Predictive Maintenance of Air Handling Units (AHU's)

1 Summary of Internship

1.1 Background

The evolution of monitoring and designing energy-efficient systems spans decades and continues with daily technological advancements. Projections indicate alarming building energy consumption growth, with buildings using 40% of global primary energy and contributing 33% of carbon emissions. Studies propose up to 30% energy reduction via enhanced monitoring protocols. HVAC systems excel here, offering personalized control, compact equipment, and contingency planning. They employ passive cooling methods like high thermal mass and night ventilation, crucial in hot climates. A malfunctioning ventilation system in a bustling supermarket disrupts activities, underscoring HVAC's importance.

1.2 Motivation

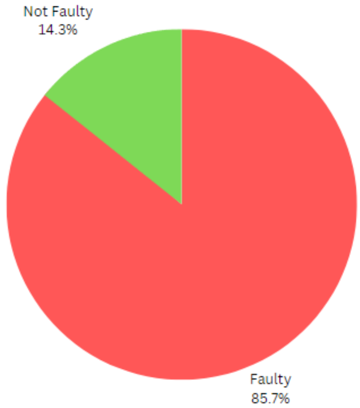
Traditional cooling systems are costlier than HVAC systems. Establishing sustainable emergency maintenance in Bangladesh, considering traffic, finances, and job loyalty, proves challenging. Manual identification of ventilation system faults is time-consuming, leading to complex decision-making. Human fallibility adds to the issue. Earlier studies used the Artificial Immune Recognition System. Implementing a closely monitored, well-trained machine system capable of prediction and prevention is preferable. This involves training both the system and personnel to interpret sensor data for anomaly detection. Timely preventive actions can prevent catastrophic disruptions and resource losses. A semi-supervised approach has also been explored in this context.

1.3 Objectives

- ➢ Developing a model to optimize and provide comfort through various HVAC modules and sensor data.
- ➢ Utilizing MPC to determine energy-efficient actions based on sensor readings and environmental conditions.
- ➢ Addressing faults and energy efficiency decline in HVAC systems through automated fault detection and diagnosis (FDD).
- ➢ Utilizing temperature, humidity, and pressure data for FDD and maintenance optimization.
- ➢ Applying a hybrid model-based approach for HVAC systems in trains, focusing on sustainability and fault modeling.
- ➢ Using synthetic data and machine learning for fault classification.

2 Contribution

2.1 Dataset Details



I have applied several machine learning (ML) techniques to the Multi-Zone Variable Air Volume AHU dataset (MZVAV-1), which comprises 272,160 rows and 18 numerical columns. The exception is the "Datetime" column. The target column, labeled "Fault Detection Ground Truth," holds values of 0 for non-faulty instances and 1 for faulty instances.
The aim of the datasets is to determine the Fault ground truth.We can get an idea about the frequency distribution about the 'Fault Ground Truth' column from the pie chart below.

2.2 Preprocessing

Feature Scaling (Standardization): By using StandardScaler, I've standardized the features by transforming them to have a mean of 0 and a standard deviation of 1. Standardization ensures that all features contribute equally to the learning process, preventing features with larger scales from dominating the model.

Feature Selection (SelectKBest): With SelectKBest and ANOVA F-value scoring, I've performed feature selection to choose the most important features. This can enhance model performance by reducing noise and focusing on the most relevant information. Selecting the top k=1-13 features helps simplify the model, potentially improving both training speed and generalization to unseen data.

Handling Class Imbalance (Oversampling): Dealing with class imbalance is crucial to ensure that the model doesn't become biased toward the majority class. By using the resample function to oversample the minority class (faulty instances), I've balanced the class distribution. This can lead to improved model performance, especially for algorithms that are sensitive to class imbalance and might not generalize well on imbalanced data.

2.3 Experimentation with ML Algorithms

Random Forest:

Random Forest algorithm was used on the dataset and experimenting with various combinations of hyperparameters like n_estimators and max_depth is a common and effective approach. Random Forest is an ensemble learning method that combines multiple decision trees to make more accurate predictions.Hence it's generally observed that increasing the values of n_estimators and max_depth in a Random Forest algorithm can lead to an increase in accuracy, up to a certain point.

N_estimators	max_depth	min_samples_split	max_features	Accuracy
100	10	2	sqrt	0.8746
200	20	2	sqrt	0.9443
300	30	2	sqrt	0.9868
400	40	2	sqrt	0.9894

2.4 Model Development :

2.4.1 AdaBoost with Decision Tree:

The approach of combining AdaBoost with decision trees and experimenting with various combinations of hyperparameters is a solid methodology for improving model performance and achieving better results on your dataset.

N_estimators	max_depth	Accuracy
50	1	0.8588
100	3	0.9201
100	5	0.9699
100	7	0.9891
100	9	0.9930
100	11	0.9936
100	13	0.9945
100	15	0.9940

N_estimators	Accuracy	max_depth
200	0.8588	1
200	0.9201	3
200	0.9863	5
200	0.9914	7
200	0.9929	9
200	0.9933	11
200	0.9939	13
200	0.9940	15

GaussianNB :

Method 1 :
Accuracy achieved with gaussianNB method is 0.14 (14%) with no preprocessing techniques.

Method 2 :
We use feature scaling, SelectKBest preprocessing techniques to boost the accuracy of the model, but still we find it difficult to to cross 80 %.

K	**Accuracy**	**Confusion Matrix**
1	0.5971	[[2958 4735] [17191 29548]]
2	0.7804	[[890 6803] [5146 41593]]
3	0.7762	[[921 6772] [5409 41330]]
4	0.7704	[[998 6695] [5801 40938]]
5	0.7711	[[978 6715] [5742 40997]]

By the above table we can identify the most valuable column would be “AHU: Cooling Coil Valve Control Signal”

Method 3 :

RandomizedSearchCV performs hyperparameter tuning through a random search over specified parameter values. hyperparameters, evaluates their performance using cross-validation, and selects the best combination of hyperparameters based on the provided scoring metric. The accuracy obtained: 0.841.

3 Conclusions

3.1 Key Results :

- ➢ Random Forest showed increasing accuracy as n_estimators and max_depth increased, with an accuracy of up to 98.94%.
- ➢ AdaBoost with decision trees exhibited significant accuracy improvements with different max_depth values, achieving up to 99.45% accuracy.
- ➢ GaussianNB results varied based on preprocessing techniques, emphasizing the value of feature selection.
- ➢ Hyperparameter tuning using RandomizedSearchCV improved GaussianNB accuracy to 84.1%.

3.2 Key Insights and Contributions:

- The most valuable feature for fault detection was identified as "AHU: Cooling Coil Valve Control Signal."
- The combination of preprocessing techniques and hyperparameter tuning significantly influenced model performance.
- The experiments underscored the potential of machine learning in predictive maintenance and fault detection in AHUs.

3.3 Conclusion:

The internship successfully demonstrated the value of machine learning in predictive maintenance of Air Handling Units (AHUs). Through dataset analysis, preprocessing, and algorithm experimentation, we achieved notable improvements in fault detection accuracy. The findings emphasize the potential for technology-driven solutions to enhance energy efficiency, reduce disruptions, and contribute to more sustainable building operations. Further research and implementation of these techniques could lead to substantial real-world benefits.

4 References

[1] Md. Zubayer Ahmed Fahim, Tiash Roy, Mma Rakib, Shihab Sharar. (2021). Predictive Maintenance of HVAC System using supervised Machine Learning Algorithms.