OPTICHILL – DATA SCIENCE TOOL FOR PREDICTION OF CHILLER PLANT EFFICIENCY

PROJECT SPONSOR: Fred Woo, Optimum Energy

Theodore Cohen[1, 2, 3], Caitlin Parke[4], Maitri Uppaluri[4]

1. Department of Chemistry 2. Department of Material Science and Engineering 3. Molecular Engineering and Science Department 4. Department of Chemical Engineering

INTRODUCTION

MOTIVATION

Today, efficiency models of plants are run in large excel sheets. These models can take months to create and process. Machine learning can be a powerful tool to automate, streamline and optimize the prediction of the efficiency of a chiller plant.

The efficiency of the chiller plant is determined by **kW/ton.** A tool that determines which variable from the chiller plant has the highest effect on the efficiency is useful to understand and optimize the plant.

GOALS

1. DATA CLEANING

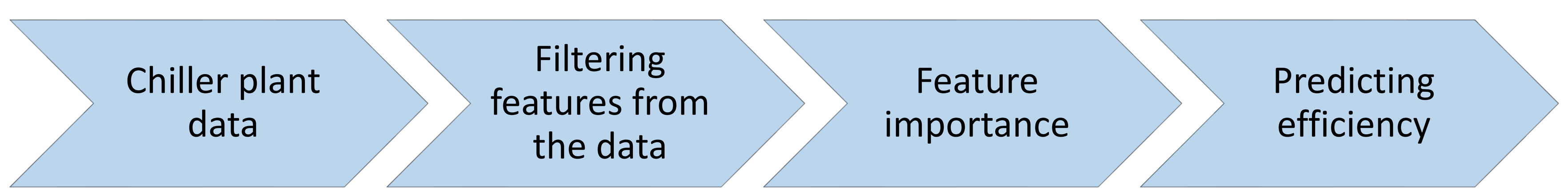
Remove unwanted features and alarms from the dataset.

2. FEATURE IMPORTANCE

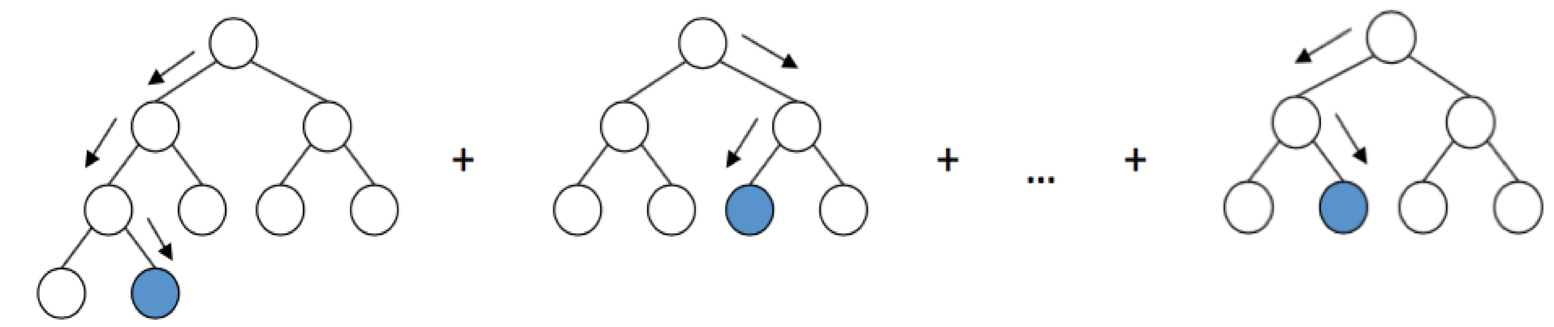
Create an algorithm that sorts out the features in the order of their importance for plant efficiency.

3. PREDICTION

To predict the efficiency of the chiller plant with machine learning.



GRADIENT BOOSTING MACHINES



Gradient boosting machines (GBM) incorporate decision trees in an additive manner with a gradient descent procedure to minimize the chosen loss function. The decision trees are called weak learners because each tree is capped at a certain depth and does not adequately describe the model well. Subsequent trees are trained to reduce the loss in the direction of the gradient and are added to the model to improve the prediction of the model, while previous trees are left unchanged. Generally, a hyperparameter, like tree depth or the minimum samples for a split, are constrained to maintain weak learners but increase the ability to tune the model; in this work, the tree depth and the number of trees were constrained. In this work, the gradient boosting machines were implemented with the *scikit-learn* package.

FUTURE WORK

- To test the Gradient Boosting model by training on a chiller plant and then using that model to predict efficiency for a different plant with a similar configuration.
- Adding and removing different features to see the effect on prediction across each feature and how the importance of these features changes with respect to the model predictions.
- Changing the control type used on a particular plant and gauging its impact on efficiency and feature importance.

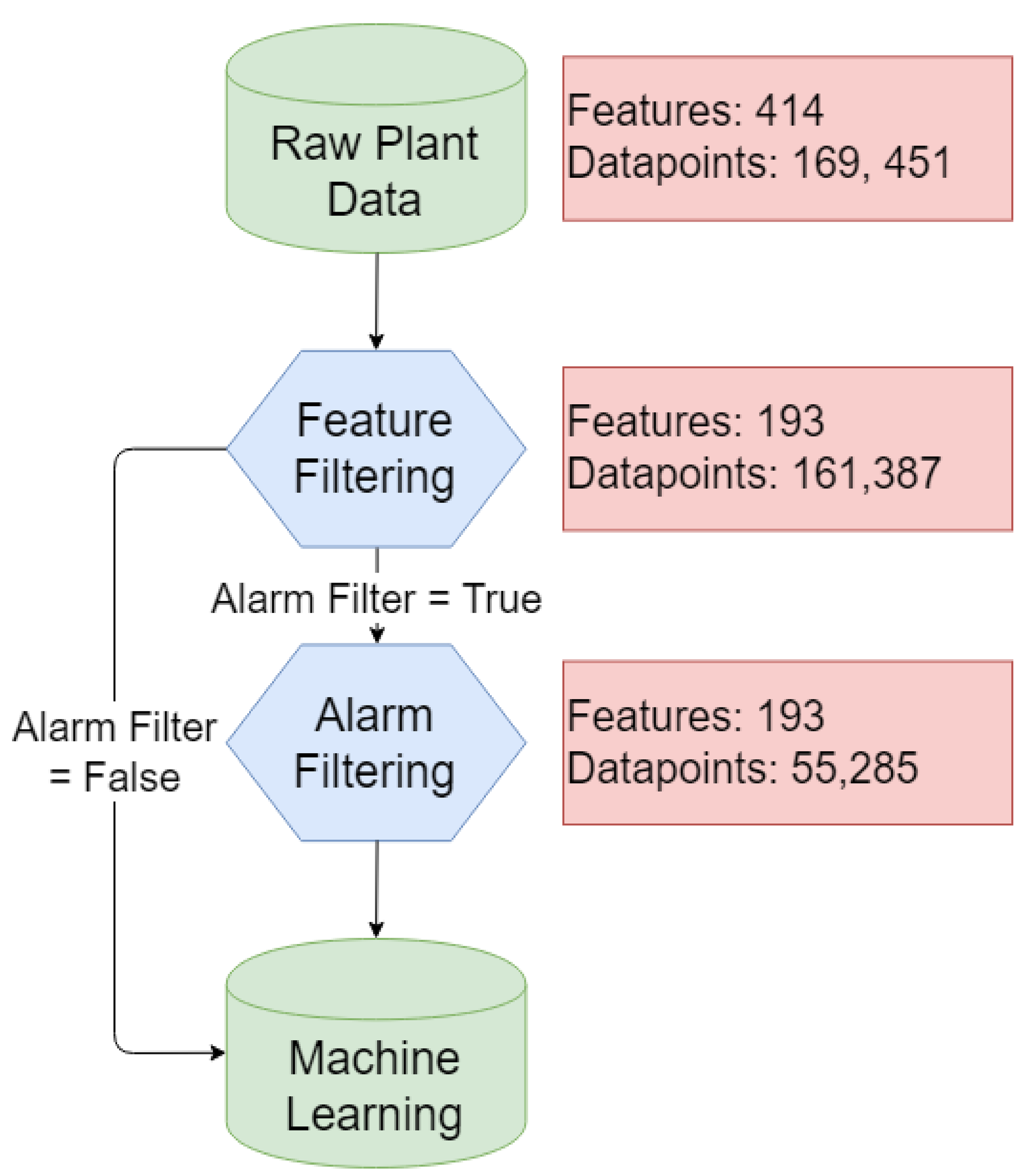
DATA CLEANING

FEATURES

Raw data from several chiller plants was filtered and optimized for machine learning implementation. This was started by only using the raw data directly from the plant. These data included readouts from the Building Automation System, chillers, cooling towers, and water pumps.

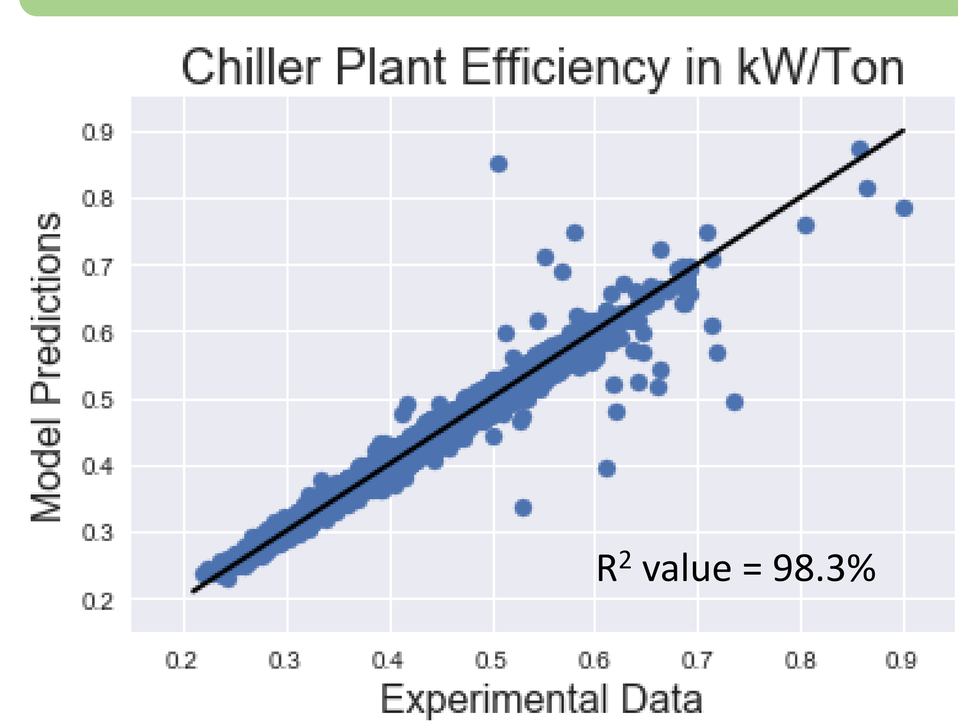
FILTERING

Various stale features were removed and any feature containing the phrase 'kW' was removed to ensure that no calculated data was used for plant performance prediction. Datapoints used for prediction were pre-optimized, complete, and in some cases had no alarms going off. Optional inputs for removing any feature were added for flexibility.



RESULTS

OVERALL PREDICTION



Feature Name	Feature Importance (%)
Chiller 5 Compressor Refrigerant Discharge Temperature (°F)	100
Maximum Chiller Pressure Lift (psi)	48.5
Minimum Chiller Pressure Lift (psi)	33.4
Chiller 5 Condenser Approach (°F)	21.5
Cooling Tower 5 Fan Speed (Hz)	20.2

With filtered data from Plant 1, a GBM model was trained and tested, and the parity plot is shown above. Additionally, a list of normalized feature importance was made by calculating the number of times a split was made on a particular feature. This tells the user which features have the largest contribution to plant efficiency. From this work, Optimum Energy has been able to explore its data more effectively; this insight will allow for collection of more representative data. For example, the parameters of a certain chiller continue to be in the top features due to its continuous use. In the future, more directed cycling of different machines can be implemented to better sample the plant feature space and may lead to improved optimization results for efficiency in the future.

TRAINING ACROSS SEASONS

For further exploration of the data, the model was trained on different seasons and then tested on the rest of the data. The transitional seasons, spring and fall, have the best testing results. These results were as expected as these seasons have more varied weather conditions, which make them better candidates for training the model.

Season	R^2 value (%)
Winter 2016-17	74
Spring 2017	76
Summer 2017	60
Fall 2017	81
Winter 2017-18	70
Spring 2018	79

DIRECT
Data Intensive Research Enabling Clean Technologies

OPTIMUMENERGY

CLEAN ENERGY INSTITUTE
UNIVERSITY of WASHINGTON

NSF

UW MEM·C

ACKNOWLEDGEMENTS

Fred Woo, Tim Wehage, Michael Huguenard, Dana Lindquist (Optimum Energy)
Professor Dave Beck and Kelly Thornton

Github URL:
https://github.com/optichill/optichill