

## Issues found:

1. Impute does not create CSession ids for new locations, which is necessary for tying tests to limits, and therefore cannot be run before agg tests
2. New wafer shapes are difficult for model to intake

## Suggestions/Options

1. Allow pipeline to run agg tests and input in either order
  - a. We impute all tests then aggregate
  - b. We aggregate all tests then impute to create full wafer

## Future work:

Throughout our explorations into this project we came across a few ideas that merited further exploration.

### **Aggregate Tests**

With aggregate tests, we used them for some training runs and compared them to outputted wafer maps. We also thought it may be beneficial to weigh the test failures even more by using it as a main feature in the training dataset.

### **Feature reduction**

We decided to try and reduce the feature set to a more manageable size by averaging correlated test columns. There could be tuning done here to figure out the best representative feature set, like PCA, or simply chucking more data into it.

### **Cross Product comparison**

It could be helpful to compare outputs from different products, or contain multiple products in a training set. As different products are not the same size, padding or rescaling may be required. The other option may be to try and assign names to clusters, such as classifying clusters as high null, or ring patterned, or so on. This could help in comparing relative size of clusters in a dataset.

### **Clustering**

We tried a multitude of clustering algorithms, at different stages in the process. We settled for KMeans for most things, but are using a force based network graph to create the final cluster visualization from the similarity score. There may be more rigorous options for clustering.

### **Continued Classification**

Once you have model weights, it is possible to ask the model to classify a single wafer map compared to the ones it has been trained on. We have not explored this area deeply.