**Already assessed:**

1. Imbalanced random forest classifier for 2 different turbines
   * T07
   * T11

**Options for further assessment:**

1. Additional turbines
   1. Probably not a value add. Not getting viable analysis results
2. Different ML models
3. Focus on the high temperature faults for T07
4. Remove the feature importances that have no measurable contribution
5. Remove feature importances that are < 0.005 because they are negligible and may increase the noise in the model
6. Import the failure data into the sensor data and take an average of the sensor values before and after the time point associated with the failure time to create a set of sensor data points for those failures.

**Lessons learned so far:**

1. The fault data only contributes a binary output of “failed or not failed.”
2. The faults that happen most often also appear highest on the list of feature importances.
3. Using the Bins technique allows predictive modeling for time series data (ie next bin failures) but because using “next bin” predicts a bin that doesn’t exist, there are many false positives
4. There is very little significant failure data so creating an accurate model is difficult
5. Other types of ML algorithms that focus on timeseries data should be investigated if there was more time.
6. T07 has a very strong correlation for the Hot Transformer minor fault