




Error Analysis - Airflow Model

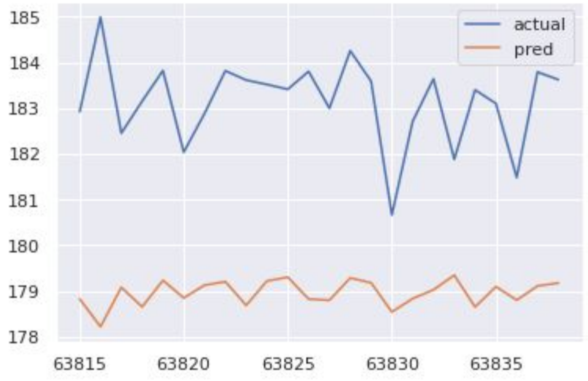


Prediction Failing to Catch Up



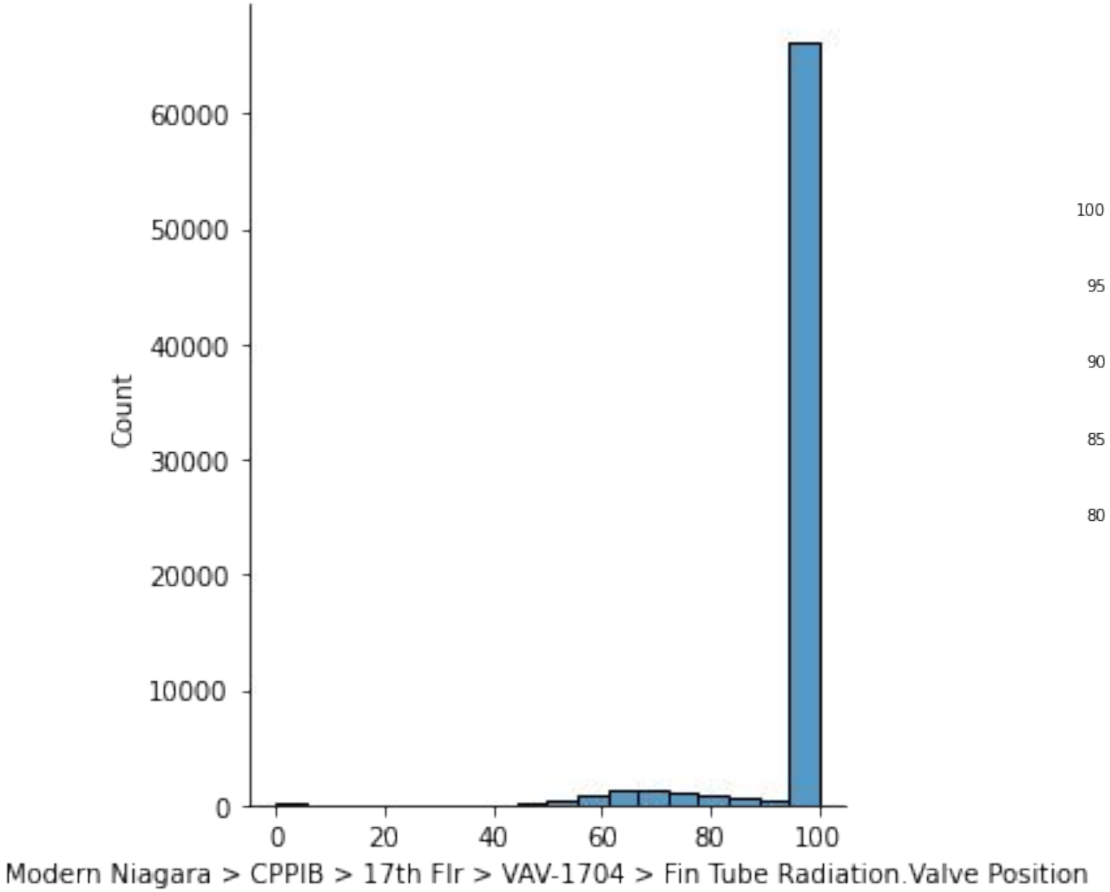
Overestimation

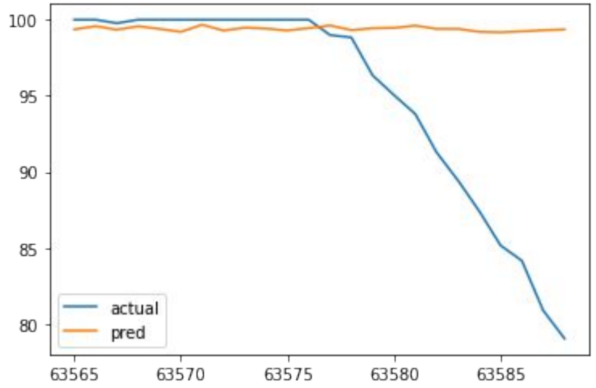
Variance

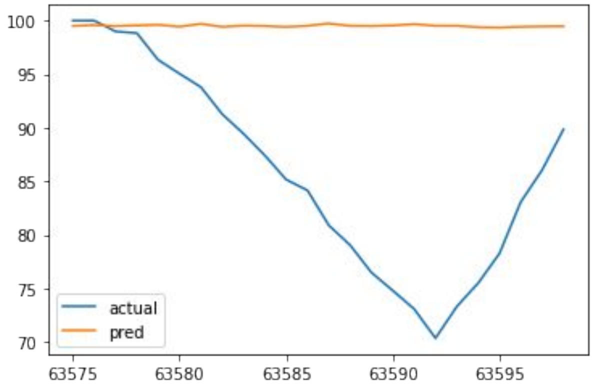


Underestimation

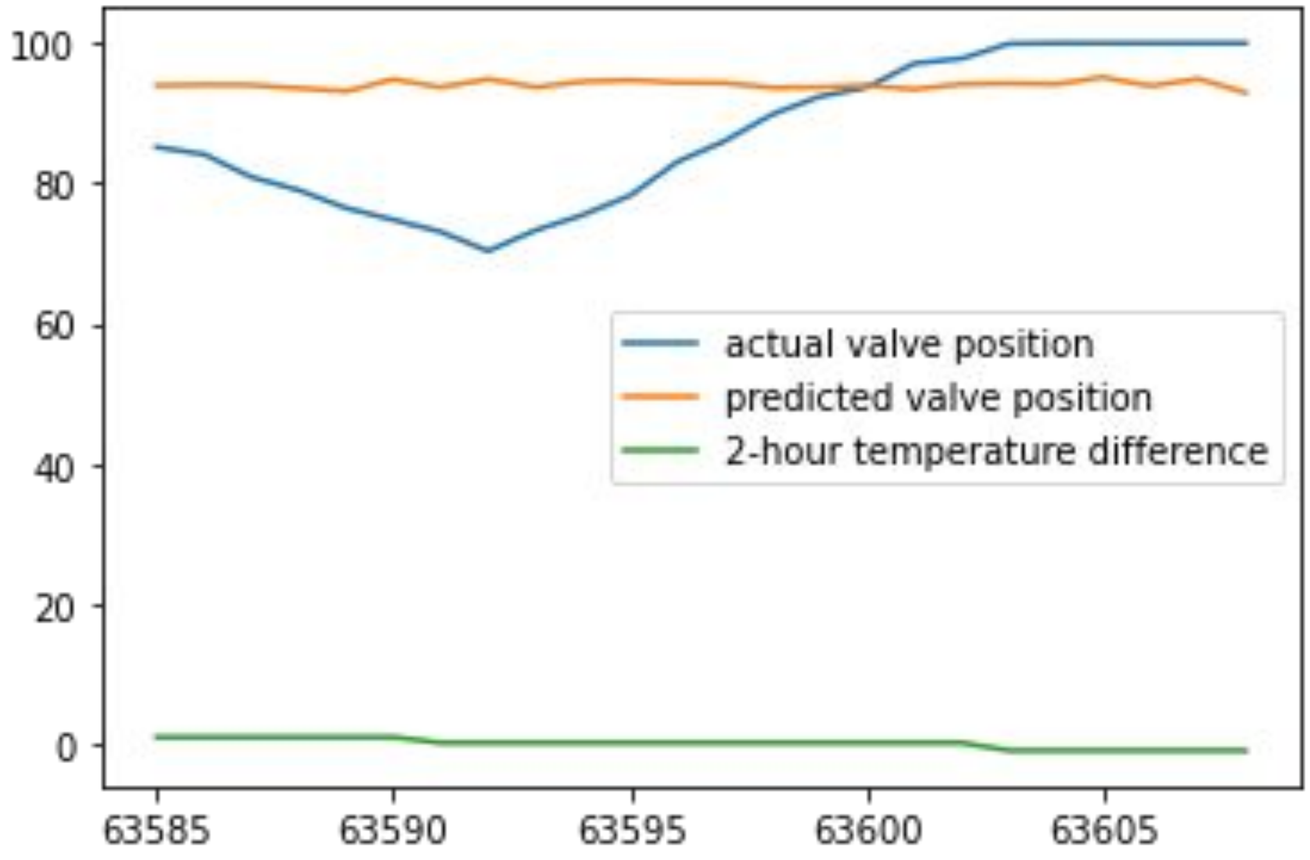
Error Analysis - Valve Model







Error Analysis - Valve Model




```
df_input
```

	Temp (°C)	Temp Diff
0	-8.2	0.9
1	-8.2	0.9
2	-8.0	1.8
3	-8.0	1.8
4	-8.0	1.8
...	...	...
73510	26.7	-1.3
73511	26.7	-1.3
73512	26.7	-1.3
73513	26.7	-1.3
73514	26.7	-1.3

Anomaly Detection Overview

Generate Predictions | Preprocess | Display Anomaly Report

Use selected ML model to predict for next time horizon and append prediction to the timeseries

- Normalize
- Add date-related features
- Add weather feature

Anomaly detection is:

- **Unsupervised**
- **Global**

Show anomalous points and corresponding information

Histogram Based Outlier Score (HBOS)

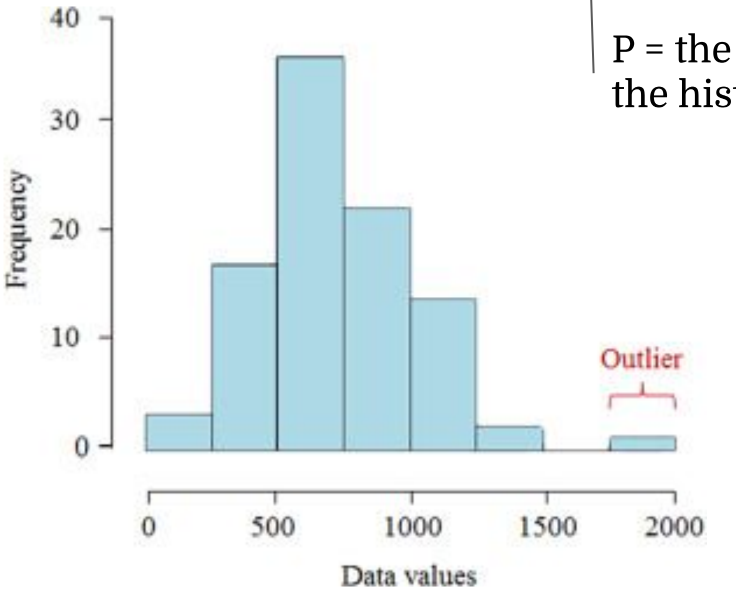
Simple statistical method based on empirical densities, but not all assumptions hold in real-life

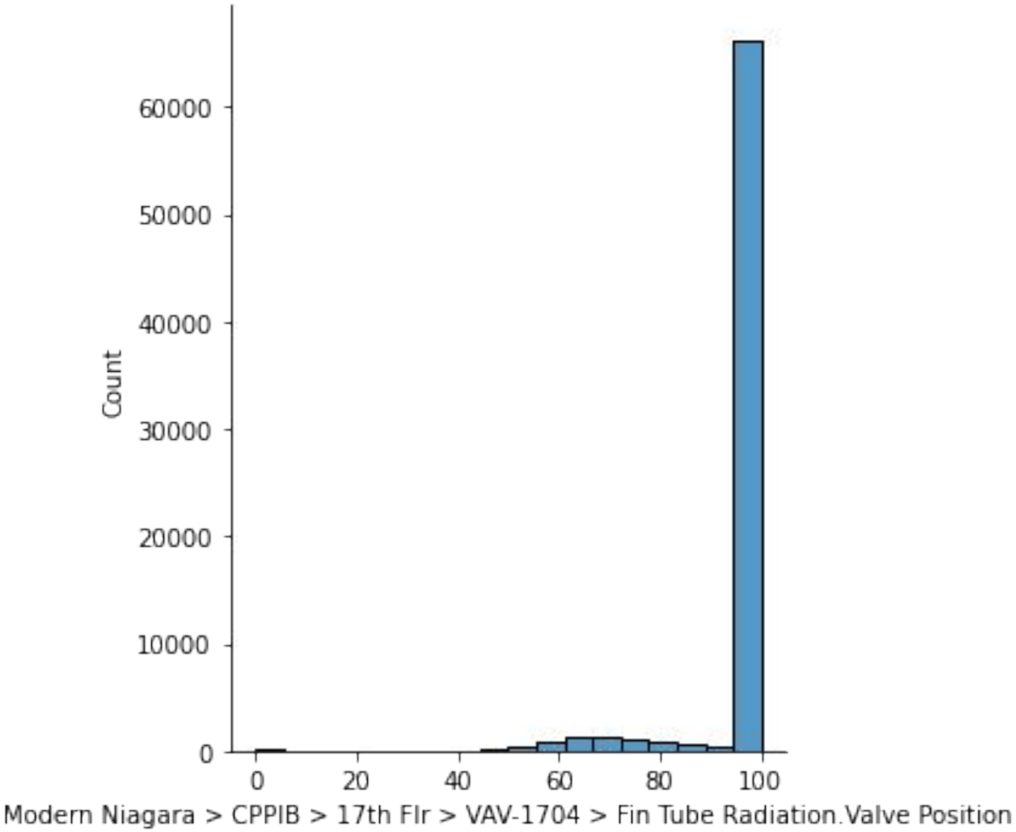
- Sensitive to scaling methods
- Only works for continuous features

$$HBOS(p) = \sum_{i=0}^{d} log(\frac{1}{hist_i(p)})$$

Larger score = more anomalous

P = the height of the histogram

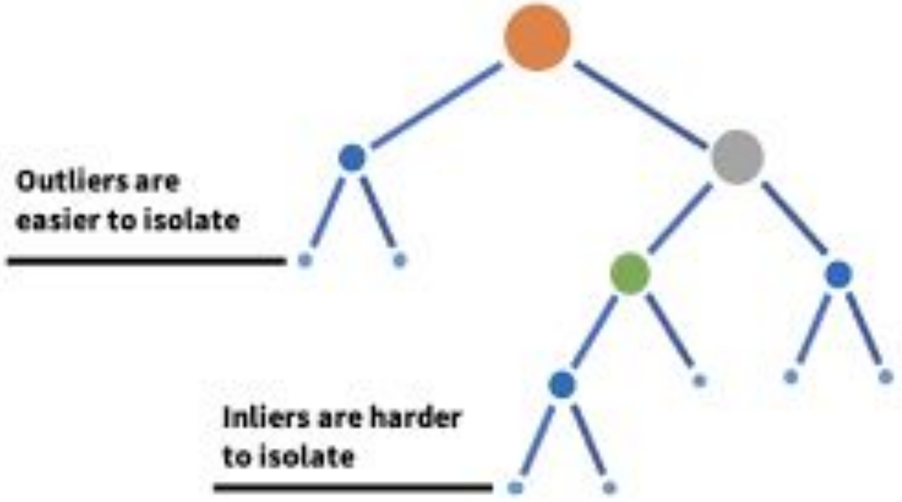


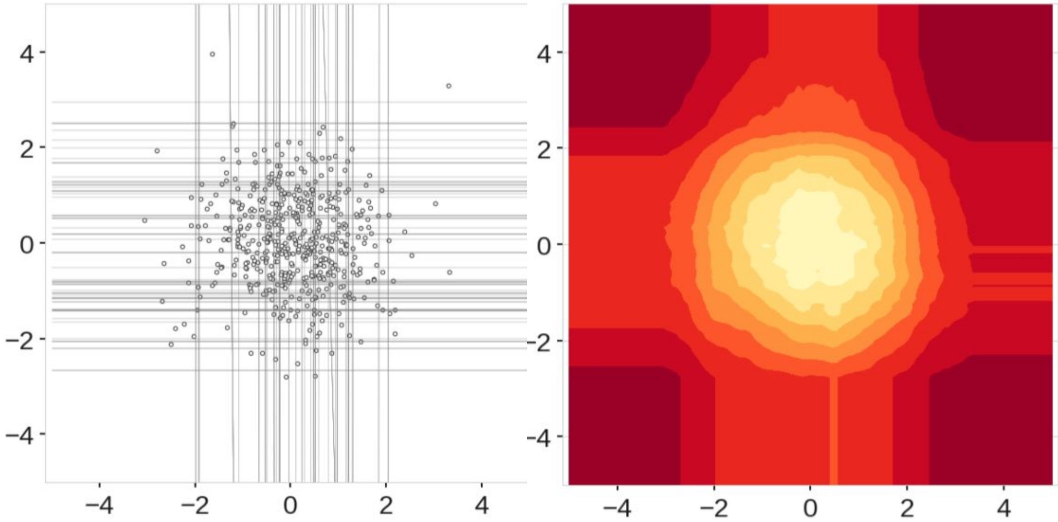


Isolation Forest (iForest)

More sophisticated, tree-based method that isolates outliers rather than models normal behaviour.

- Assumes outliers are **few** and **very different**
- Sub-sampling efficiency
- Sensitive to feature errors, hyperparameters
- Less visualizable and explainable

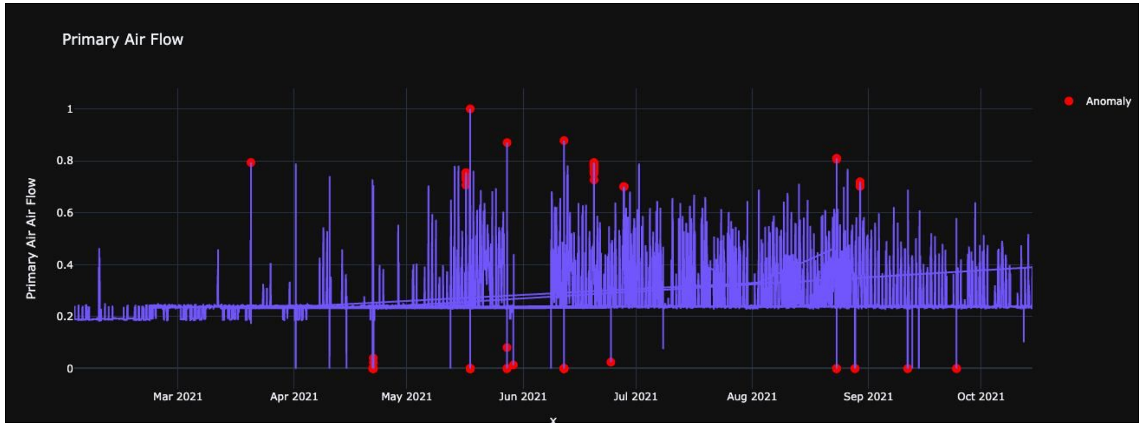


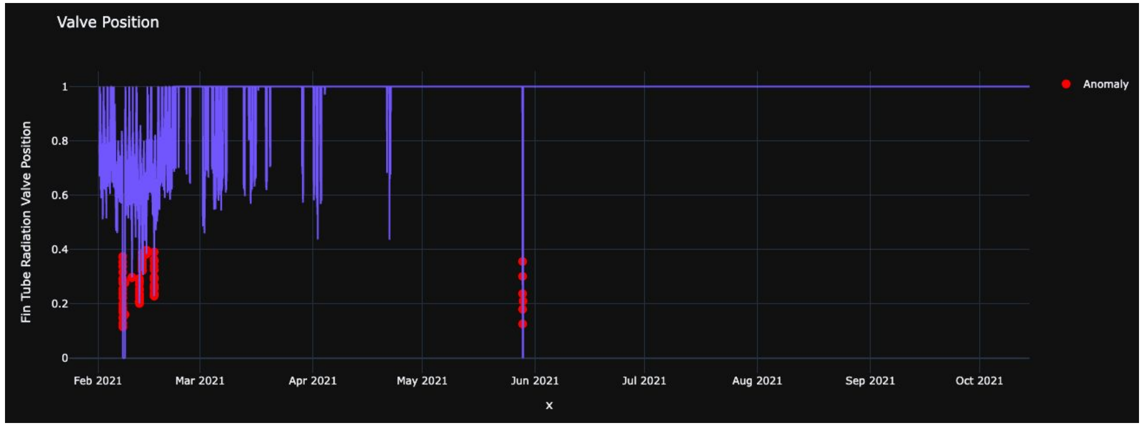


Setting the Contamination Fraction

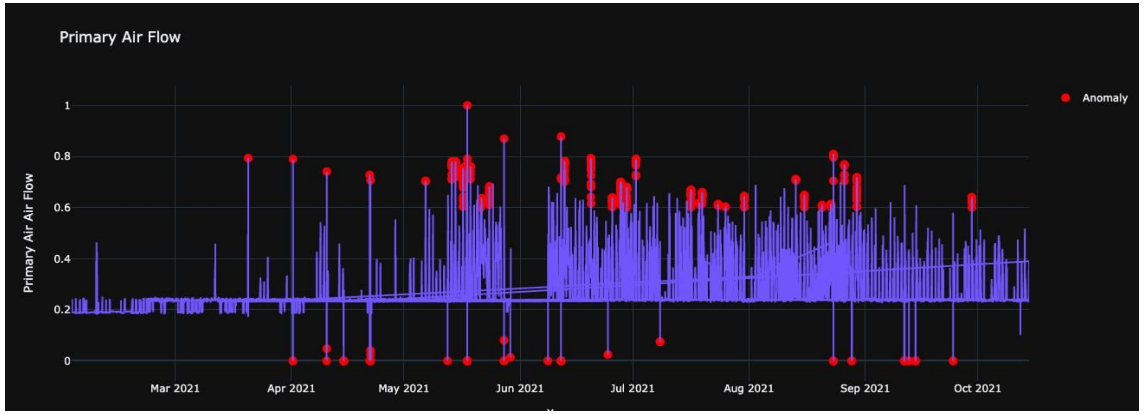
Histogram-Based Outlier Score (HBOS): Varying the contamination fraction

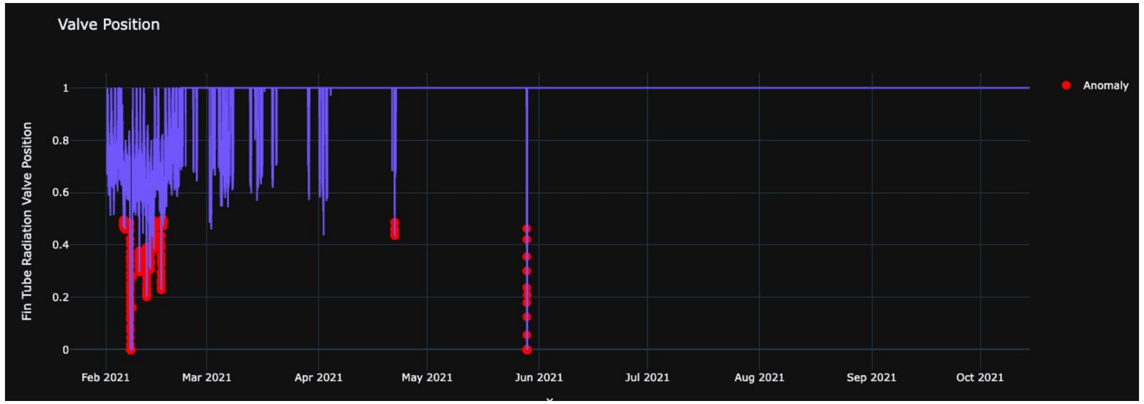
0.1%





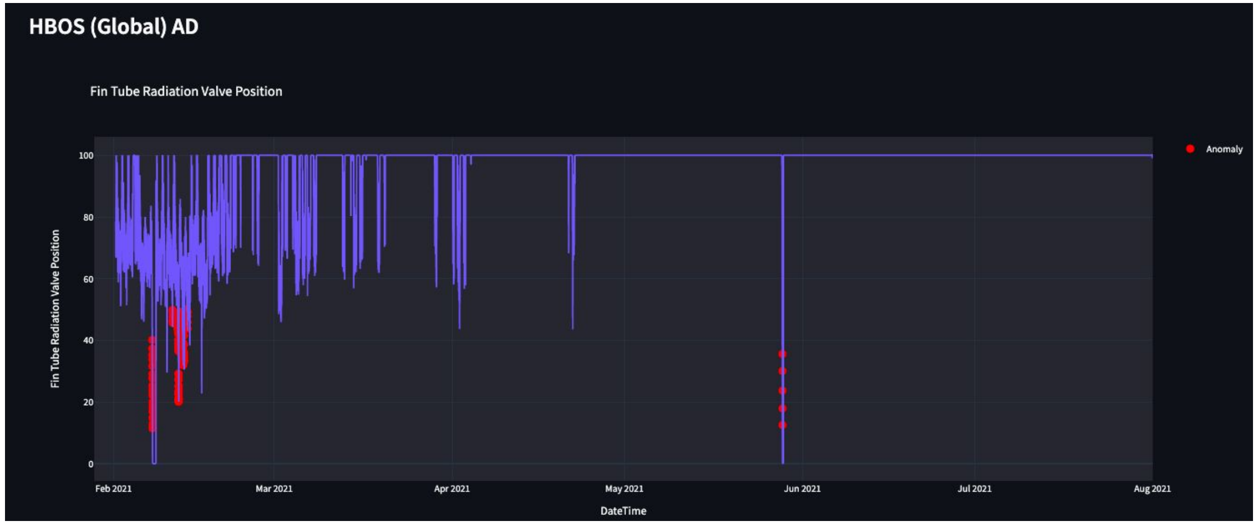
0.5%





Valve Position - Comparison of Methods

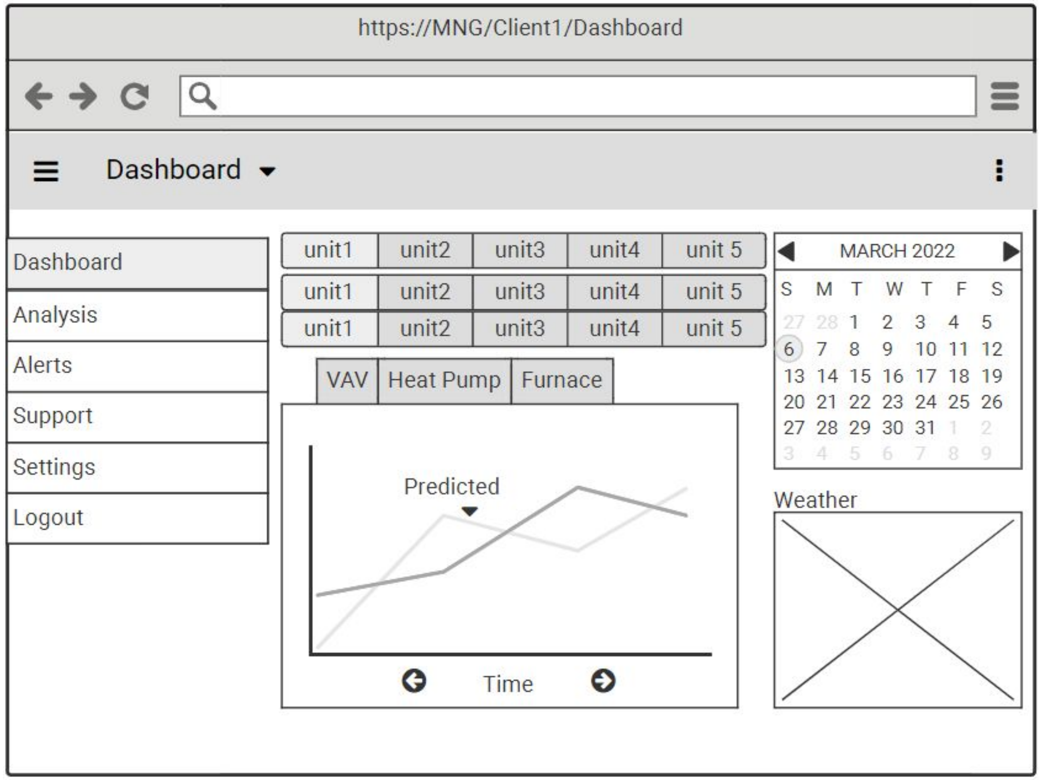
Valve Position at Fraction = 0.5%



Airflow Model - Comparison of Methods

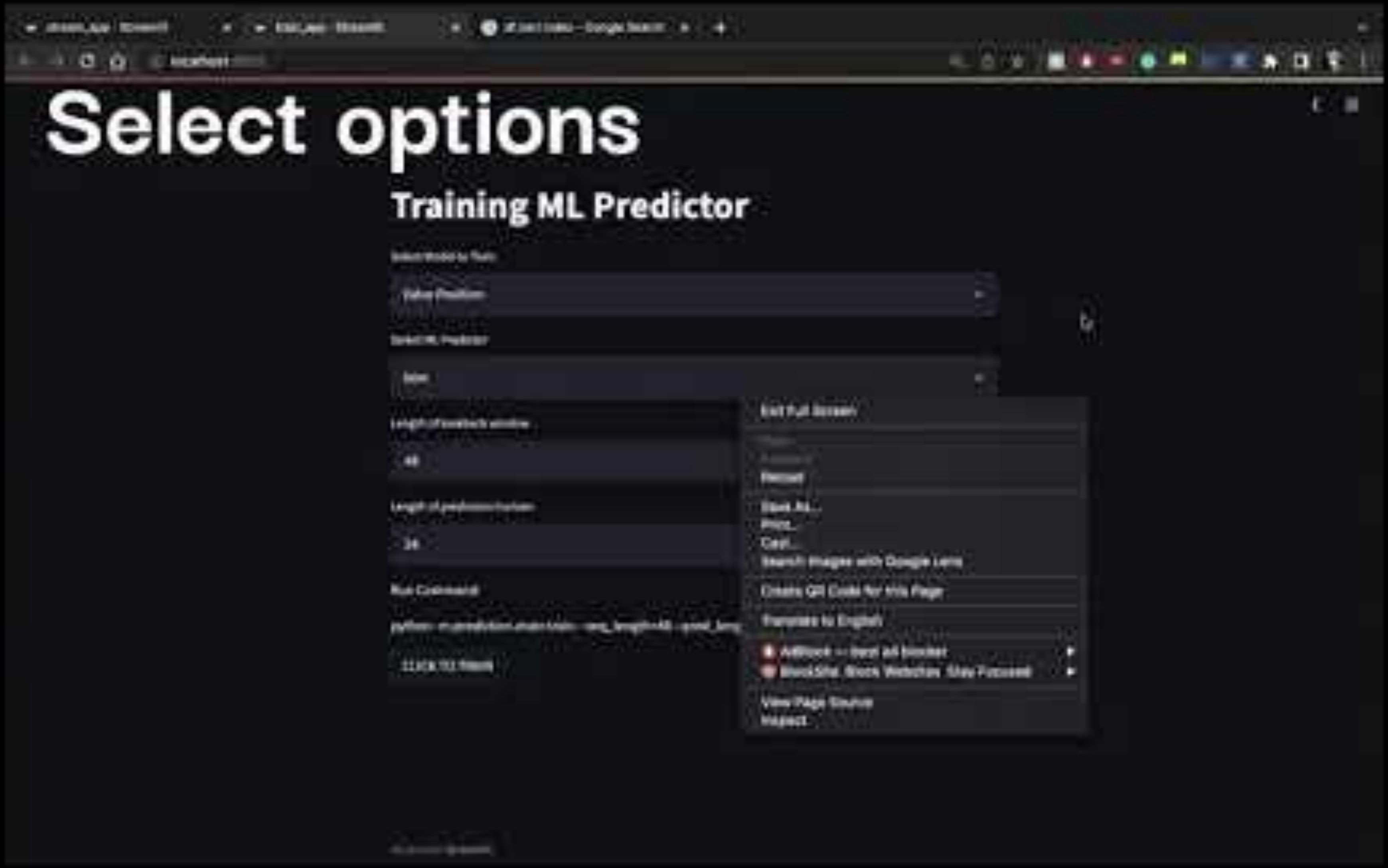
Airflow at Fraction = 0.5%

UI Design



Input Window	Showing the static data
Filters	Ability select/navigate to a date and time
Output window	Ability to download predictions
Interactive Plots	The ability to navigate the plots that have been generated.

Training Demo



Anomaly Detection Dashboard

Conclusion & 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 Prediction: Develop more models and identify more inputs & outputs

Thank you!!

Questions?

Feel free to contact chizhang0826@gmail.com