In this document, I will record some notes both on the status of the project in general, and more specifically on the final code produced.

Why RuLSIF?

As a group we tried many different methods for trying to find anomalies in the datasets. We tried a variety of models, some specifically designed for time series, and some not specifically designed for time series. We quickly found out that models not designed specifically for time series are not likely to work on this dataset; the behavior changes too much during the melt.

From observing the sensor data in the case of the known anomaly, it seems that the sensors change behavior drastically during the anomaly. This gives hope that detecting the anomaly might be reasonable to do, and it indicates why RuLSIF might be a good choice for this problem. RuLSIF is designed to detect change points, places in the data where the behavior changes. In our experiments, RuLSIF strongly detects the changes in behavior in the dataset, while not detecting the slow changes in behavior that are normal within the melt. In addition, it seemed to perform the best at marking points we believed to be anomalous, while not marking the points we do not believe are anomalous. Unfortunately, due to a lack of data we cannot confirm this model generalizes well, but given what data we have we believe that RuLSIF makes sense for this problem.

What does the final code do?

The final version of the code performs a few things. Firstly, it grabs all of the data from 150 channels; the selection of the channels is not part of any particular design, but it just takes every 4th to get a good spread of sensors, starting from channel 100 to exclude non-sensor channels. Then, this data goes through a hierarchical agglomeration step, which converts the 150 channels into 6 new channels, designed to preserve information from the original channels while reducing dimensionality. Finally, the RuLSIF algorithm is ran on these new channels, which generates data on the change points detected. A high score from the RuLSIF algorithm indicates the distribution of values is changing quickly, which from our observations we believe correlates well with a change in behavior in the furnace. We also note that there are no significant change points found on datasets that do not have anomalies, so the change point score does seem to distinguish between anomalous and non-anomalous behaviors well. Note that the algorithm provides data on whether or not each point is anomalous, not on whether their was an anomaly in the melt as whole; doing this would require more testing data then was available to produce good results.