AIDI 1002 Project

Business Understanding & Problem Discovery

Statement of Work V1

Gavin Rowsell - (100792801)

Date: November 6th, 2020

Business problem

Undershoot and overshoot are common problems when working with control systems. For a large greenhouse, the temperature inside is critical to the growth and yield of the product. The control system is given a range of temperatures and will issue instructions to various systems (Ventilation, Insulation curtains, Circulation fans, etc.) in order to heat or cool the greenhouse. Upon receiving a measurement outside the desired temperature range, the system will respond by shutting off heaters or closing vents to move the temperature back within the desired range. The problem is these actions are not instantaneous, the vents take time to open or close and the heaters do not immediately start or stop emitting heat. This leads to over or under shoot of the target temperature which can disrupt the integrity of the crop and is a source of wasted energy. Current solutions are to reduce the range of acceptable temperatures given to the control system so that when over or undershoot occurs the temperature will still be within the actual desired range; however, this leads to over-cycling and an increased utility cost.

Rational statement

A machine learning model can be used to keep the temperature of the greenhouse within the desired range more consistently. By using measurements of humidity, wind speed, time of day, month of the year (amount of daylight), and outside temperature the algorithm will be able to predict the point at which the temperature inside the greenhouse will move outside the desired range. The model will also account for the affect of the weather conditions slowing or speeding up the rate of temperature change due to a response (opening vents, rotating fans, moving curtains, etc.). While the system already is designed to react to temperature changes, the actions of the system change the temperature within the greenhouse at varying rates depending on the previously mentioned factors (humidity, wind speed, time of day etc.). As an example, on a windy day opening the vents will cool the greenhouse faster than if there was no wind. A machine learning model can use weather measurements to more accurately determine when and for how long a cooling or heating command should be issued. Since we are attempting to predict the future temperature of the greenhouse according to various features, linear regression is the method that will be used.

Data Requirements & Sources

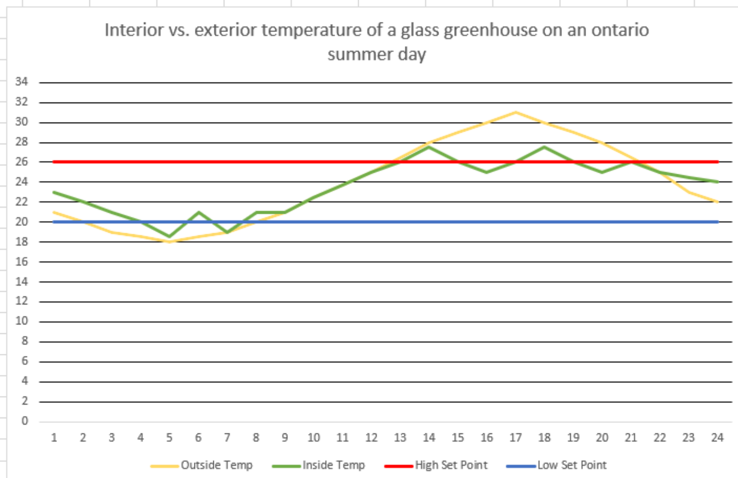
Data will be created with the help of Mitchel Grant, a colleague of mine whom is an assistant project estimator and designer at Thermo Energy Systems Inc. Historical weather data for Ontario, which is widely available on the web, will be used as a base and (with insights from industry) can be augmented with datapoints describing how a control system would react and affect the environment simulated by the collected weather data.

In the example below, we can see how the temperature of the greenhouse temporarily leaves the desired range, again this is due to the lag of the system performing cooling or heating actions. This lag will vary depending on the weather conditions at that time.

<u>Example Data:</u>

Hour	Outside Temp	Inside Temp	Low Set Point	High Set Point
1	21	23	20	26
2	20	22	20	26
3	19	21	20	26
4	18.5	20	20	26
5	18	18.5	20	26
6	18.5	21	20	26
7	19	19	20	26
8	20	21	20	26
9	21	21	20	26
10	22.5	22.5	20	26
11	23.75	23.75	20	26
12	25	25	20	26
13	26.5	26	20	26
14	28	27.5	20	26
15	29	26	20	26
16	30	25	20	26
17	31	26	20	26
18	30	27.5	20	26
19	29	26	20	26
20	28	25	20	26
21	26.5	26	20	26
22	25	25	20	26
23	23	24.5	20	26
24	22	24	20	26

Tomatoes optimal between 20 and 26 degrees celsius



Data Limitations, Assumptions & Constraints

For this problem we are assuming that we already know the actions that need to be taken (open vents, start boiler, start fans, etc.) what we are interested in is at what point should that instruction be given and for how long since there is a varying interval of time before that action begins to affect the temperature and when it stops affecting it. Different materials are used to construct greenhouses and these materials have different heat transfer coefficients which will affect the rate of change of the temperature inside, for this project it will be assumed that the material in use is glass. Another assumption is that exterior temperatures will be within a range that the system has been designed for (no extreme heat or cold). Datapoints are listed per hour, not all measurements provided to the control system are recorded and available for use. Typical control systems receive new measurements every 30 seconds however we will be looking at hourly intervals.

Test Process

Model accuracy can be determined by the number of measurements within the desired temperature range divided by the total number of temperature measurements within the greenhouse. More useful for a regression model however, is a measurement of the bias and variance of the model. If the model is measured to have high variance or bias it will not be performing optimally and can be an indication of overfitting or underfitting. By using validation curves, the optimal ratio of variance and bias can be found. To show the effectiveness of the model, different crops requiring different temperature ranges can be tested using the gathered weather data.