## **HVAC Data Processing and PPO Model Training – Project Report**

There are two jupyter files:  
**1.** **preprocessing\_and\_feature\_engineering.ipynb** It contains preprocessing of data and feature engineering.  
**2. env\_and\_model\_training.ipynb** It contains Custom OpenAI Gymnasium Environment, model training and evaluation.

### **Data Preprocessing & Feature Engineering**

Started by loading and cleaning the equip and weather dataset.

Data Analysis & cleaning:

Performed EDA on both data sets, checked nulls, fixed timestamps (datetime columns), combined equip and weather data.

Feature engineering:

Time-based features: hour\_sin, hour\_cos, minute\_sin, minute\_cos, dayofweek, is\_weekend for periodic patterns. This is done because when time changes from 23:00 to 00:00 it looks like there is a difference of 23 but infact the difference is of 1 hour.

Delta features: delta\_supply\_return, delta\_outdoor\_indoor, delta\_humidity to capture environmental differences.

Lag features: One-step lags for Valve, RaTemp, SaTemp, RaHumidity to give the agent short-term memory.

### **Custom OpenAI Gymnasium Environment**

Created a custom HVACEnv environment to simulate HVAC control:

Action space: Continuous control of the HVAC valve between 0–100%.

Step function:

Updates the environment with the chosen action.

Returns next observation, reward, and done flag.

Episode termination: Ended when dataset was exhausted.

### **Custom Reward Strategy**

Designed a reward function balancing:

1. Comfort – Penalized large deviation of room temperature from the target comfort range (e.g., 22–24°C).
2. Energy Efficiency – Penalized high thermal energy usage.
3. Smoothness – Penalized sudden large changes in valve position to protect equipment.

This encouraged the agent to maintain comfort with minimal energy consumption.

### **PPO Model Initialization**

Used Stable-Baselines3 PPO:

Training environment wrapped in DummyVecEnv for vectorized operation.

PPO hyperparameters:

* + policy= "MlpPolicy",
  + verbose=1,
  + learning\_rate=5e-5,
  + batch\_size=256,
  + n\_epochs=15,
  + clip\_range=0.15,
  + clip\_range\_vf=None,
  + ent\_coef=0.001,
  + vf\_coef=0.25,
  + max\_grad\_norm=0.5,
  + gae\_lambda=0.98,

### **Model Training**

Trained the PPO agent using the train split (80% of dataset).

* Input: Observations from dataset.
* Output: Continuous valve position.
* Training lasted for defined timesteps until reward convergence.

### **Model Evaluation on Test Data**

Evaluated the trained agent on the test split (20% of dataset):

* The agent interacted with the HVACEnv using unseen data.
* Collected:
  + Rewards per timestep
  + Energy usage trends
  + Room temperature behavior
* Plotted reward curves and energy consumption patterns to visualize performance.

### **Performance, Improvement & Suggestions**

After first training the model the MAE was 59, then after changing hyperparameters and timesteps the MAE was 8 which can further be decreased by:

1. Feature Expansion – Include more lag steps and weather forecast data for better anticipation.
2. Reward Tuning – Adjust comfort/energy penalty balance for specific optimization goals.
3. Longer Training – More PPO updates and fine-tuned hyperparameters could improve stability.