Wei Ruan

Final Report:
Automated Fault Detection and Diagnosis (AFDD) of HVAC Systems

1. Problem statement

The energy consumption in the building sector accounts for approximately 20-40% of the total energy consumption, and heating, ventilation, and air conditioning (HVAC) systems consume approximately 50-60% of the total energy consumed by the buildings. This project will develop a data-driven AFDD scheme for air handling units (AHUs) with the precision of over 80% to detect common faults of HVAC, which will detect faults early that could be repaired to save energy, improve the reliability and thermal comfort of buildings.

The dataset, generated by PNNL, includes the operation data of three types of AHUs for a large office building under the normal conditions and fault conditions. After a variety of exploratory data analysis (EDA) methods, a few machine learning models are compared and KNN model is selected to be an optimized ML model to detect one of faults.

After reducing the dataset from 17 to 12 features, the random forest model was able to achieve an average accuracy of 0.988 for the first fault - outdoor air temp offset 2 degree higher. Similar models could be implemented for the faults to enhance monitoring the operations of three types of AHUs.

2. Data Wrangling

The raw dataset of the first fault from PNNL contained 272,1612 rows with 18 columns. The experiments were conducted to mimic the fault scenarios and these data were collected and simulated by PNNL. Although they were cleaned and organized, they needed to check the data types and drop some missing data. For every dataset, there was the date type of data. For this AFDD problem, we don’t need the date and time of the data, so they were dropped from the raw data. There were also some blanks in the column names, which were cleaned in the dataset.

The histograms of features are dependent variables were plotted for further investigation. The last sub-plot showed that there were 5 times more fault scenarios (230,000 rows) than the normal scenarios in the whole dataset. These provide sufficient train data to establish a robust fault detection model and sufficient test data to validate the accuracy of that model. For other data, we also have more than 3000 rows of data at the fault scenarios, which provide sufficient train and test data for the future investigation.

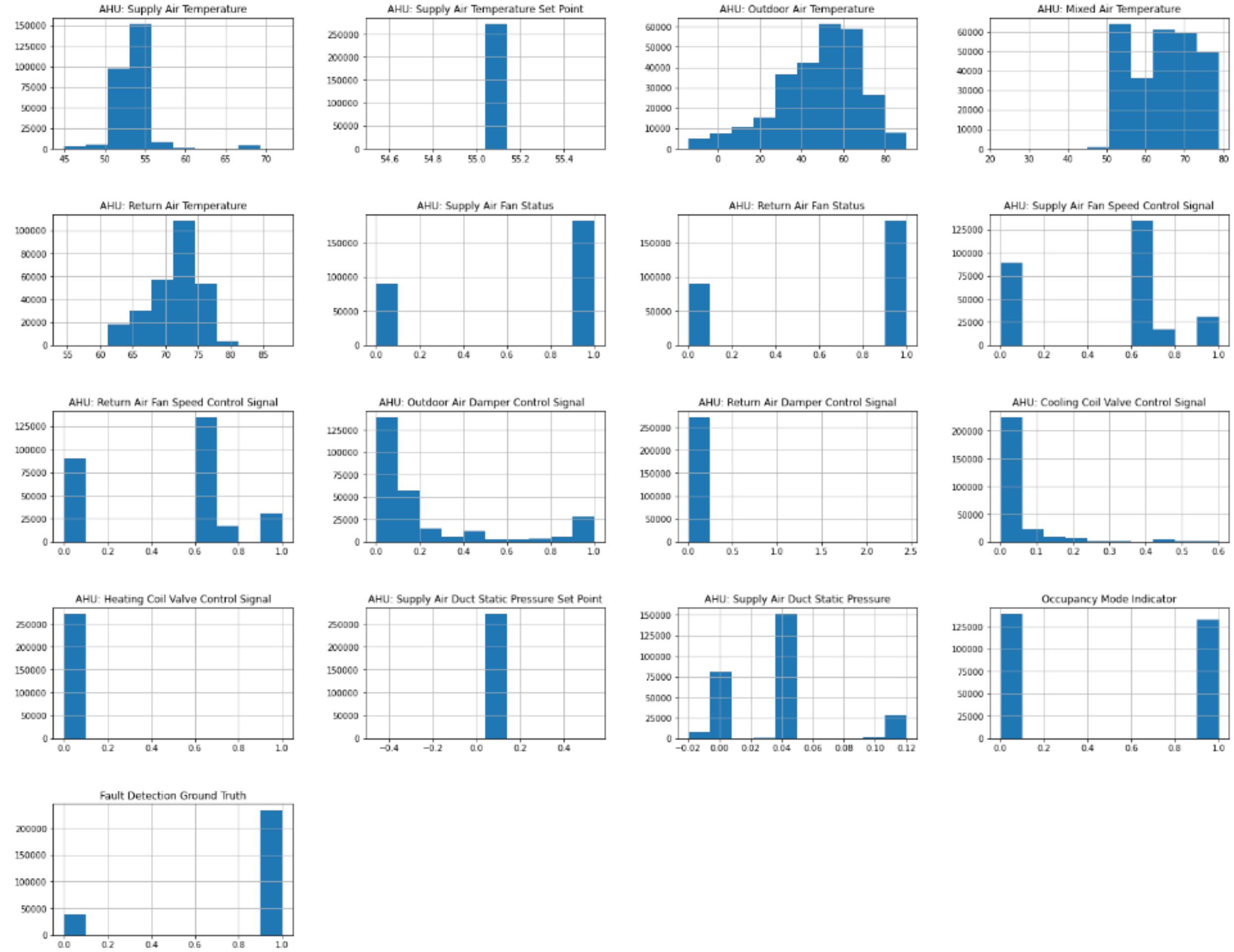


Figure 1: Subplots of the features and dependent variable

Exploratory Data Analysis

In the exploratory data analysis, we found there are six types of faults that were the outdoor air temperature (OAT) sensors off set +/- 1, 2, and 3 degree Fahrenheit than the true values. We picked up the scenarios that the OAT sensors off set 2 degree Fahrenheit than the true values as the fault scenarios for this project. From Figure 2, we could see that there were about 38,000 rows of observation for both the normal and the fault scenarios that we investigated in this project.

The histograms of outdoor air damper control signal (0-1) were also plotted for the normal conditions and the fault conditions in Figure 3a and 3b. The data collecting days were very close and the weather was similar for the normal conditions and the fault conditions. The difference of the outdoor air damper control signal was also plotted in Figure 4. It was checked that 0.00% of the differences were at least as extreme as (95%) our observed difference. We reject the H_$_{\text{Null}}$ and accept H_$_{\text{Alternatative}}$. We conclude that the fault1 does impact the OA damper control signal.

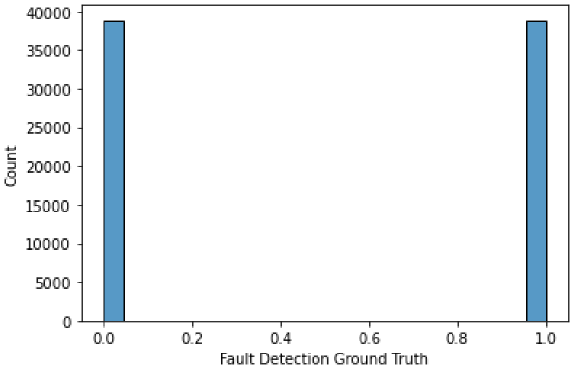


Figure 2: Count of Normal (Fault Truth = 0) and Fault (Fault Truth = 1) scenarios

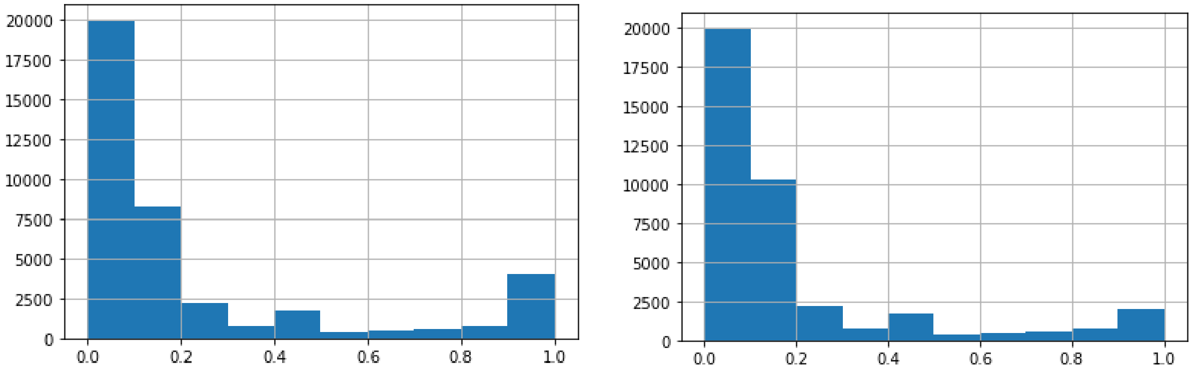


Figure 3: Histograms of OA damper control signal (a) normal conditions (b) fault condition

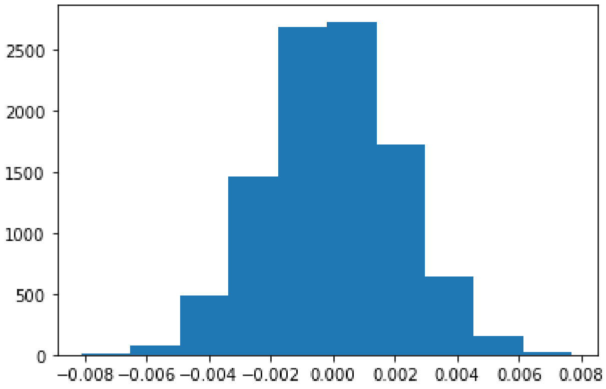
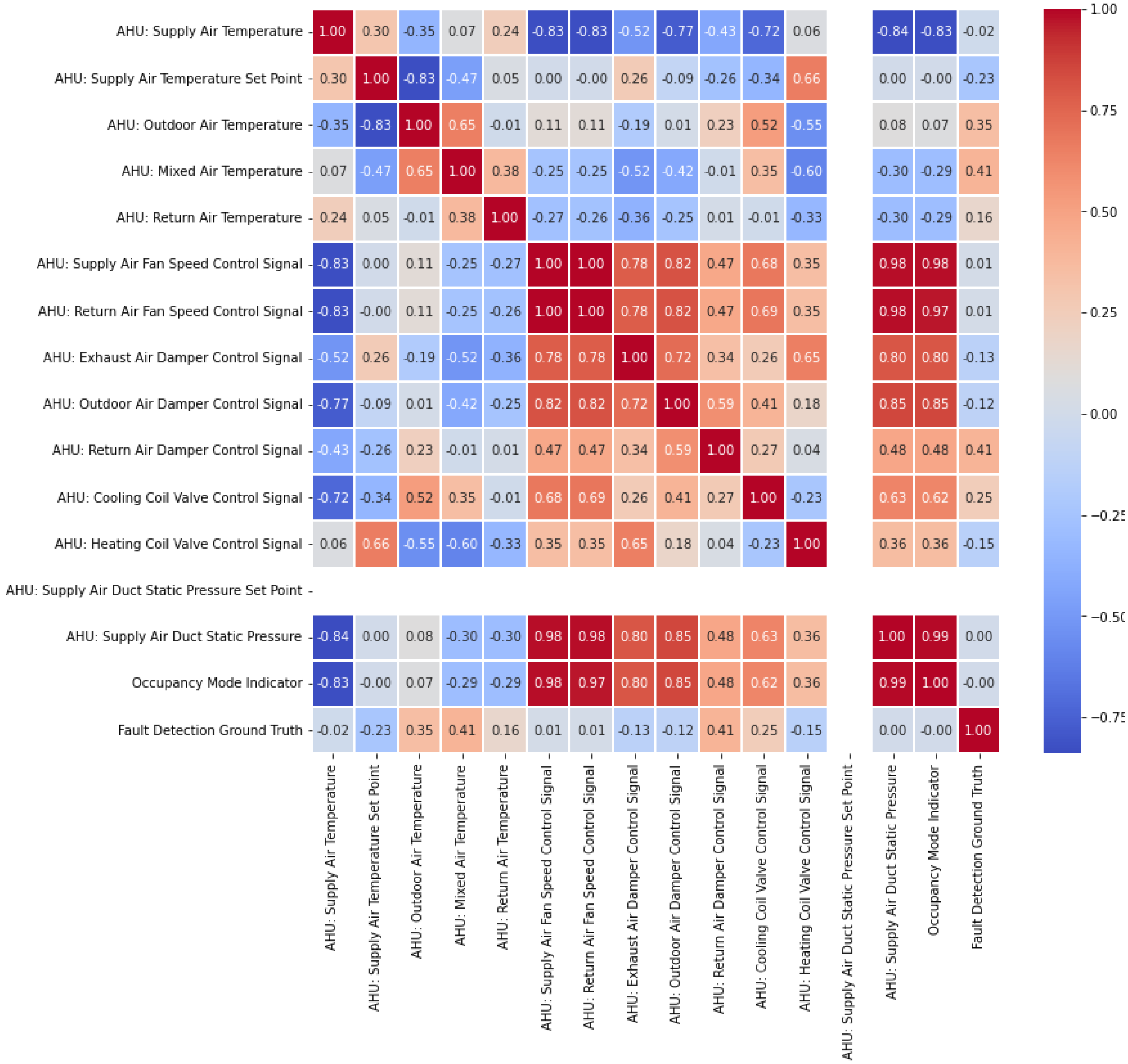


Figure 4: Histograms of OA damper control signal difference

To visualize the relationships of the features with the fault truth, the heatmap of the features was plotted in Figure 5. The top four features on the heatmap are listed here:

- the mixed air temperature (0.41)
- return air damper control signal (0.41)
- outdoor air temperature (0.35)
- cooling valve control signal (0.25)

From the domain knowledge of HVAC, theoretically we knew that the OAT sensors have impacts, maybe significantly, on these four features.



	AHU: Supply Air Temperature	AHU: Supply Air Temperature Set Point	AHU: Outdoor Air Temperature	AHU: Mixed Air Temperature	AHU: Return Air Temperature	AHU: Supply Air Fan Speed Control Signal	AHU: Return Air Fan Speed Control Signal	AHU: Exhaust Air Damper Control Signal	AHU: Outdoor Air Damper Control Signal	AHU: Return Air Damper Control Signal	AHU: Cooling Coil Valve Control Signal	AHU: Heating Coil Valve Control Signal	AHU: Supply Air Duct Static Pressure Set Point	AHU: Supply Air Duct Static Pressure	Occupancy Mode Indicator	Fault Detection Ground Truth
AHU: Supply Air Temperature	1.00	0.30	-0.35	0.07	0.24	-0.83	-0.83	-0.52	-0.77	-0.43	-0.72	0.06		-0.84	-0.83	-0.02
AHU: Supply Air Temperature Set Point	0.30	1.00	-0.83	-0.47	0.05	0.00	-0.00	0.26	-0.09	-0.26	-0.34	0.66		0.00	-0.00	-0.23
AHU: Outdoor Air Temperature	-0.35	-0.83	1.00	0.65	-0.01	0.11	0.11	-0.19	0.01	0.23	0.52	-0.55		0.08	0.07	0.35
AHU: Mixed Air Temperature	0.07	-0.47	0.65	1.00	0.38	-0.25	-0.25	-0.52	-0.42	-0.01	0.35	-0.60		-0.30	-0.29	0.41
AHU: Return Air Temperature	0.24	0.05	-0.01	0.38	1.00	-0.27	-0.26	-0.36	-0.25	0.01	-0.01	-0.33		-0.30	-0.29	0.16
AHU: Supply Air Fan Speed Control Signal	-0.83	0.00	0.11	-0.25	-0.27	1.00	1.00	0.78	0.82	0.47	0.68	0.35		0.98	0.98	0.01
AHU: Return Air Fan Speed Control Signal	-0.83	-0.00	0.11	-0.25	-0.26	1.00	1.00	0.78	0.82	0.47	0.69	0.35		0.98	0.97	0.01
AHU: Exhaust Air Damper Control Signal	-0.52	0.26	-0.19	-0.52	-0.36	0.78	0.78	1.00	0.72	0.34	0.26	0.65		0.80	0.80	-0.13
AHU: Outdoor Air Damper Control Signal	-0.77	-0.09	0.01	-0.42	-0.25	0.82	0.82	0.72	1.00	0.59	0.41	0.18		0.85	0.85	-0.12
AHU: Return Air Damper Control Signal	-0.43	-0.26	0.23	-0.01	0.01	0.47	0.47	0.34	0.59	1.00	0.27	0.04		0.48	0.48	0.41
AHU: Cooling Coil Valve Control Signal	-0.72	-0.34	0.52	0.35	-0.01	0.68	0.69	0.26	0.41	0.27	1.00	-0.23		0.63	0.62	0.25
AHU: Heating Coil Valve Control Signal	0.06	0.66	-0.55	-0.60	-0.33	0.35	0.35	0.65	0.18	0.04	-0.23	1.00		0.36	0.36	-0.15
AHU: Supply Air Duct Static Pressure Set Point																
AHU: Supply Air Duct Static Pressure	-0.84	0.00	0.08	-0.30	-0.30	0.98	0.98	0.80	0.85	0.48	0.63	0.36		1.00	0.99	0.00
Occupancy Mode Indicator	-0.83	-0.00	0.07	-0.29	-0.29	0.98	0.97	0.80	0.85	0.48	0.62	0.36		0.99	1.00	-0.00
Fault Detection Ground Truth	-0.02	-0.23	0.35	0.41	0.16	0.01	0.01	-0.13	-0.12	0.41	0.25	-0.15		0.00	-0.00	1.00

Figure 5: Heatmap of the features with the Fault Truth

Figure 6 showed the outdoor air damper control signal with OAT for a typical air handling unit (AHU) serving a large office building that is the same type of AHU as investigated in this project. The data of outdoor air damper control signals were plotted with OAT in Figure 7 for the normal (blue dots) and fault (orange dots) conditions. We knew that most of the operation was in the free cooling mode, in which the outdoor air damper modulated the openness to maintain the supply air temperature at its setpoint (55 °F). Except for some operation points that AHU switched between the occupancy and un-occupancy modes, the fault conditions (orange dots) shifted 2 °F to the right side than the normal conditions (blue dots). It indicated that the OAT sensor offset 2 °F higher than the actual values.

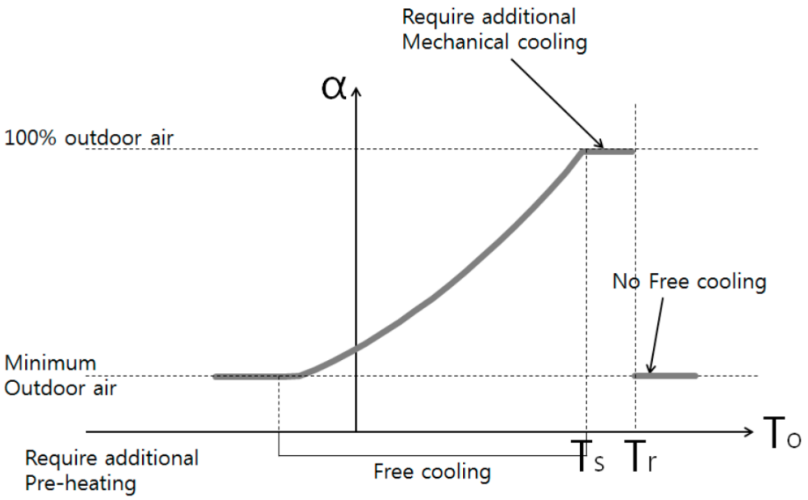


Figure 6: Outdoor air damper control signal vs OAT for a typical AHU

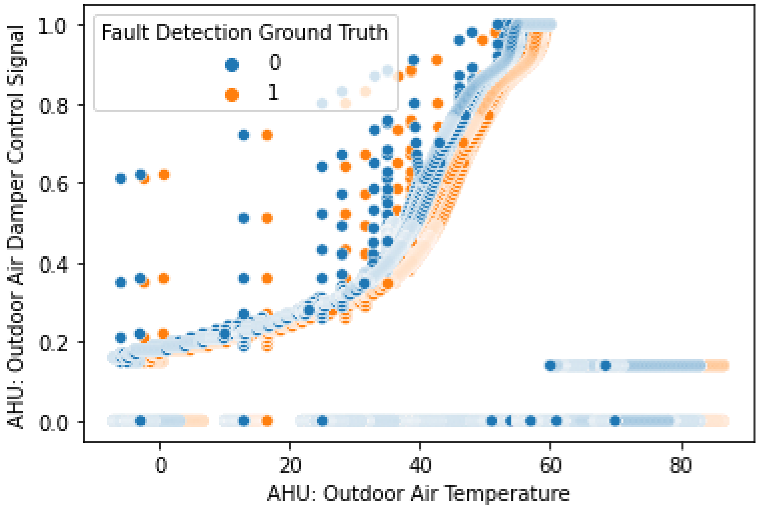


Figure 7: Outdoor air damper control signal vs OAT under the normal and fault conditions

3. Model Selection

For the selected fault detection, six machine learning classification models were compared. They were logistic regression, KNN, SVM, Random Forest, Naive Bayes, and Gradient

Boost. The model accuracy scores and the ROC-AUC scores were plotted in Figure 8. Obviously, KNN, Random Forest, and Gradient Boost had the model accuracy scores above 0.985, which were significantly higher than the other three models (about 0.55). It has a similar trend for ROC-AUC scores as the model accuracy scores. From Figure 8(b), the training scores were quite close to the test scores for each model. The models investigated were neither overfit or underfit the dataset .

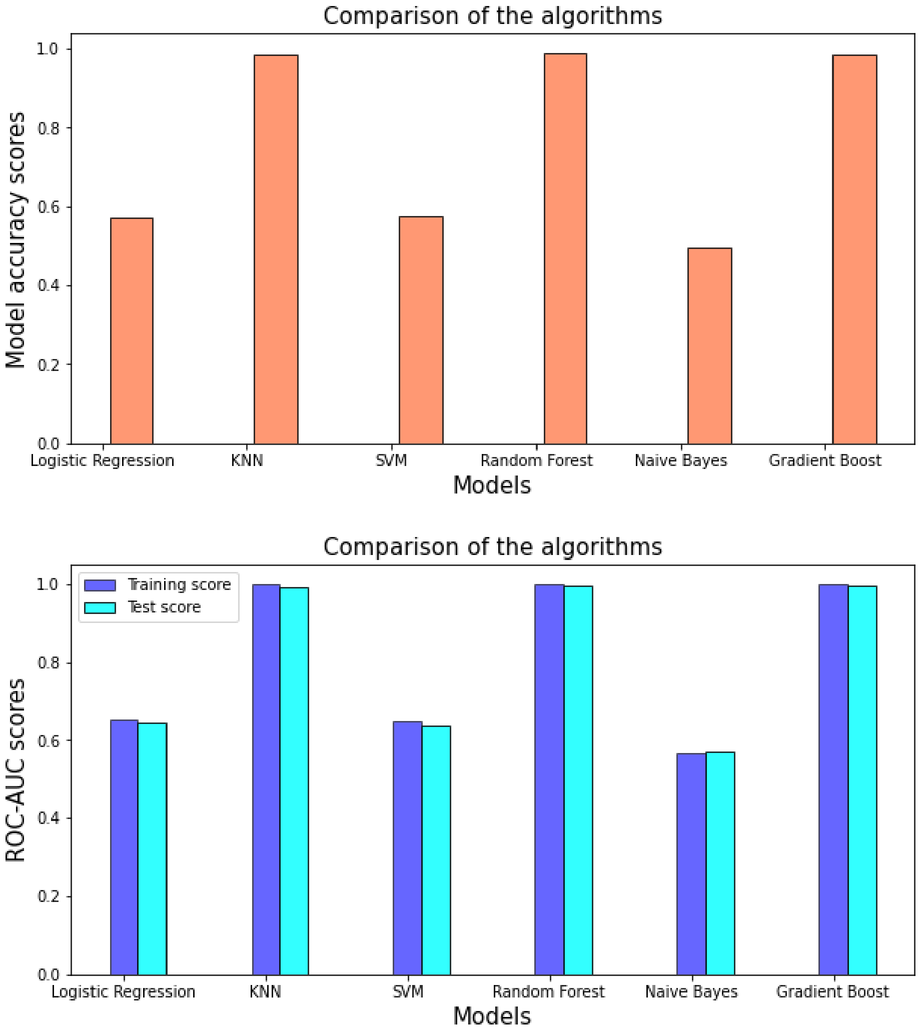


Figure 8: Model scores comparison (a) model accuracy scores (b) ROC-AUC scores

We selected two models - Random Forest (RF) and Gradient Boost (GB) to optimize the hyperparameters by grid search cv, which increased the model accuracy scores and the ROC-AUC scores slightly. Then, the features were ranked based on the RF and GB models, which were plotted in Figure 9. The relative importances was slightly different in values, but the rankings were exactly same for the two models.

The ROC curve for each model almost vertically shot up from zero to nearly one (about 0.9992) with the false positive rate (FPR), which showed that the models of RF and GB are robust and accurate as one of the ML classification models in this project.

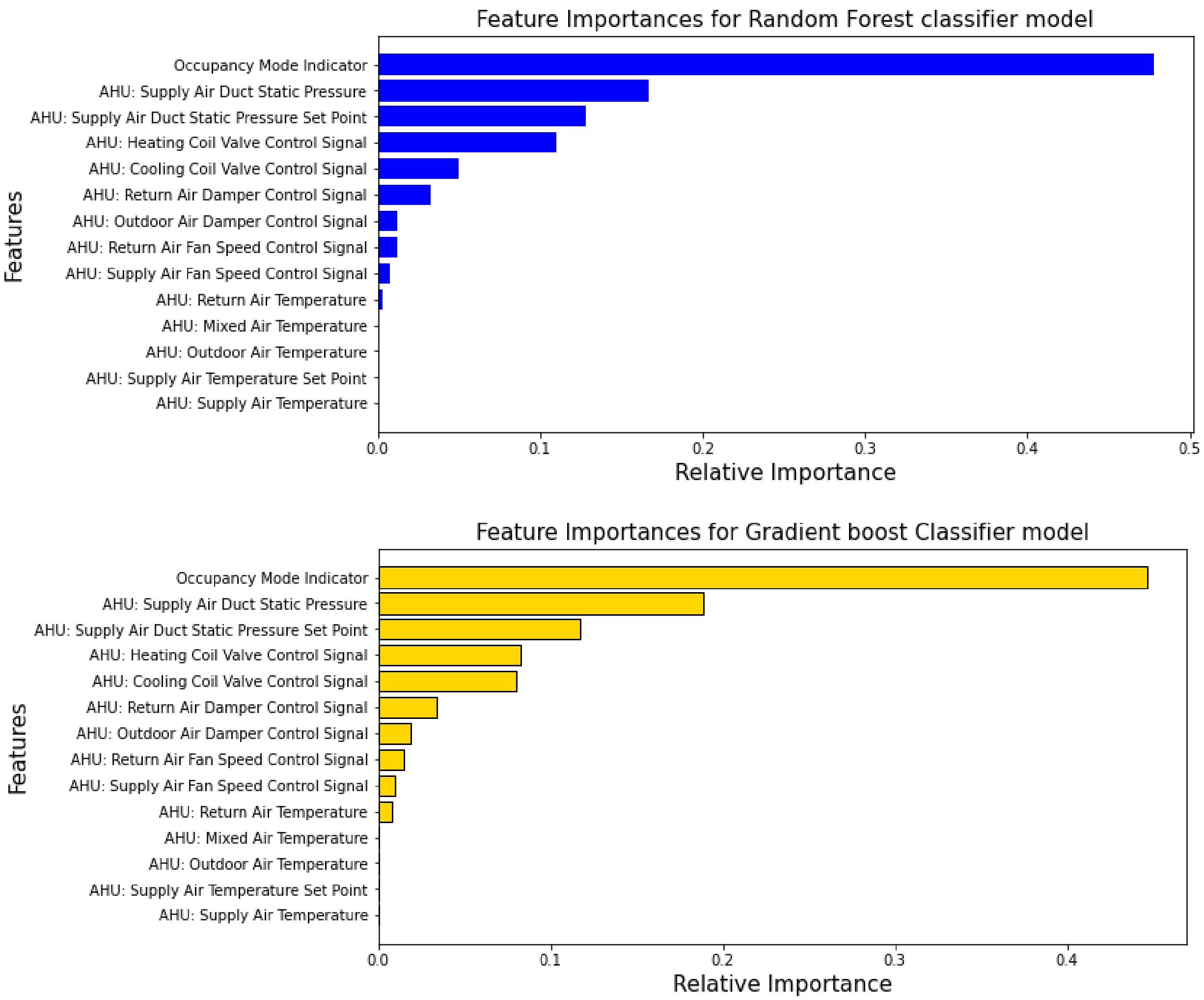


Figure 9: Feature ranking based on ML models (a) Random Forest (b) Gradient Boost

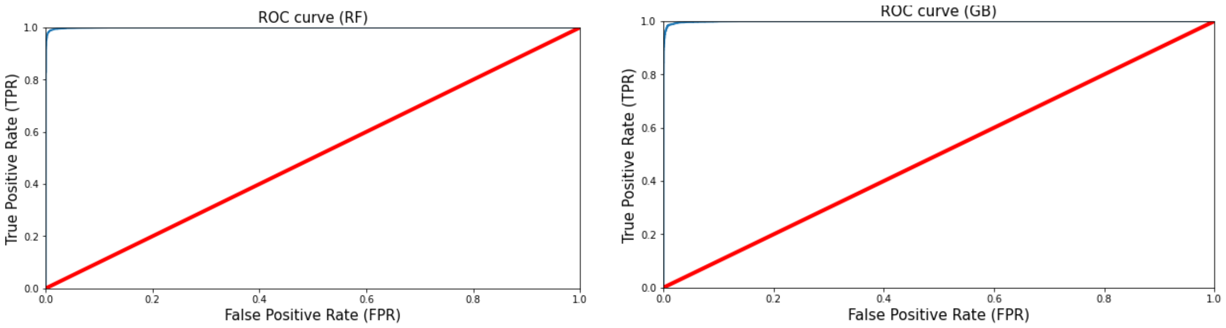


Figure 10: ROC curves based on ML models (a) Random Forest (b) Gradient Boost

4. Takeaways

Evaluating the performance of a model by training and testing on the same dataset can lead to the overfitting. Hence the model evaluation is based on splitting the dataset into train and validation set. But the performance of the prediction result depends upon the random choice of the pair of (train,validation) set. Inorder to overcome that, the Cross-Validation procedure is used where under the k-fold CV approach, the training set is split into k smaller sets, where a model is trained using k-1 of the folds as training data and the model is validated on the remaining part.

We have evaluated each model in terms of model accuracy score, and 'ROC-AUC' score for both the training and test data, and plotted them. The two best performing models are the Random forest and the Gradient boost. Both are the ensemble model, based on decision trees.

Next, we have carried out the grid search CV for the hyperparameter tuning for both the models separately. This step was the most time consuming one in terms of computation. (The RF model took much longer time). With the result of the optimized hyperparameters, we have again fitted the two models, and got the predictions separately.

We have evaluated the ROC-AUC scores with the optimized hyperparameters. Clearly, the model performance improved with the optimized parameters. The final ROC-AUC scores fro both RF and the GB are 0.99941 and 0.99916.

5. Future Research

This project enlightened me on the classification models to detect faults in the air handling units for a large office building. While my models were valuable to this specific project or the dataset that were built by experiments and simulation, it is difficult to say how useful they would be in the real application if one fault occurs. We would like to know how much data is needed to detect a single fault.

The faults in this project were mimicked by experiments or simulation. If a single fault is not steady or stable, it is interesting if the models developed in this project are applicable. Furthermore, we would like to know how the models respond to the multi faults occurring at the same time.

Finally, the current classification models were developed for fault detection. A question will arise: how to diagnose a single fault or integrated faults in the real application. Especially for the integrated faults, how to decouple the faults and find the root cause will be more challenging.