Boiler Efficiency Drop – Synthetic Faults

Name: Priyanshu Kumar
Roll No. 210107066

Submission Date: 25th April 2024

Final Project submission
Course Name: Applications of AI and ML in chemical engineering
Course Code: CL653

Executive Summary

Overview

Boilers are the main source of heat for District Heating and Cooling (DHC) systems, which are essential for the distribution of energy in cities. However, these boiler's increasing operational costs and negative environmental effects are a result of their heat losses and efficiency reduction. We aim to identify early indicators of thermal inefficiencies and efficiency reductions by utilising data-driven methodologies. This allows for timely adjustments to achieve optimal performance and sustainability. Within the complex framework of DHC infrastructures, this integrated strategy strives to improve energy efficiency, lower costs, and limit environmental effect.

Problem Statements

1) Detection of efficiency drop of a reboiler
2) Detection of thermal losses for a DHC boiler

Introduction

Background

A boiler is a device that is used in many industrial activities, such as heating and electricity generation. Its purpose is to turn liquid into vapour. Boilers date back to the first century AD, but they became well-known in the 17th century when Denis Papin designed the safety valve. Modern alloy steel boilers are designed to endure high temperatures and pressures, unlike the wrought iron boilers of the past.

There are two main types of boilers: water-tube and fire-tube. Whereas water-tube boilers circulate hot furnace gases outside of water-filled tubes, fire-tube boilers transmit heat through steel tubes submerged in water. The latter was developed in response to the need for high-pressure steam generation and is widely used in industrial settings for things like manufacturing operations and ship propulsion.

Boilers are important in areas other than industry; in District Heating and Cooling (DHC) systems, they are indispensable. DHC systems provide localised areas with heating or cooling through networked infrastructure, lowering carbon emissions and providing an energy-efficient alternative. An essential component of DHC systems, boilers guarantee the supply of thermal energy needed for room heating and other uses.

In conclusion, boilers—which have been a part of industrial processes and DHC systems for centuries—remain essential, representing the fusion of historical development and modern engineering concepts to meet energy demands in a sustainable manner.

Description of Problem

The primary focus of the issue statement is maximising boiler dependability and efficiency in industrial operations as well as District Heating and Cooling (DHC) systems. Improving heat transfer mechanisms, combustion processes, and operational performance are all part of this strategy to optimise energy use and reduce environmental effect. The objective is to handle issues with thermal energy generation and distribution in a sustainable way by utilising modern technical methods.

It is imperative to address the issues of boiler efficiency and reliability for a number of reasons. First off, improved boiler performance lowers operating costs and has a positive environmental impact by increasing energy efficiency. Second, better boiler performance boosts output and quality of the final product in industrial processes. Thirdly, effective boilers guarantee dependable thermal energy supply in District Heating and Cooling (DHC) systems, promoting environmentally friendly urban heating and cooling options. In general, resolving these problems advances environmental stewardship, cost effectiveness, and energy sustainability.

Objectives

This project has 2 major objectives:

1) Detection of efficiency drop of a reboiler
2) Detection of thermal losses for a DHC boiler

We intend to use machine learning methodologies to solve these problems.

Methodologies

Data Description

This dataset contains data generated in the AI DHC project. This dataset contains synthetic fault data for efficiency drop and thermal losses of a DHC boiler. The IEA DHC Annex XIII project “Artificial Intelligence for Failure Detection and Forecasting of Heat Production and Heat demand in District Heating Networks” is developing Artificial Intelligence (AI) methods for forecasting heat demand and heat production and is evaluating algorithms for detecting faults which can be used by interested stakeholders (operators, suppliers of DHC components and manufacturers of control devices).

Data Characteristics

1) **Time_s:** This variable indicates the time at which data points were recorded , expressed in seconds.
2) **P_fuel_kW:** This indicates the fuel source's power output, which is usually expressed in kilowatts (kW). The energy input into the boiler system is quantified by this parameter.
3) **P_thermal_kW:** This variable, which is likewise expressed in kilowatts (kW), represents the boiler's thermal power output. It shows how much heat energy is moved from the boiler to the application or system.
4) **T_out_degC:** This indicates, in degrees Celsius (°C), the fluid's (often water or steam) output temperature as it leaves the boiler. This parameter shows the fluid's temperature following heating in the boiler.
5) **G_losses:** This variable shows the losses that occur within the boiler system. These losses may be the result of incomplete combustion, radiation losses, or heat losses to the environment. It is a parameter without dimensions.
6) **Eta:** This variable, which stands for efficiency, shows how well the boiler system performs overall in transforming fuel energy into useful thermal energy. It is usually stated as a percentage and is a dimensionless parameter.
7) **T_ext_degC:** This variable, expressed in degrees Celsius (°C), indicates the outside or ambient temperature around the boiler system. Boiler performance can be affected by external temperature, particularly in terms of heat exchange efficiency.
8) **P_demand_kW:** This stands for the network thermal load, which is the amount of thermal energy that the boiler-connected system or application needs. Kilowatts (kW) is the unit of measurement for this characteristic.

9) **T_supply_degC:** This variable shows the temperature of the thermal fluid provided to the system or application that is linked to the boiler; it is a representation of the network supply temperature. Degrees Celsius are used to measure it (°C).
10) **T_return_degC:** This is an acronym for the network return temperature, which is the temperature of the thermal fluid that is returning to the boiler from the application or system. This parameter is essential for evaluating the performance and efficiency of the system; it is measured in degrees Celsius (°C).

Nature of Data

Given that the project's data includes time-dependent variables like power, temperature, and thermal load, it appears that the data is dynamic in nature. Dynamic data indicates that the parameters of the system are always changing, which is representative of real-world situations where conditions are constantly changing. For this project, a thorough grasp of the boiler's behaviour under various operating situations is made possible by the analysis of dynamic data, which makes it easier to construct prediction models and control strategies. It improves the sustainability, dependability, and effectiveness of the boiler system within District Heating and Cooling (DHC) networks by enabling the detection of transient states, optimisation of operational parameters, and proactive maintenance interventions.

Data Preprocessing

Several data pretreatment procedures can be required for this project in order to guarantee the precision and potency of the analysis:

1) **Cleaning:** This is the process of locating and dealing with incorrect or missing data points. The integrity of the dataset is ensured, for instance, by eliminating outliers or imputing missing values in variables like temperature or power measurements.
2) **Normalisation:** To bring all the variables in the dataset to a consistent scale, normalisation techniques like Min-Max scaling or z-score normalisation can be used. This is because the dataset may contain variables with different scales or units (e.g., temperature in Celsius and power in kilowatts).
3) **Transformation:** Changing some variables could improve the analysis's readability or efficacy. For example, changing skewed distributions of some variables logarithmically or converting temperature values from Celsius to Kelvin can enhance the performance of the model.
4) **Feature engineering:** This can enhance a dataset by generating new features from preexisting ones or by extracting pertinent data. For instance, obtaining further

elements from temperature data, such as heating or cooling degree days, could offer insightful information on seasonal changes in energy consumption.

5) **Managing Categorical Variables:** To transform the categorical variables in the dataset into a numerical format that can be analysed, encoding techniques like one-hot encoding or label encoding may be used. Examples of these variables include boiler kinds and fuel sources.
6) **Temporal Aggregation:** Because time-series data are dynamic, it may be useful to aggregate them into meaningful intervals (such as hourly, daily, or monthly) for analysis purposes. This will enable the detection of trends and patterns.

The dataset can be improved and optimised for further analysis by putting these preprocessing methods into practice. This will allow for more precise modelling, prediction, and decision-making when it comes to boiler performance and efficiency in District Heating and Cooling (DHC) systems.

Correlation Heatmap

Tools and Technologies

Important open-source tools and libraries for the boiler performance anomaly detection project in DHC systems are: TensorFlow for deep learning flexibility; Pandas for structured data manipulation; NumPy for effective numerical computing; Matplotlib with Seaborn for perceptive data visualisation; and Scikit-learn for flexible machine learning algorithms and model evaluation. The tools and methods provided by Scikit-learn help with model selection and evaluation, and TensorFlow allows for more complex deep learning experiments. Pandas and NumPy efficiently manage the preprocessing and manipulation of data, while Matplotlib in conjunction with Seaborn offers visualisation features for data exploration. When combined, these instruments strengthen the creation of strong models and improve boiler performance anomaly detection.

Implementation Plan

Model Selection

Random Forest Classifier

RFC is selected because to its efficiency in managing classification assignments, which corresponds with the issue of identifying abnormalities in boiler operation (denoted by the 'anomaly' target variable). RFC is renowned for its resistance to overfitting, high-dimensional feature space handling capabilities, and robustness to noisy data.

RFC is an excellent choice for detecting patterns and anomalies in the boiler efficiency data because of the binary categorization nature of the issue (normal vs. anomalous). The number of estimators, maximum features, and minimum samples per leaf—all examples of hyperparameters selected through optimization—show how efforts have been made to strike a compromise between model complexity and performance.

Logistic Regression

For binary classification applications, LR is a well-liked linear classification technique. It was selected because to its ease of use, readability, and effectiveness with big datasets. LR is appropriate for situations when the decision boundary is linear or nearly linear because it makes the assumption that features and the log-odds of the target variable have linear relationships.

LR can be useful for binary classification jobs with well-separated classes even though it is more straightforward than ensemble techniques like RFC. Within the project, thermal conductance is represented by the 'G_losses' variable, which is the basis for anomaly detection using LR. Because LR can produce interpretable coefficients, it can shed light on how different factors affect anomaly detection.

Support Vector Machines

Support Vector Machines (SVM) are powerful supervised learning models used for classification and regression tasks. They work by finding the optimal hyperplane that separates classes in a high-dimensional space. SVMs aim to maximize the margin between classes, making them effective for handling complex datasets with non-linear relationships.

Training

Model Fitting

The fit() method is used to fit the chosen machine learning model to the preprocessed training data. At this stage, the model's parameters are changed to reduce the difference between the expected and actual results.

Hyper-Parameter Tuning

To determine the combination that produces the highest performance, grid search and randomised search approaches are used to optimise the models' hyperparameters.

Evaluation Metrics

Matthews Correlation Coefficient (MCC)

Since MCC accounts for true positives, true negatives, false positives, and false negatives, it is appropriate even for datasets that are unbalanced.
When it comes to binary classification jobs, like identifying performance abnormalities in boilers, MCC is especially pertinent.

Accuracy Score

An overall indicator of the model's accuracy in differentiating between anomalies and non-anomalies is its accuracy.

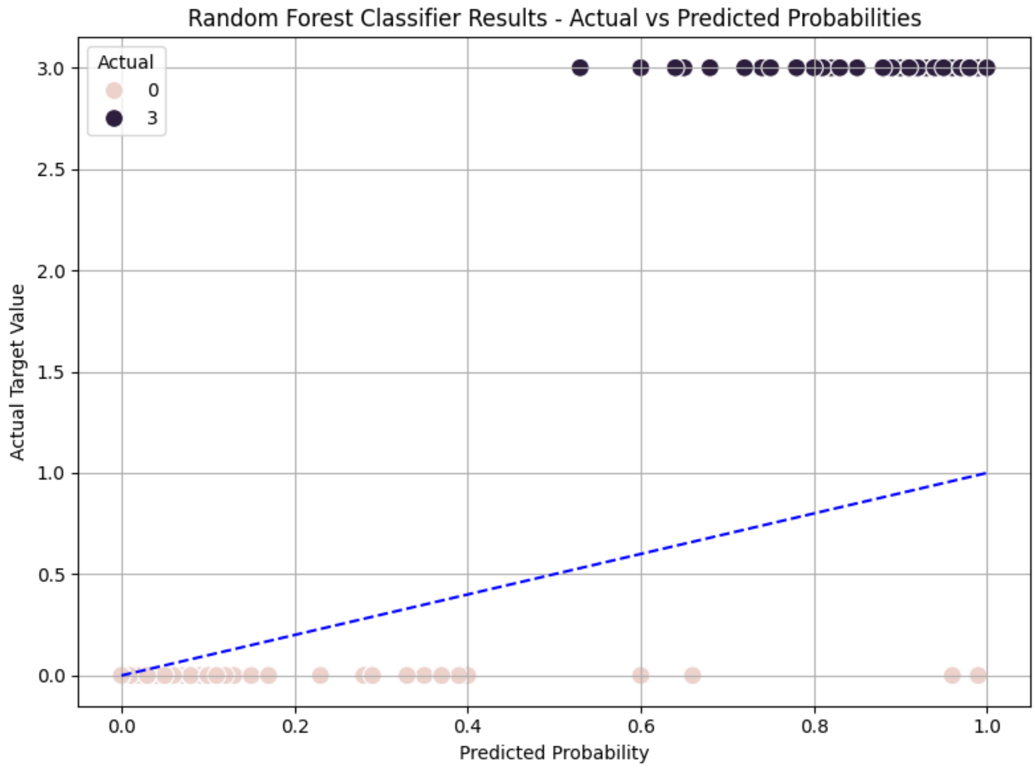
It provides a clear understanding of the model's performance, which is a supplement to MCC.

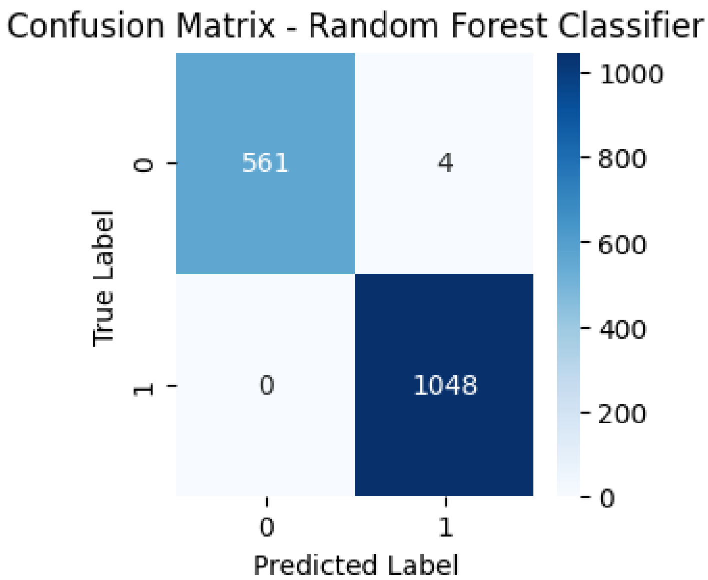
Testing and Deployment

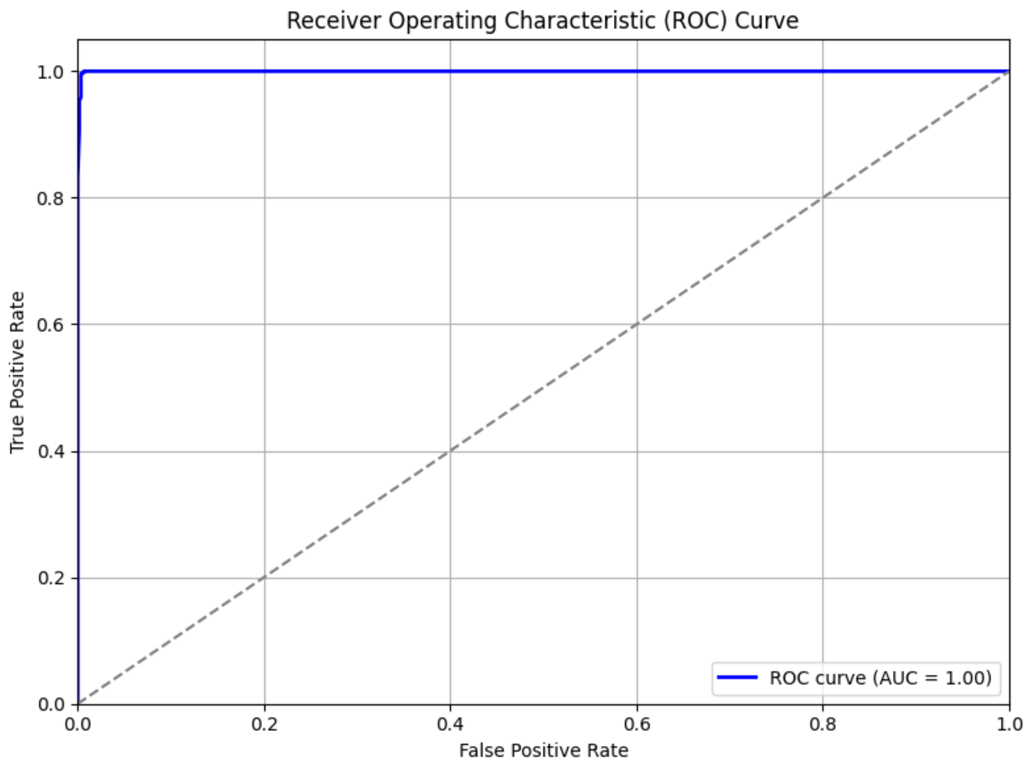
There's no deployment part included, but it's assumed that the project will be deployed in the future for real-world use. Deployment involves making the trained model accessible for predictions in a production environment. This step is crucial for utilizing the model's predictive power on new data. However, at this stage, the focus is solely on model optimization, leaving deployment for a later phase.

Results

Random Forest Classifier







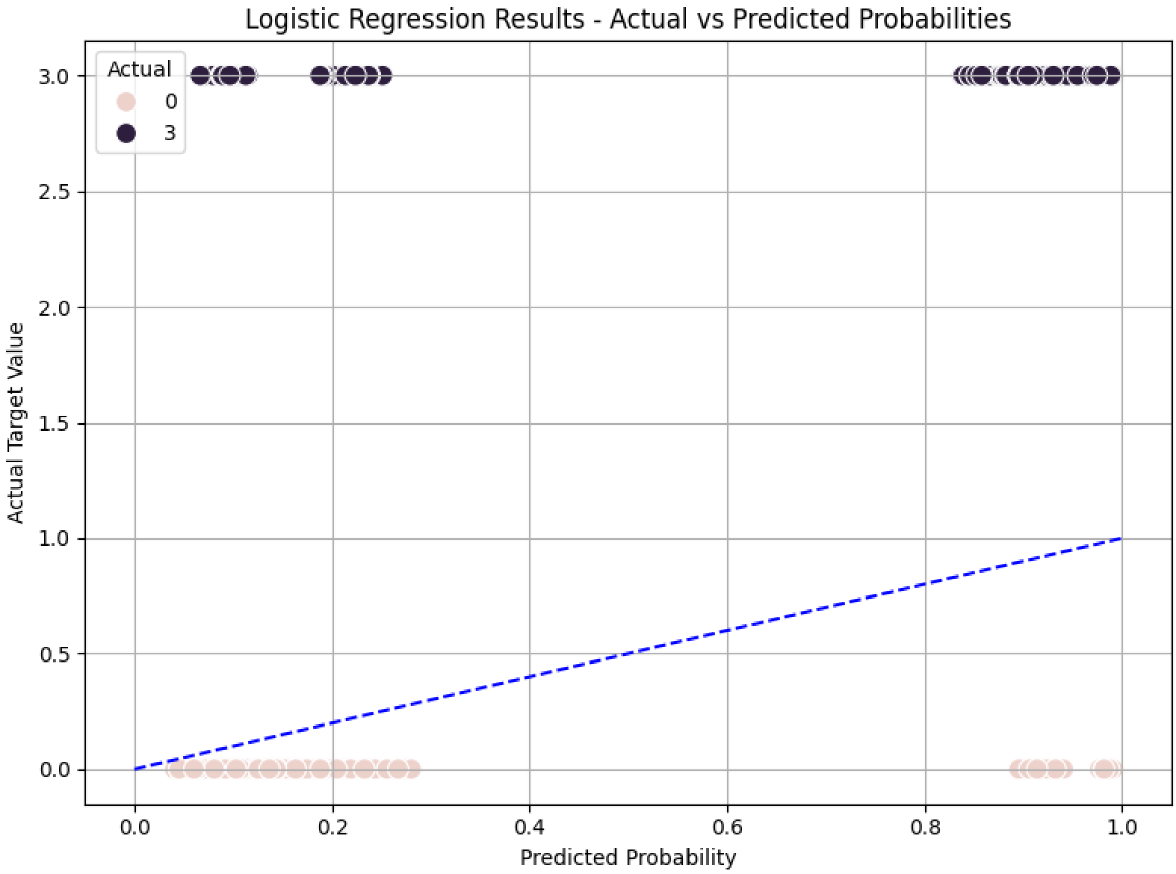
Accuracy: 0.9975201487910725

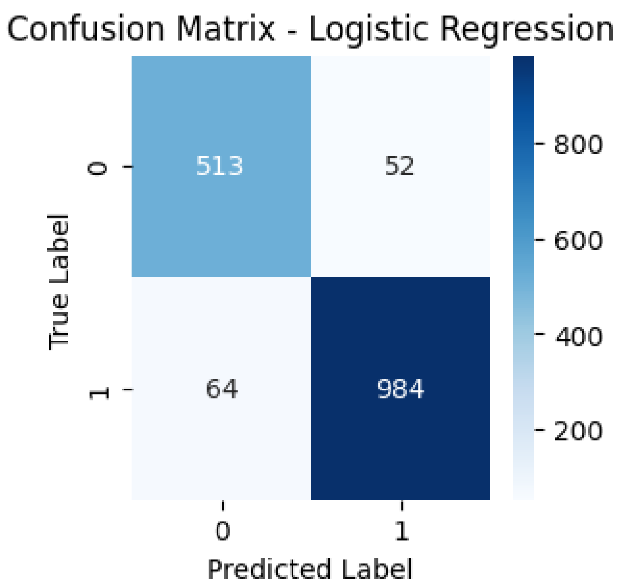
MSE: 0.02231866088034178

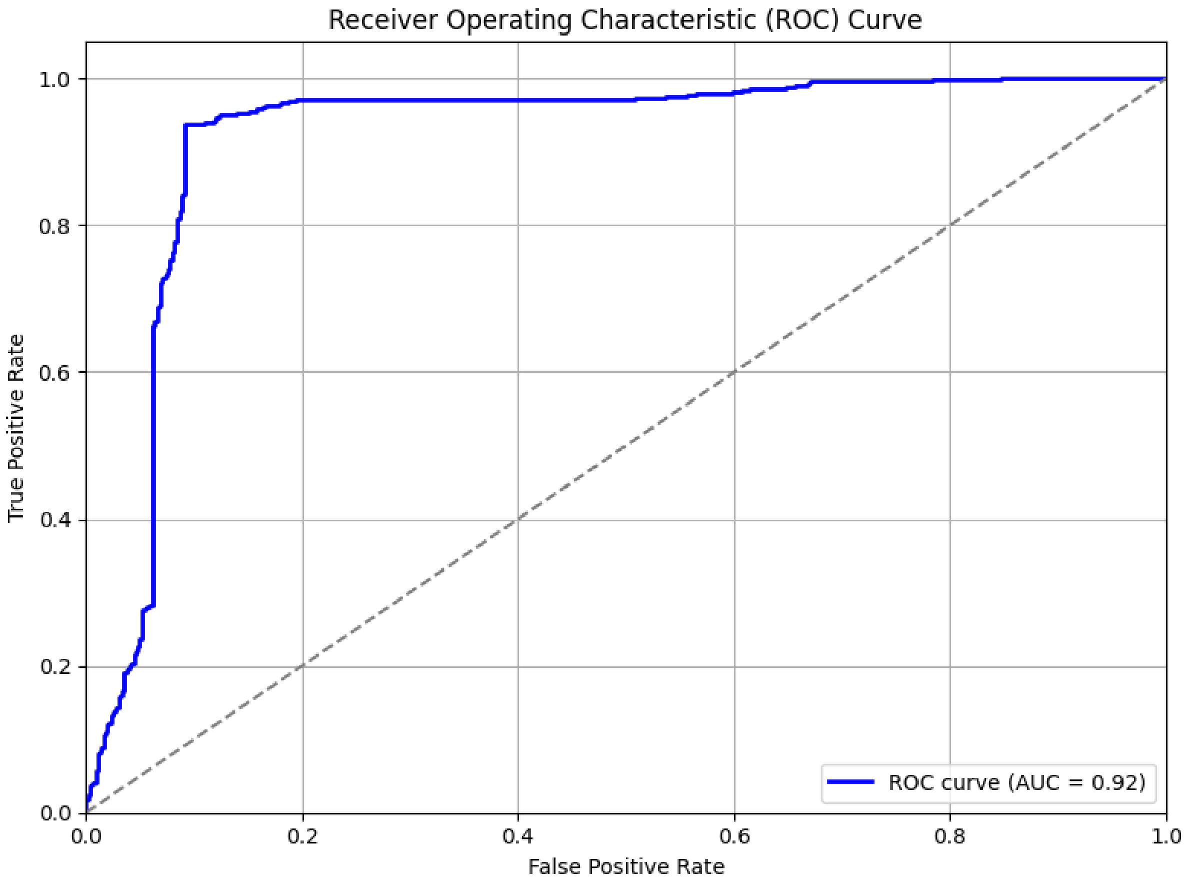
RMSE: 0.149394313413678

AUC: 0.9996816523677632

Logistic Regression







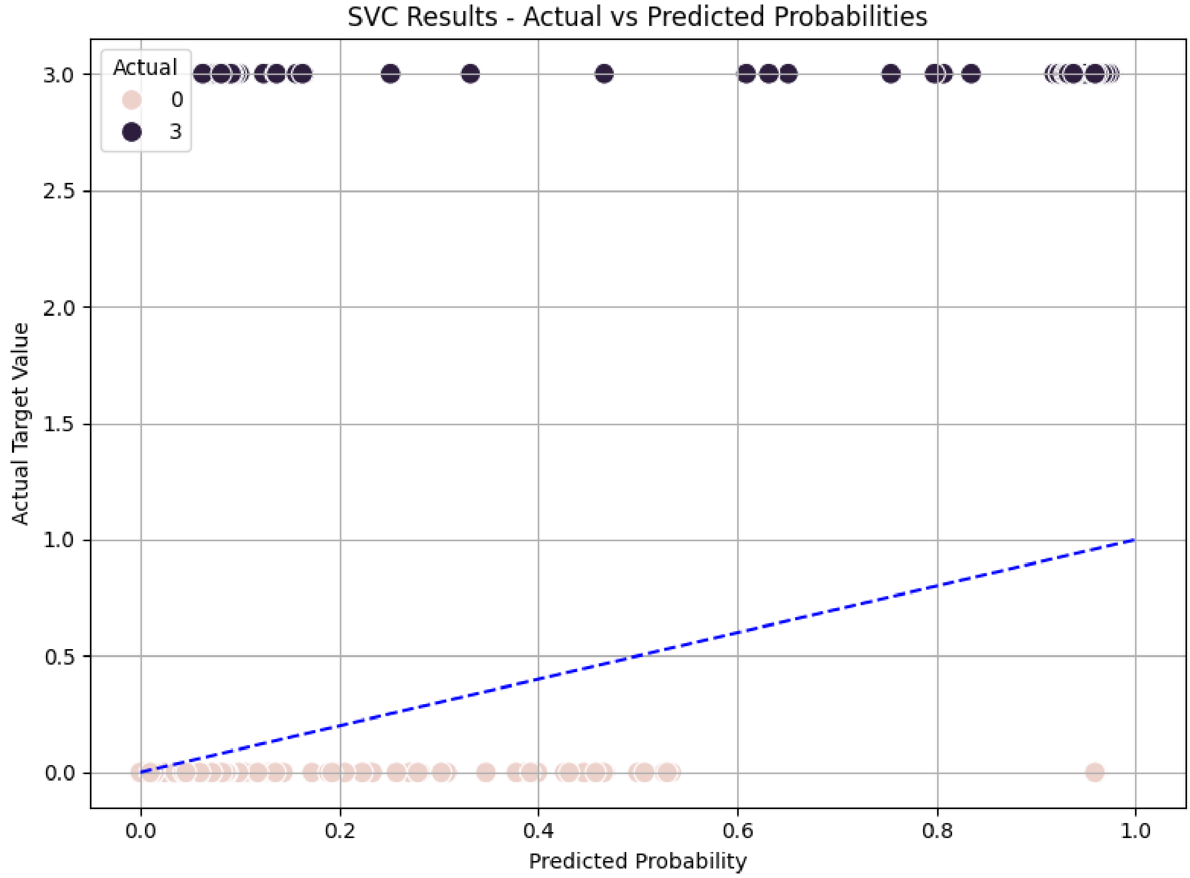
Accuracy: 0.9280843149411035

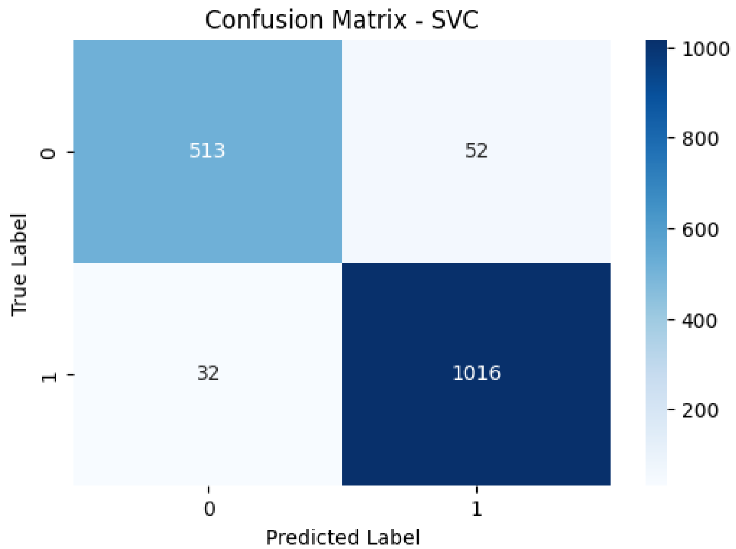
MSE: 0.6472411655300682

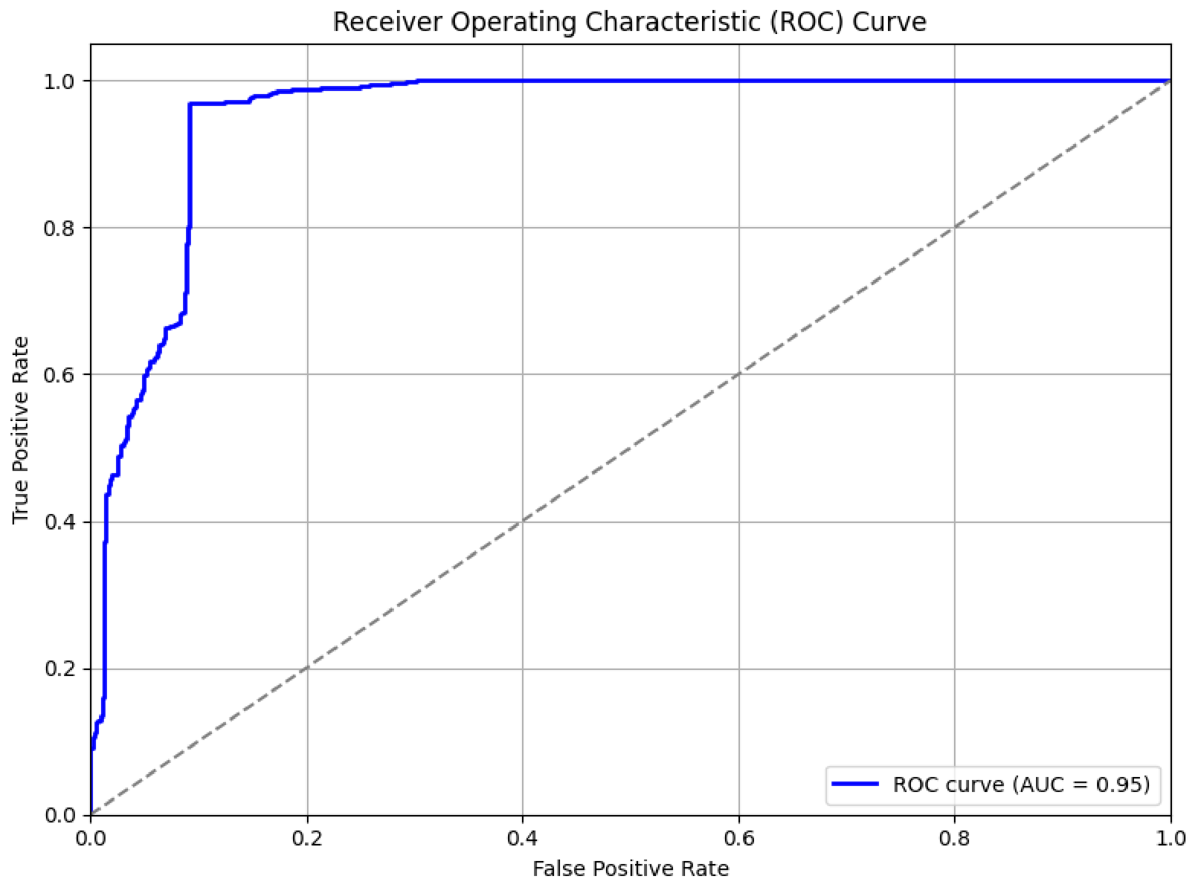
RMSE: 0.8045129989814137

AUC: 0.923260487738972

Support Vector Machines

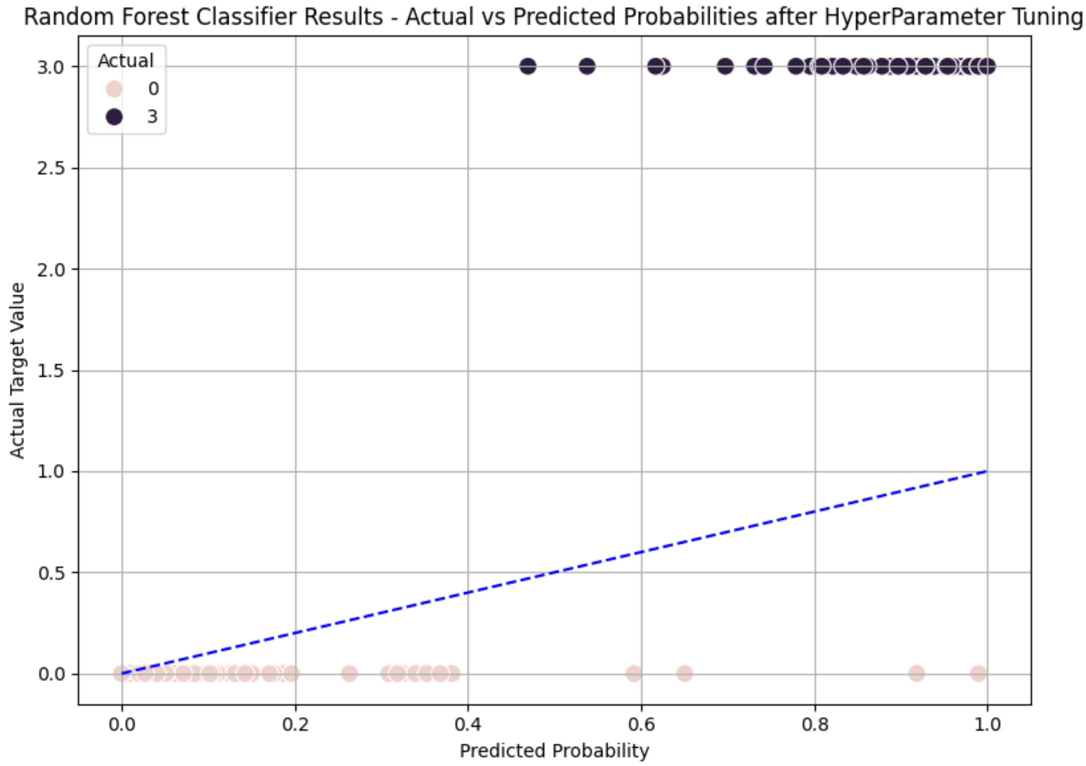


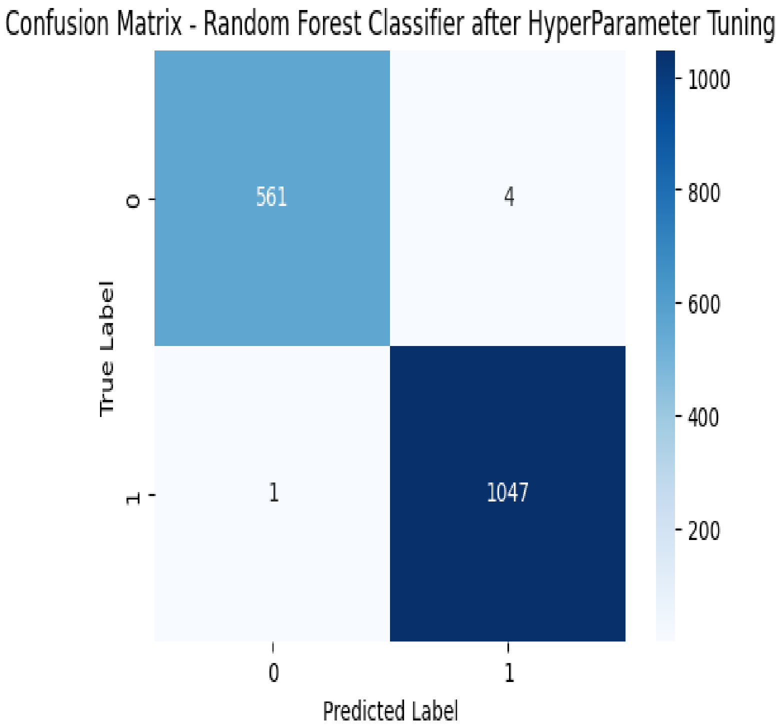


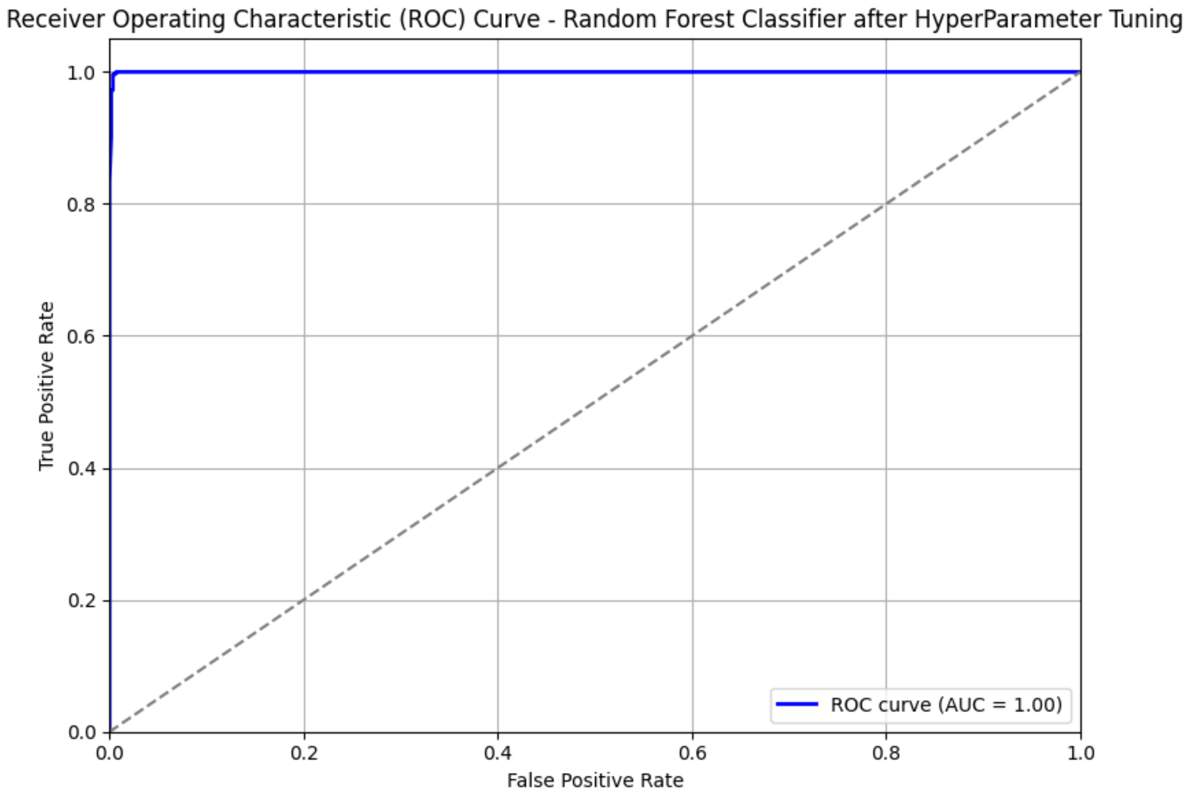


Accuracy: 0.9479231246125233
MSE: 0.4686918784879077
RMSE: 0.684610749614181
AUC: 0.953131220698507

Hyper Parameter Tuning



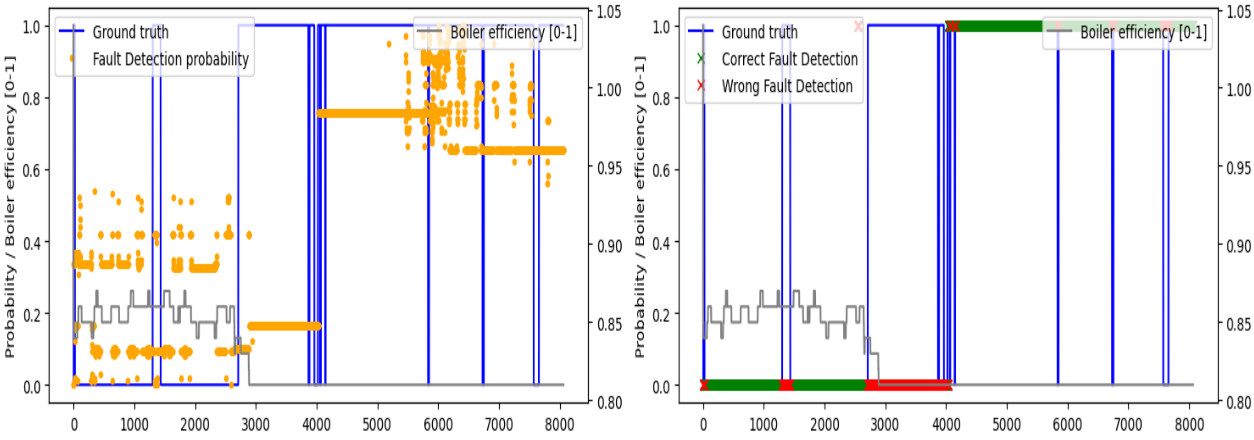




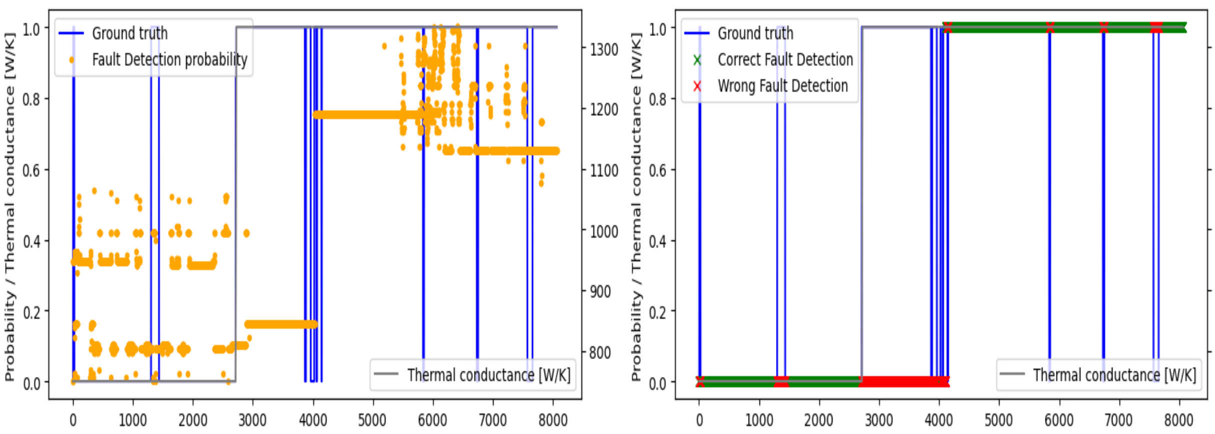
Accuracy: 0.9969001859888407
MSE: 0.0.27898326100433975
RMSE: 0.1670279201224573
AUC: 0.9997078294940215

Graphs

Detection of boiler efficiency drop



Increase of heat losses



Conclusion

In order to solve significant issues with energy efficiency and system dependability, the project focuses on creating an AI/ML model for anomaly detection in boiler performance inside District Heating and Cooling (DHC) systems. The model seeks to analyse operational data to identify deviations from typical behaviour, enabling timely intervention to prevent equipment failures and optimise energy usage. It does this by utilising open-source tools such as TensorFlow, Pandas, and Scikit-learn. The project's relevance stems from its potential to have an impact on several sectors. Specifically, it offers operators proactive maintenance, energy providers increased system efficiency, and building managers enhanced sustainability. The project is an example of a forward-thinking strategy to improve system performance and reliability in urban heating and cooling infrastructure by merging cutting-edge technologies with current systems. All things considered, it provides a scalable, effective, and affordable way to deal with practical issues and promote improvements in energy conservation and environmental sustainability.

Applications

The AI/ML model created for boiler performance anomaly detection will be implemented in District Heating and Cooling (DHC) systems in a real-world setting. It will keep an eye on boiler operations all the time, looking for departures from the standard that could point to irregularities. Through the examination of multiple operational factors including heat losses, efficiency declines, and temperature swings, the model will detect anomalies instantly, allowing for timely action to avert equipment malfunctions, maximise energy efficiency, and preserve system dependability.

Impact

There is a great deal of potential for the initiative to assist stakeholders in several sectors. Early anomaly detection can help boiler operators and maintenance staff by enabling them to proactively address problems, minimise downtime, and save maintenance expenses. By minimising energy waste and lessening their impact on the environment, energy providers and facility managers will increase sustainability, optimise resource allocation, and improve overall system performance. The project's true worth ultimately rests in its capacity to guarantee the continuous functioning of DHC systems, hence promoting energy efficiency, financial savings, and environmental sustainability in the infrastructure of urban heating and cooling.

References

https://www.intechopen.com/chapters/78400

https://www.britannica.com/technology/boiler

Auxiliaries

https://www.kaggle.com/datasets/mathieuvallee/ai-dhc-boiler-eta-hls
https://www.iea-dhc.org/the-research/annexes/annex-xiii/annex-xiii-project-03