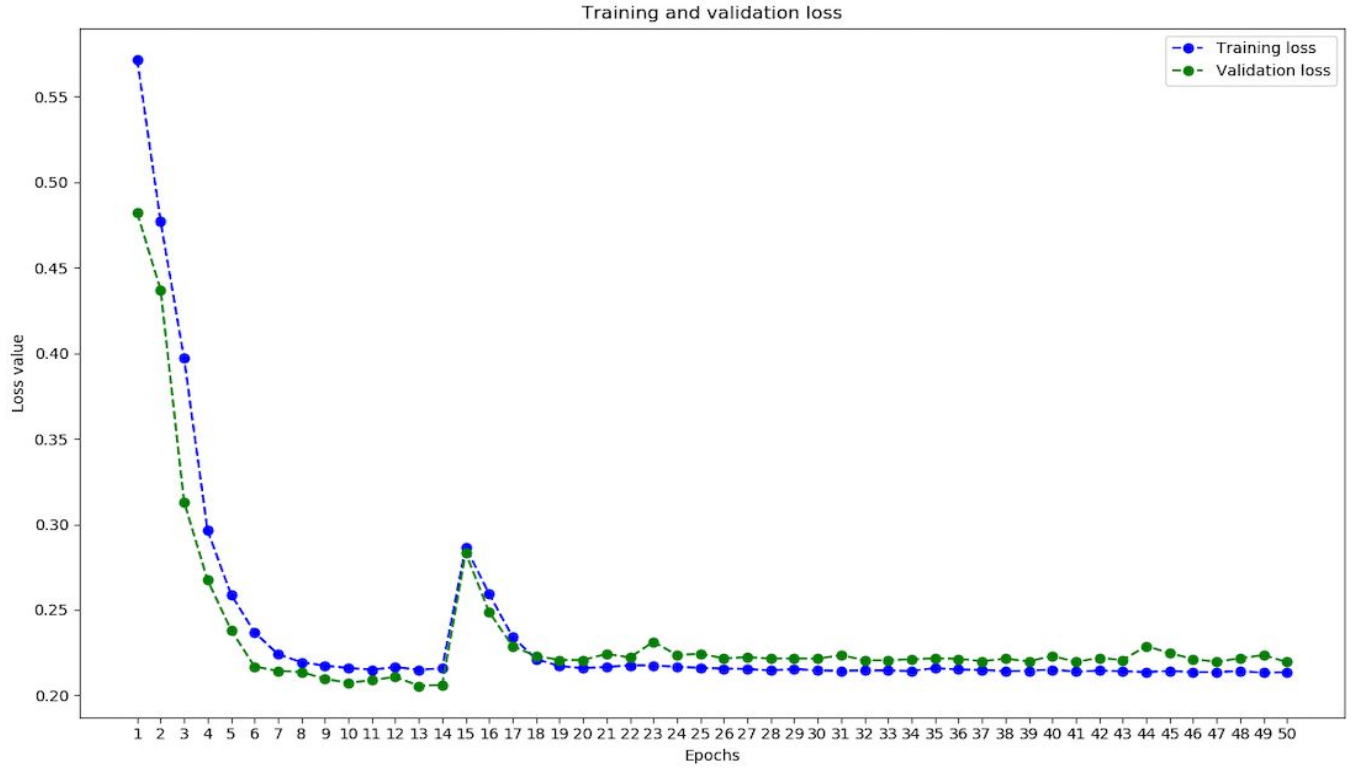
Exploring estimated features from Autoencoders

Autoencoder Loss value Plot

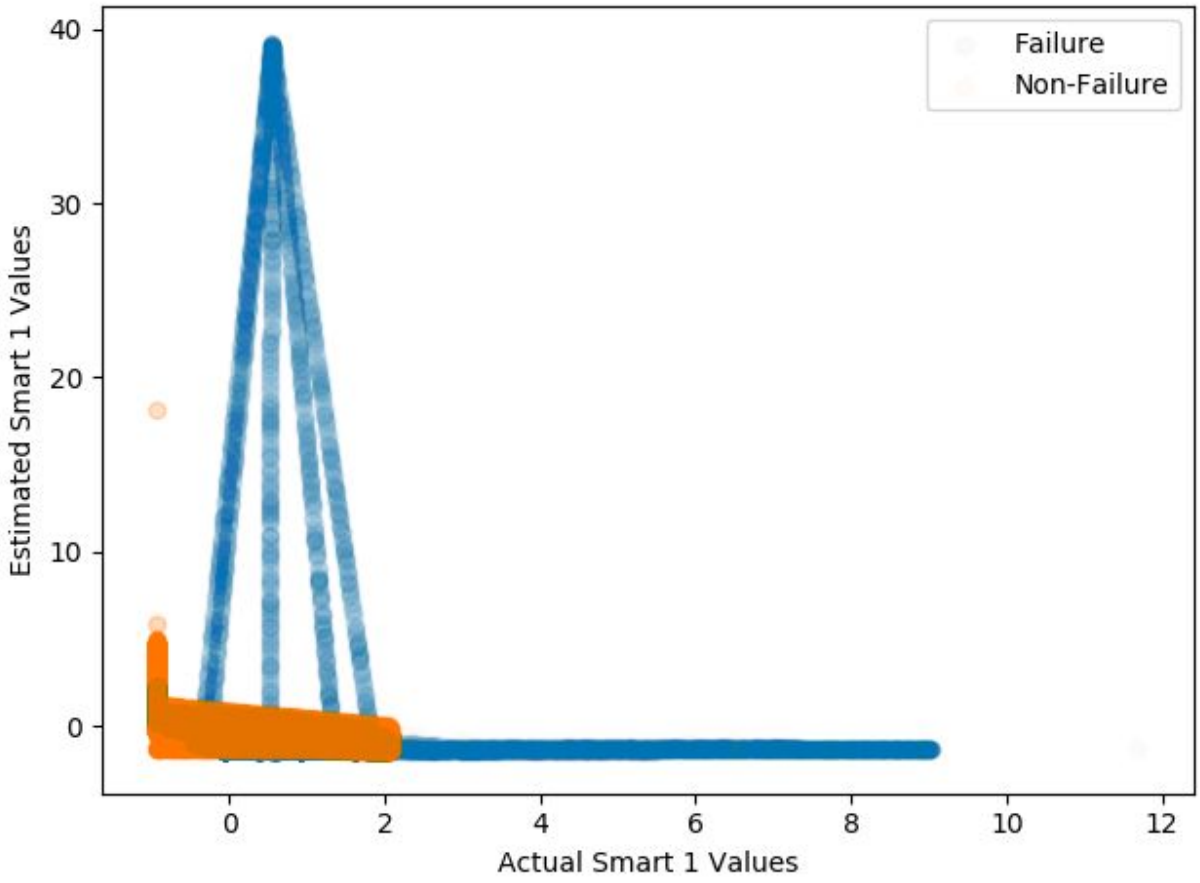
Plot after cross validation over autoencoder, we can see initially validation error is even lesser than training loss but as epoch value increases the autoencoder learns better and hence validation loss becomes a bit more than training loss.
The average loss for training data is - 0.3164
The average loss for validation data is - 0.3128



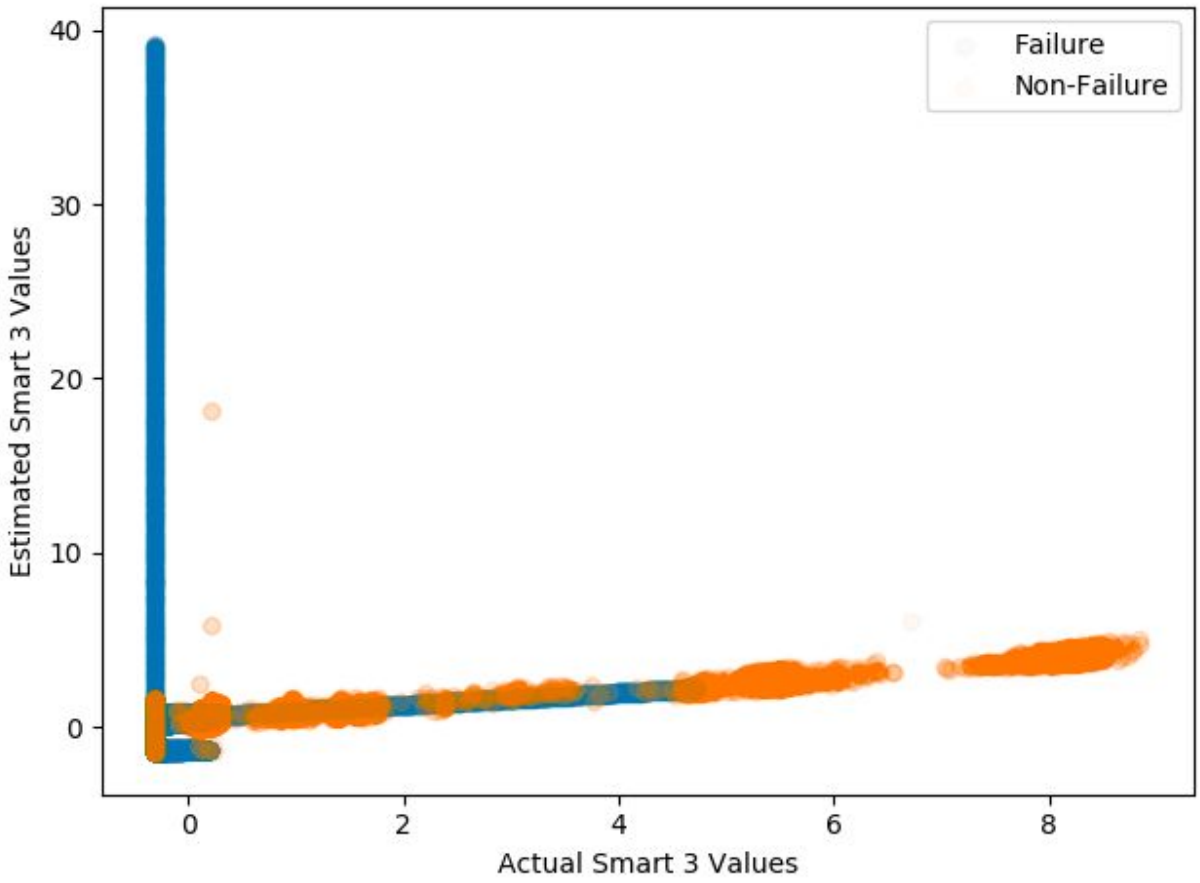
Exploring estimated features from Autoencoders

Estimated Versus Actual Feature Plots

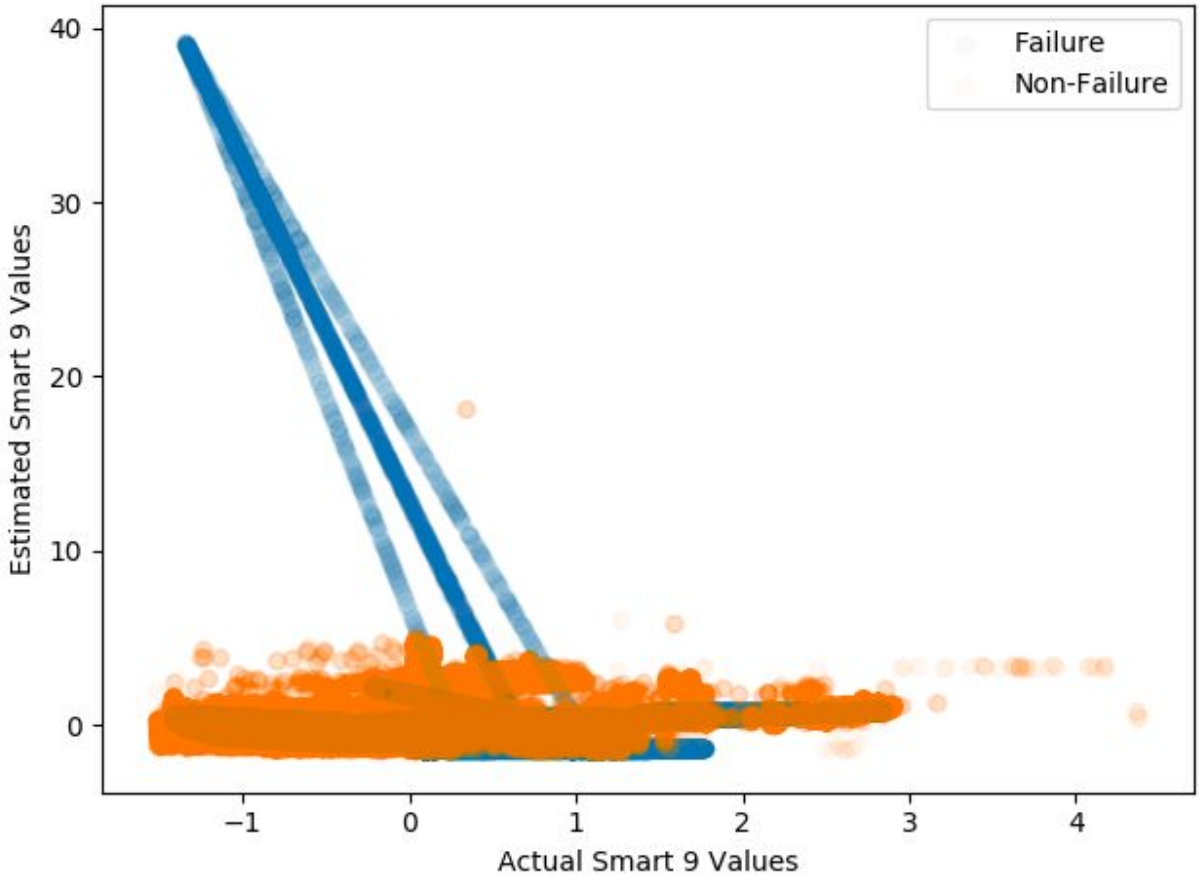
The plots for input and output features estimated by trained autoencoder for new data, used 1 million, 400 data points obtained after performing oversampling using SMOTE



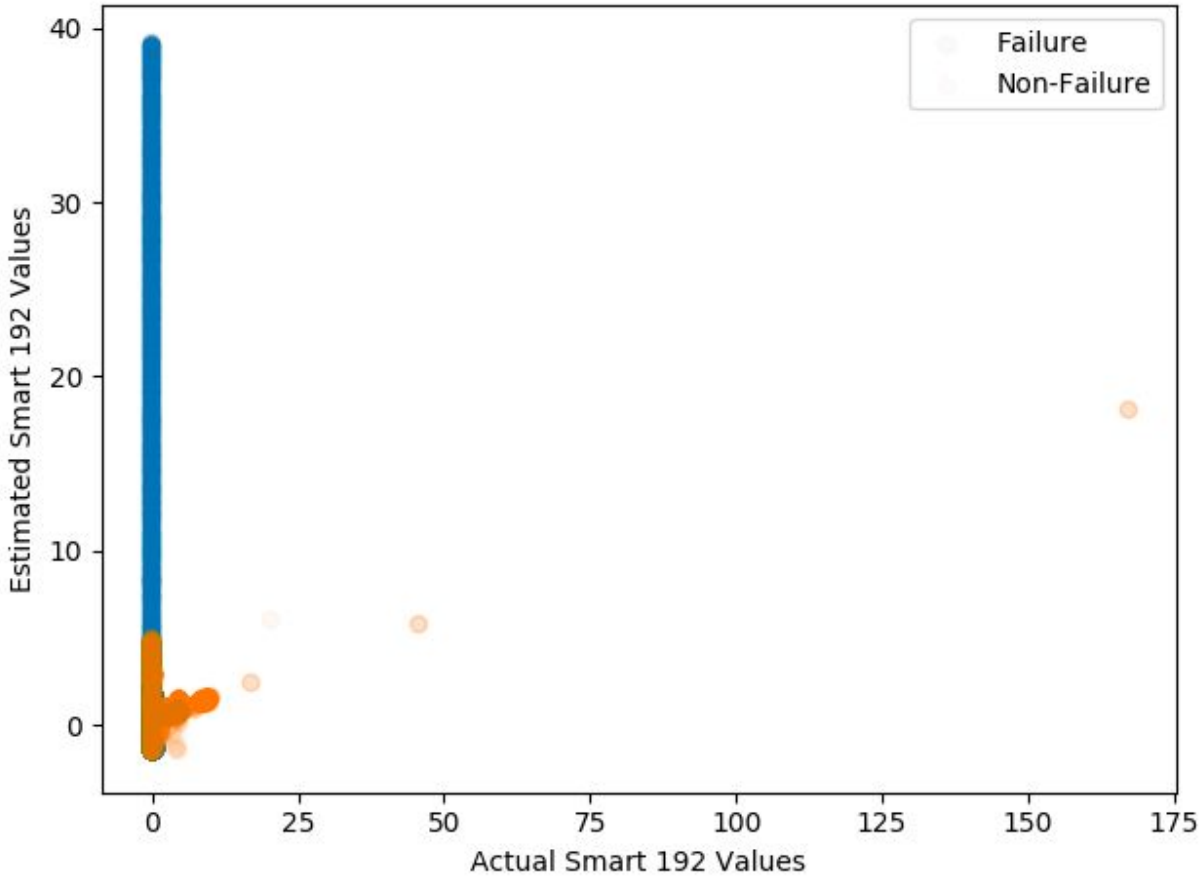
Exploring estimated features from Autoencoders



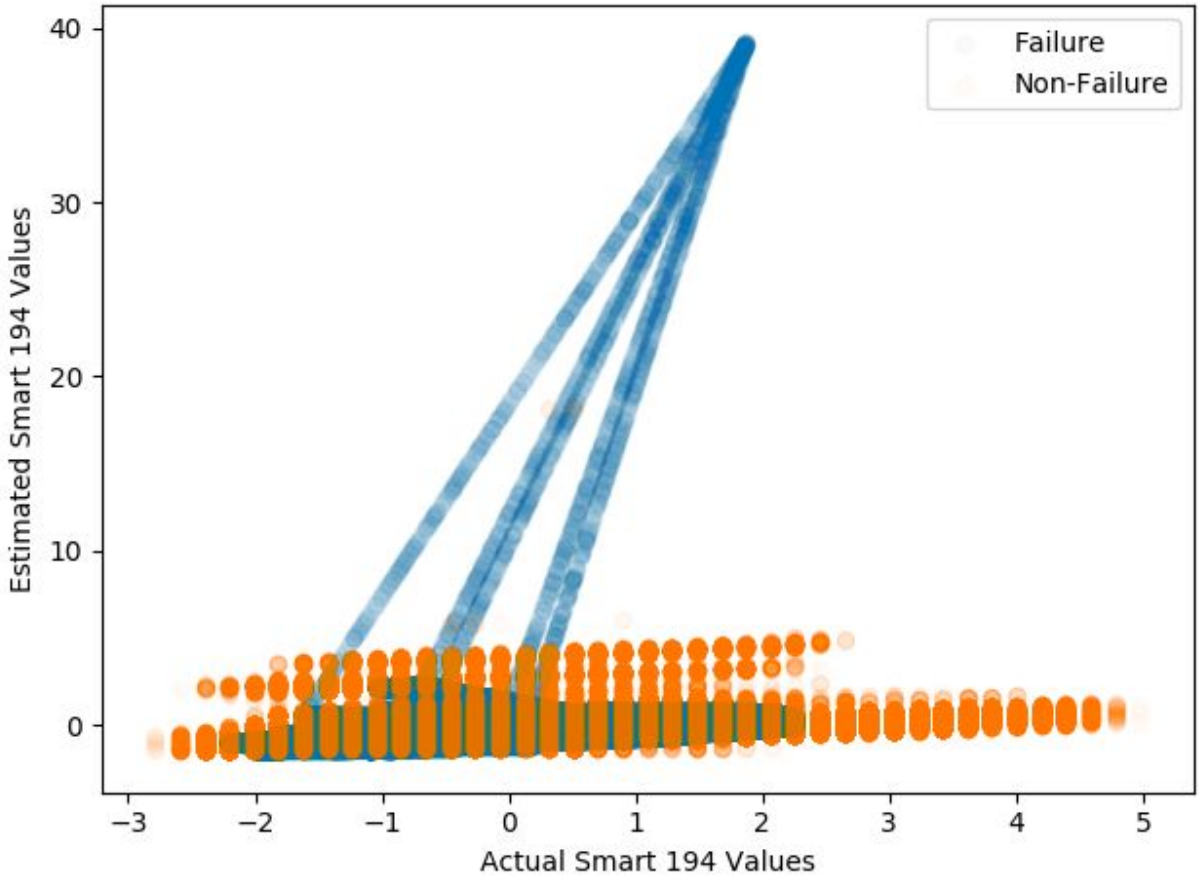
Exploring estimated features from Autoencoders



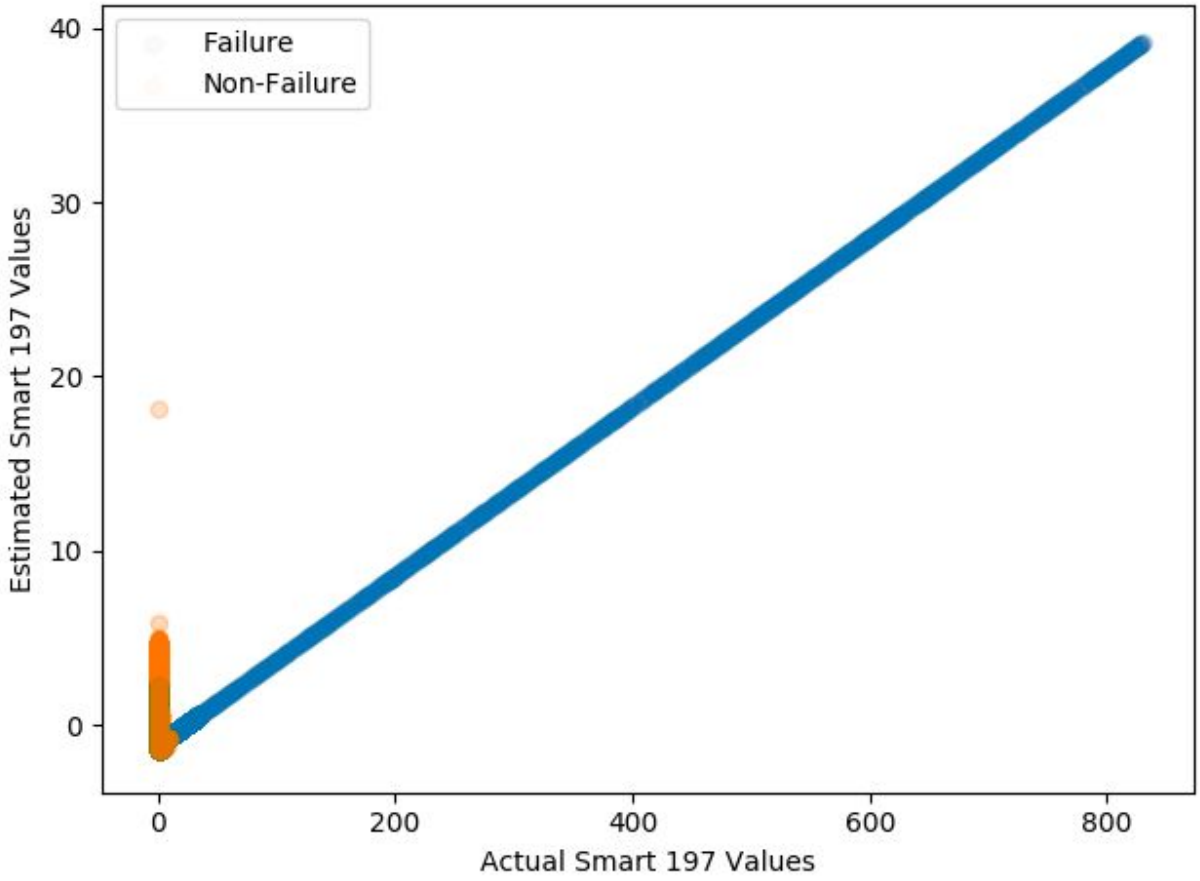
Exploring estimated features from Autoencoders



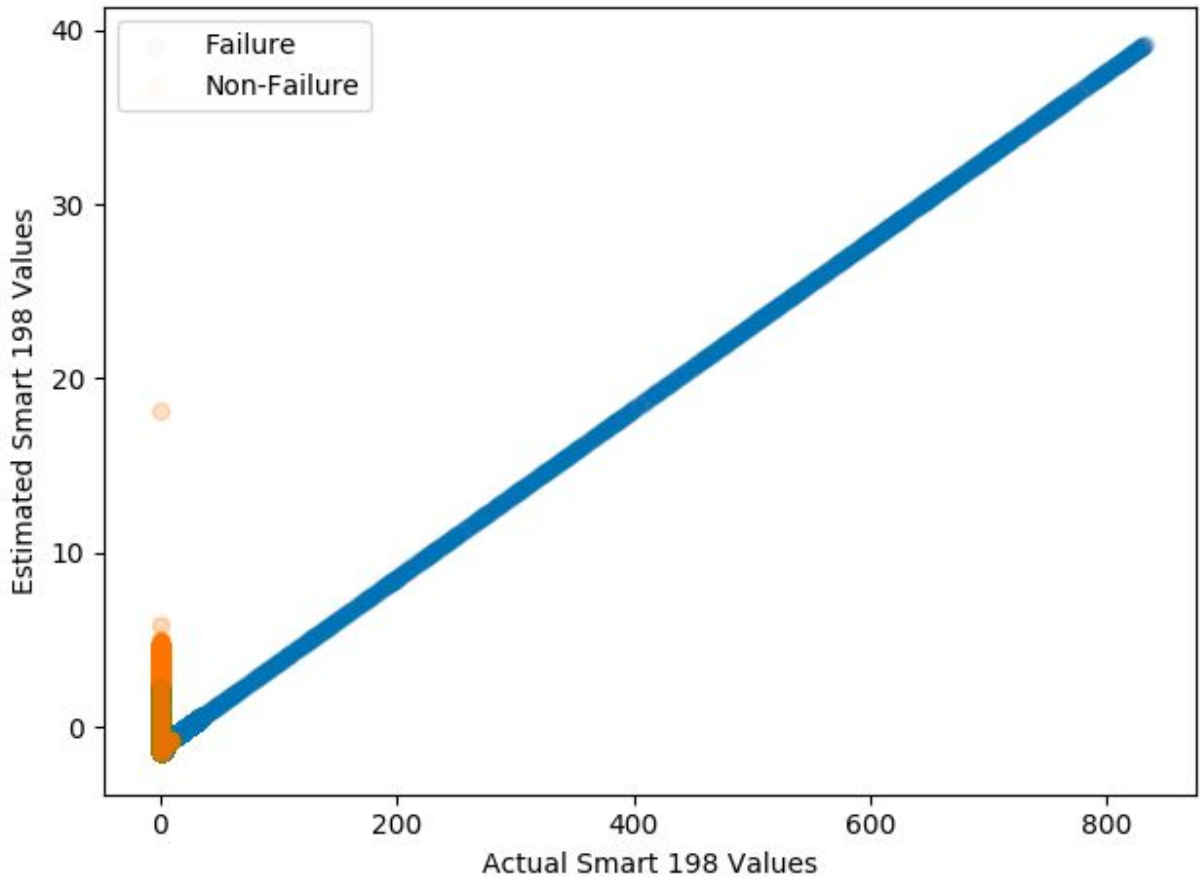
Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders



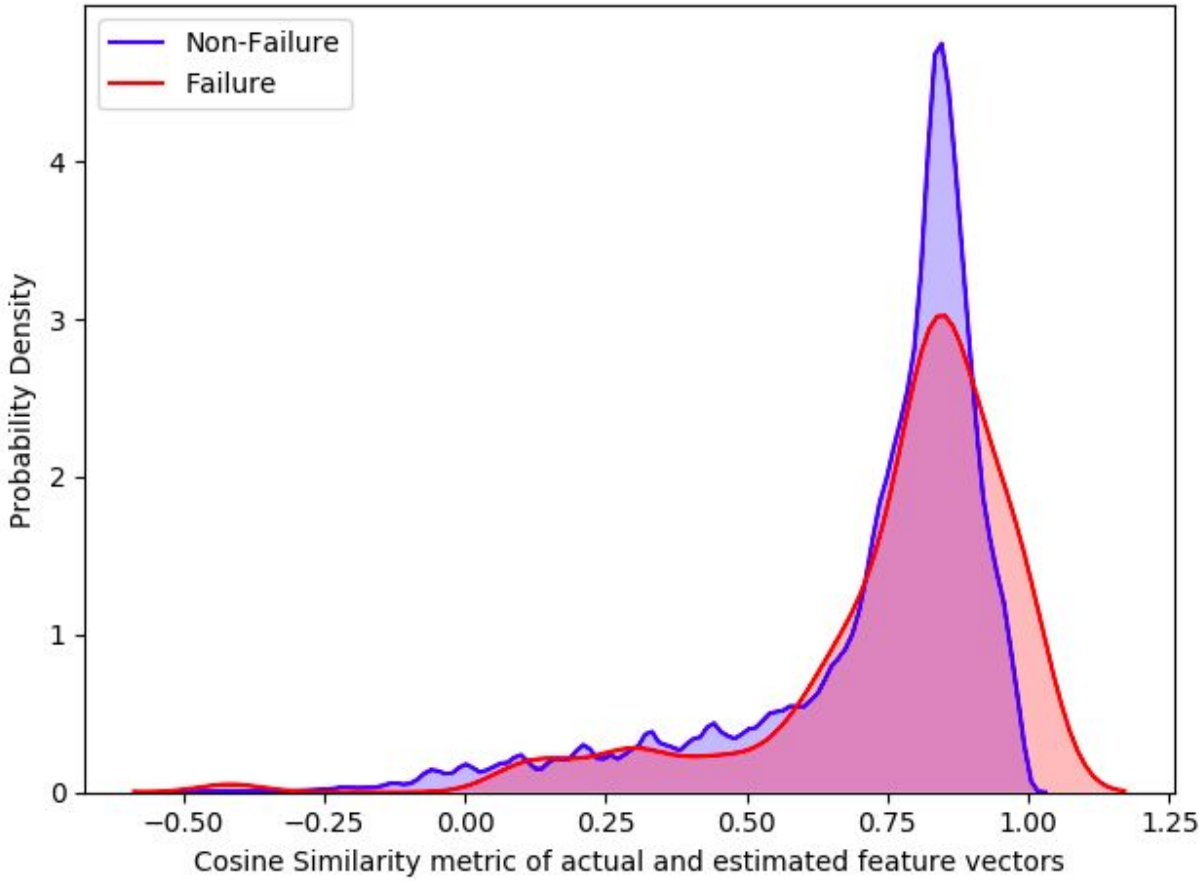
Exploring estimated features from Autoencoders



Cosine Similarity Plot

Density plot for COSINE SIMILARITY of estimated and actual features

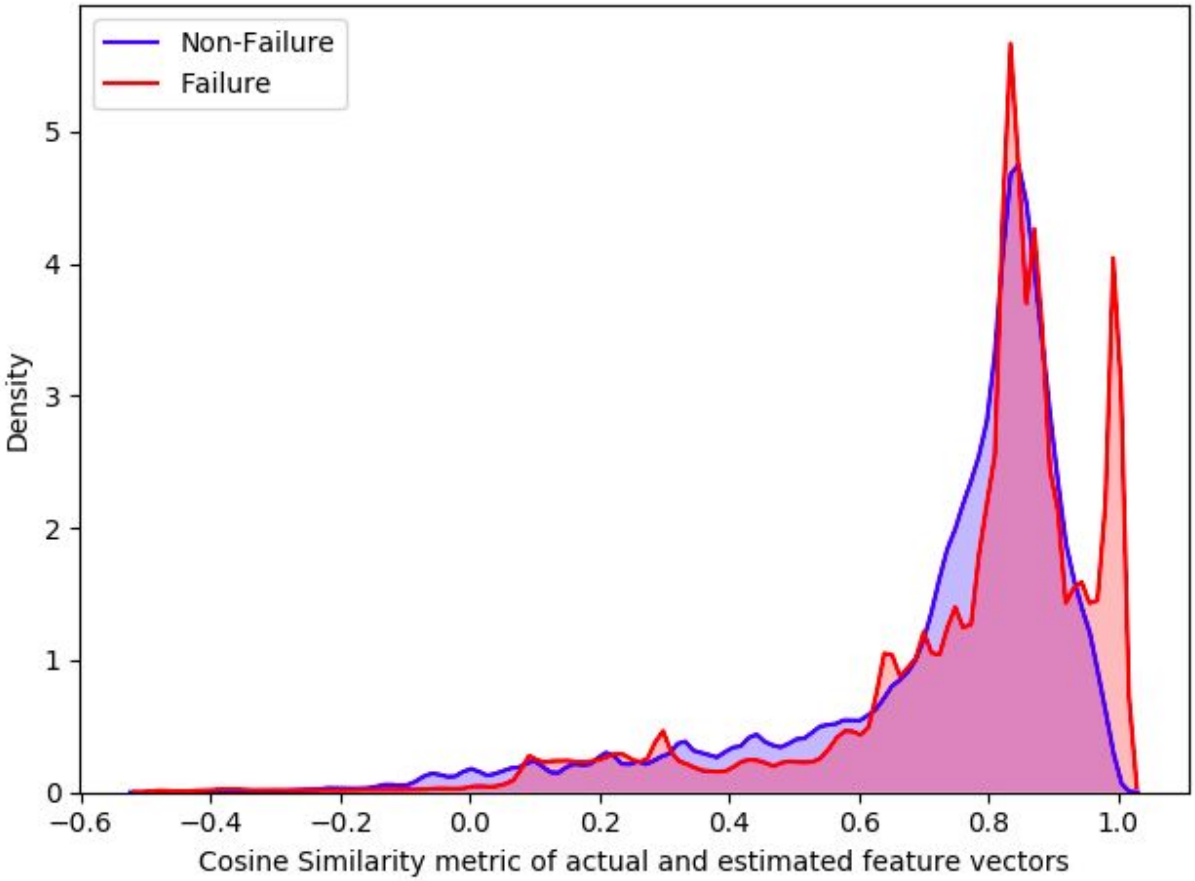
Exploring estimated features from Autoencoders



Cosine Sim Plot after Oversampling

After SMOTE oversampling and then plotting for predicted as above plot had very small sample for positive datapoints

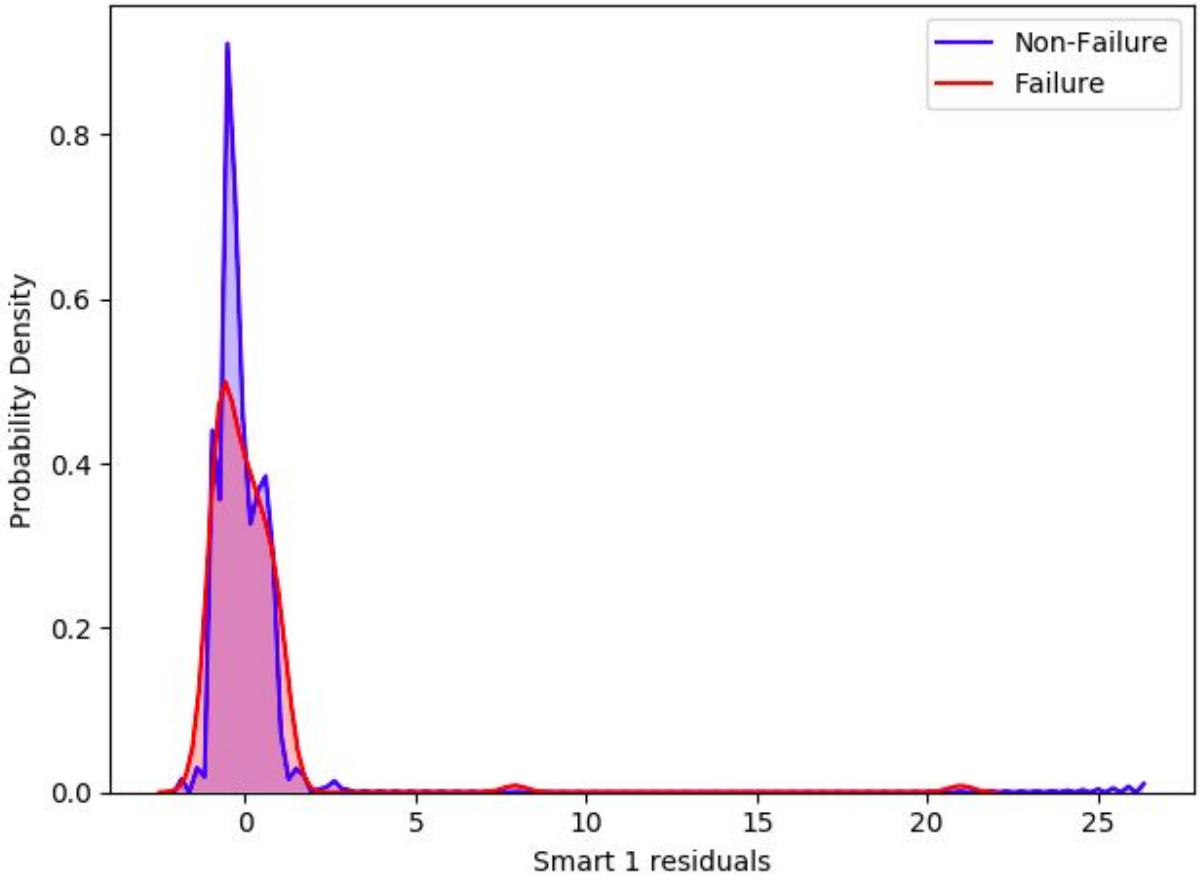
Exploring estimated features from Autoencoders



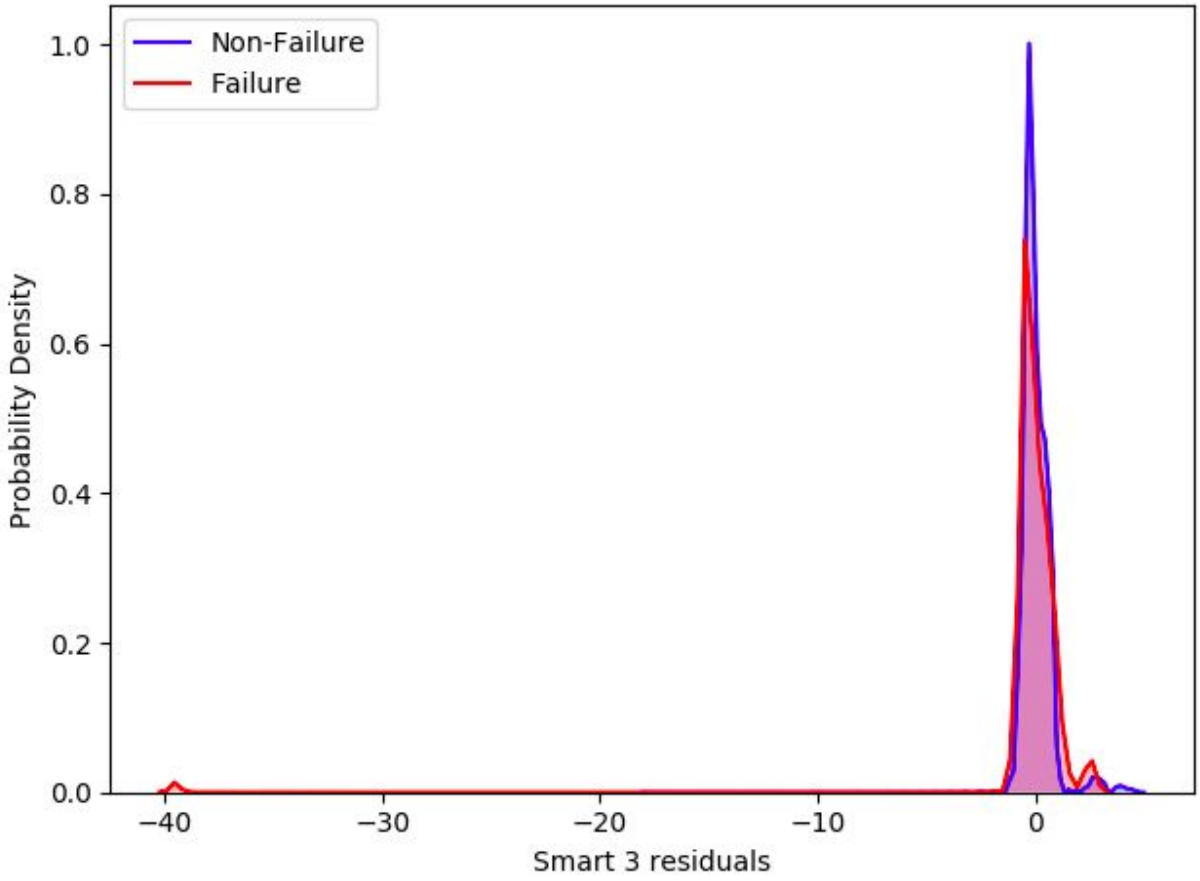
Residuals Plots

Plots for residuals for smart values- 1,3,9, 192, 194, 197, 198 in blue with labels as operational and red as failure versus probability density.

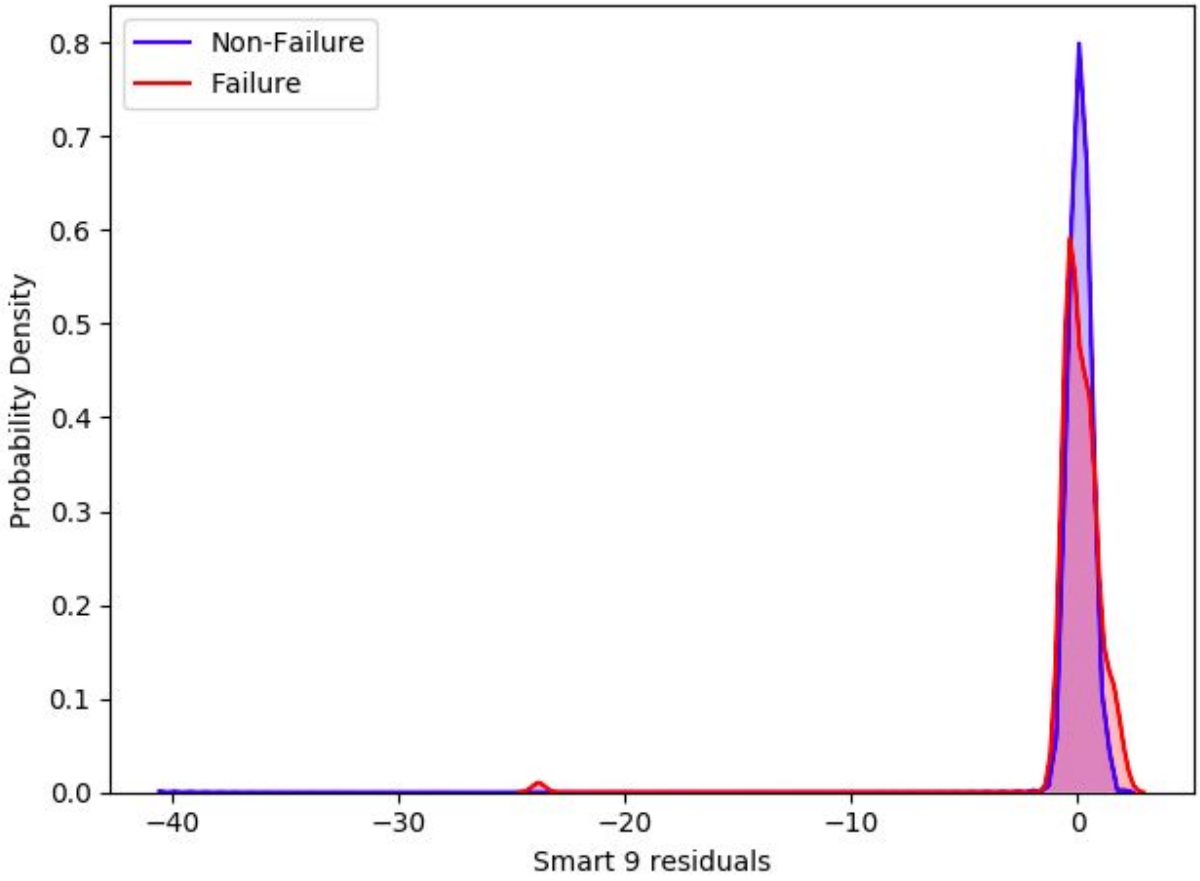
Exploring estimated features from Autoencoders



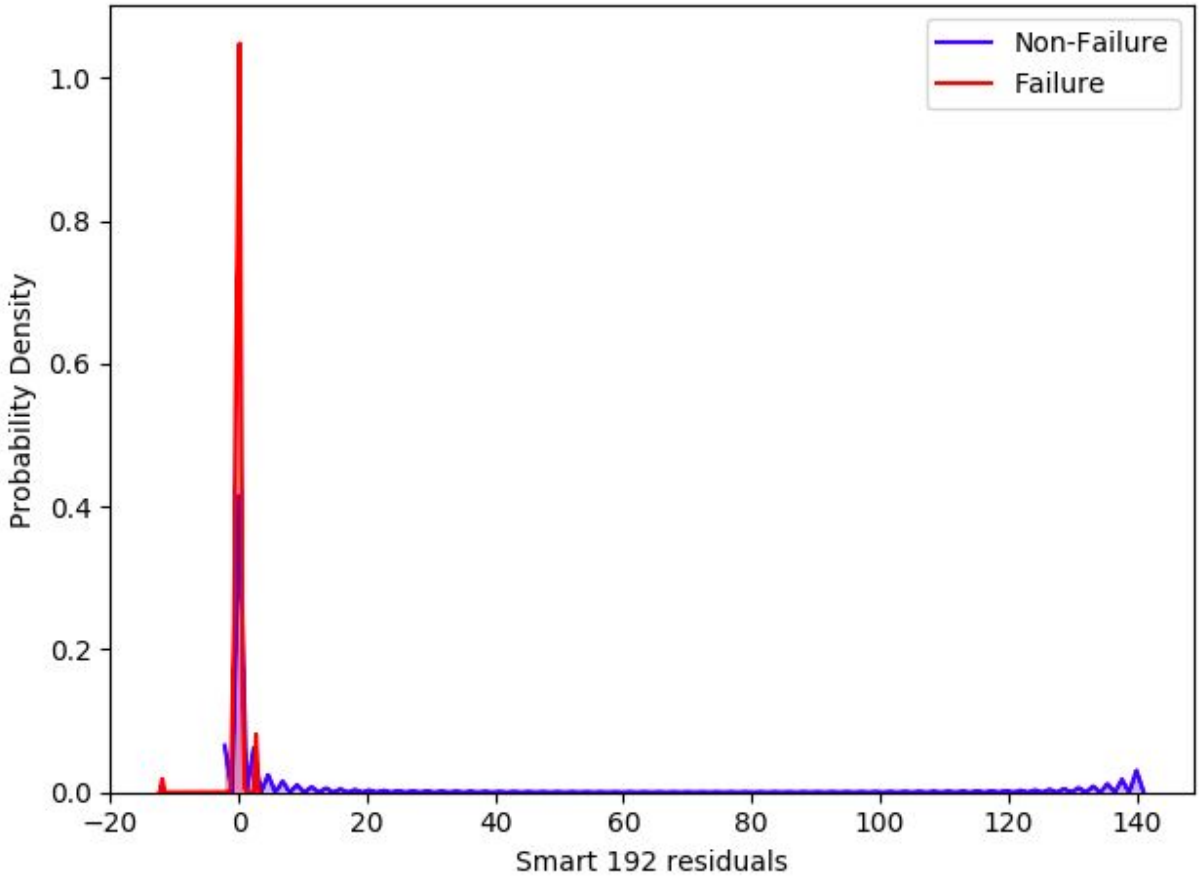
Exploring estimated features from Autoencoders



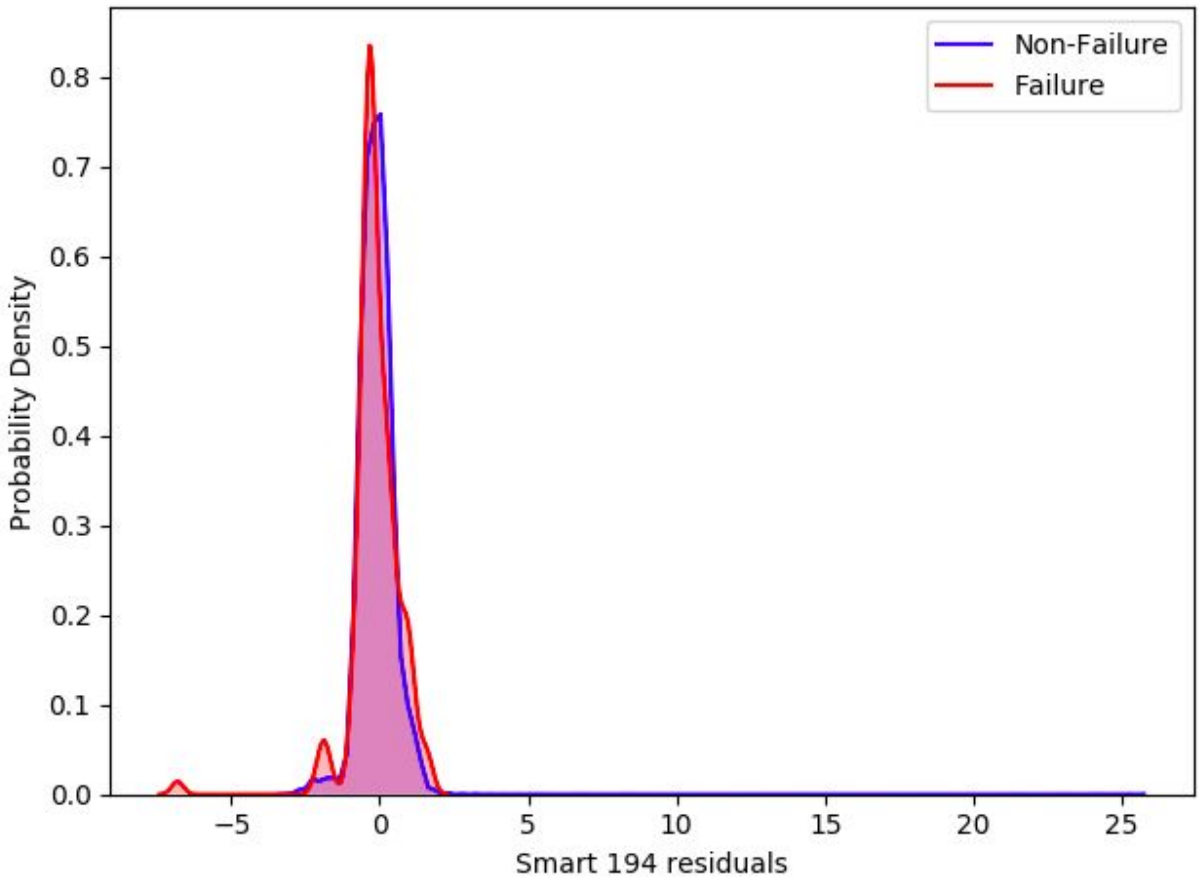
Exploring estimated features from Autoencoders



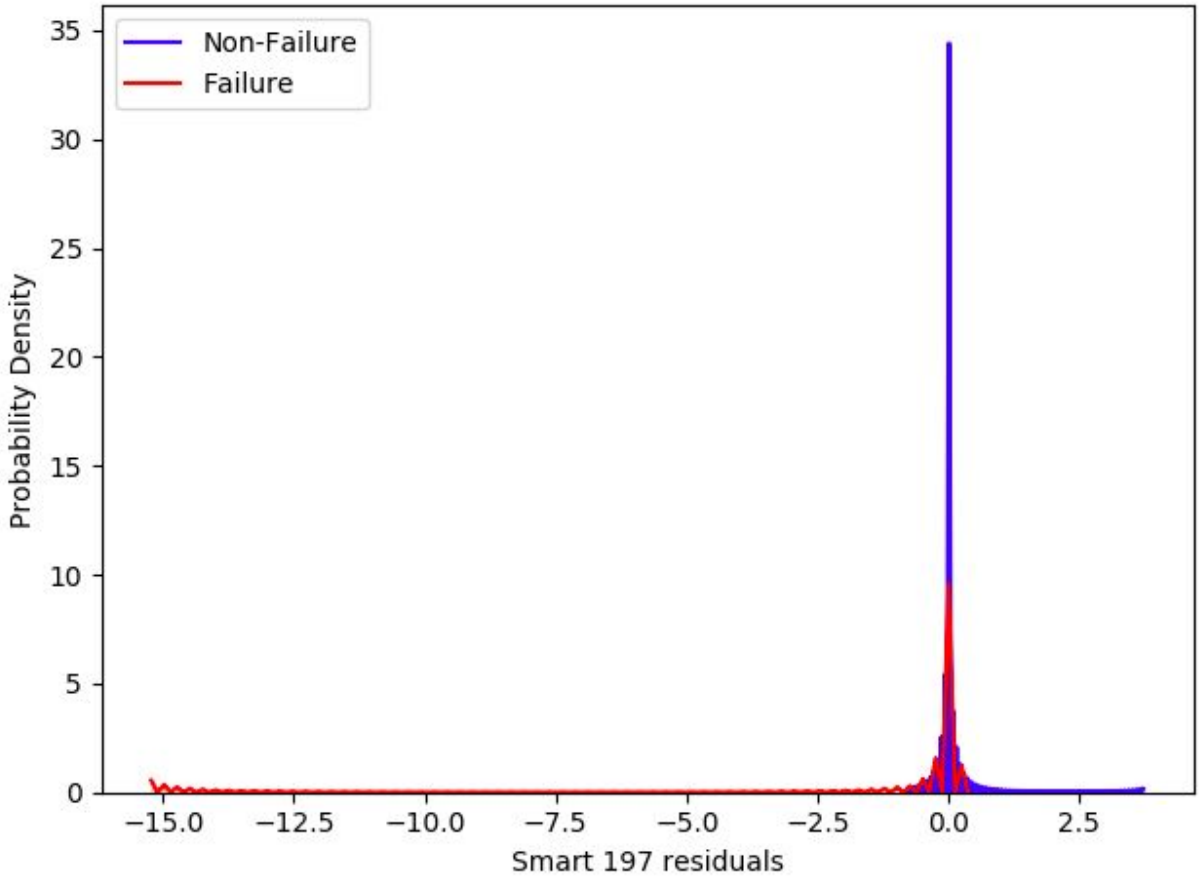
Exploring estimated features from Autoencoders



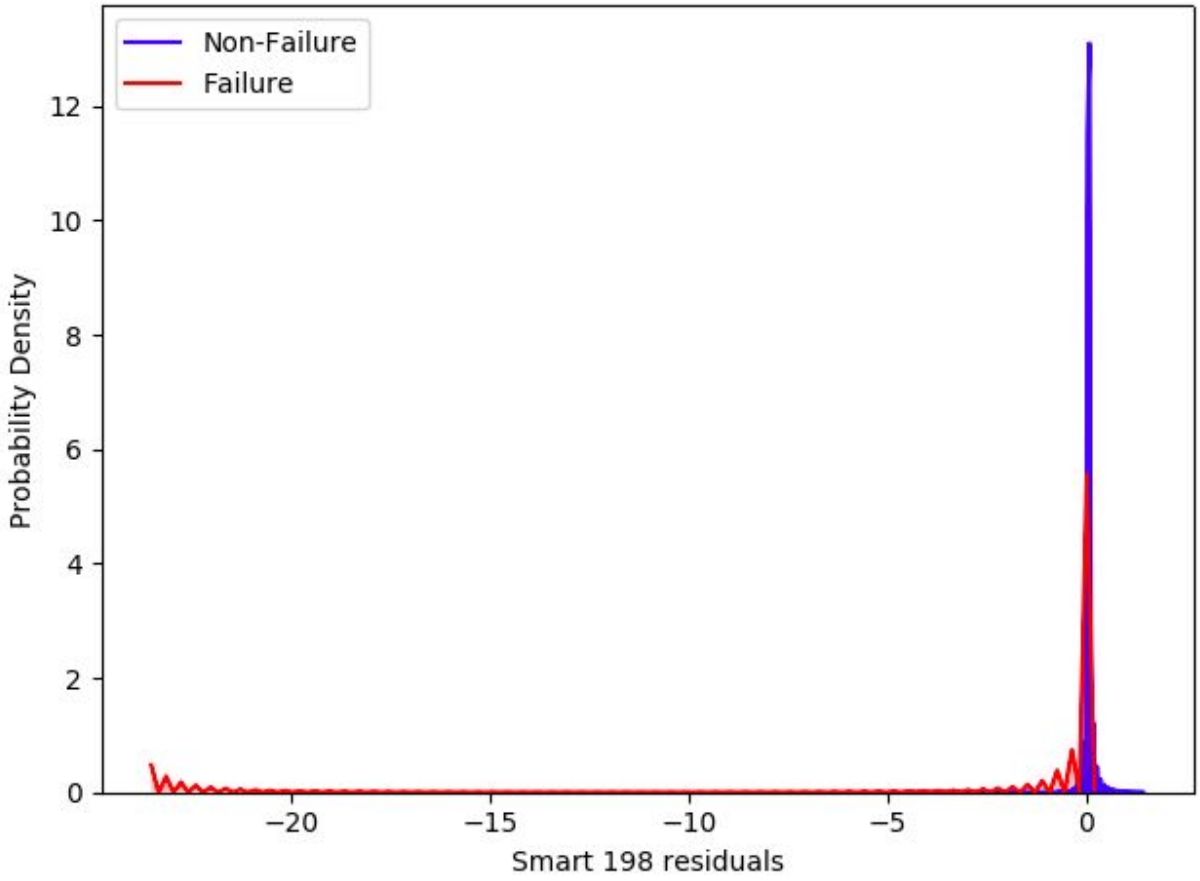
Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders

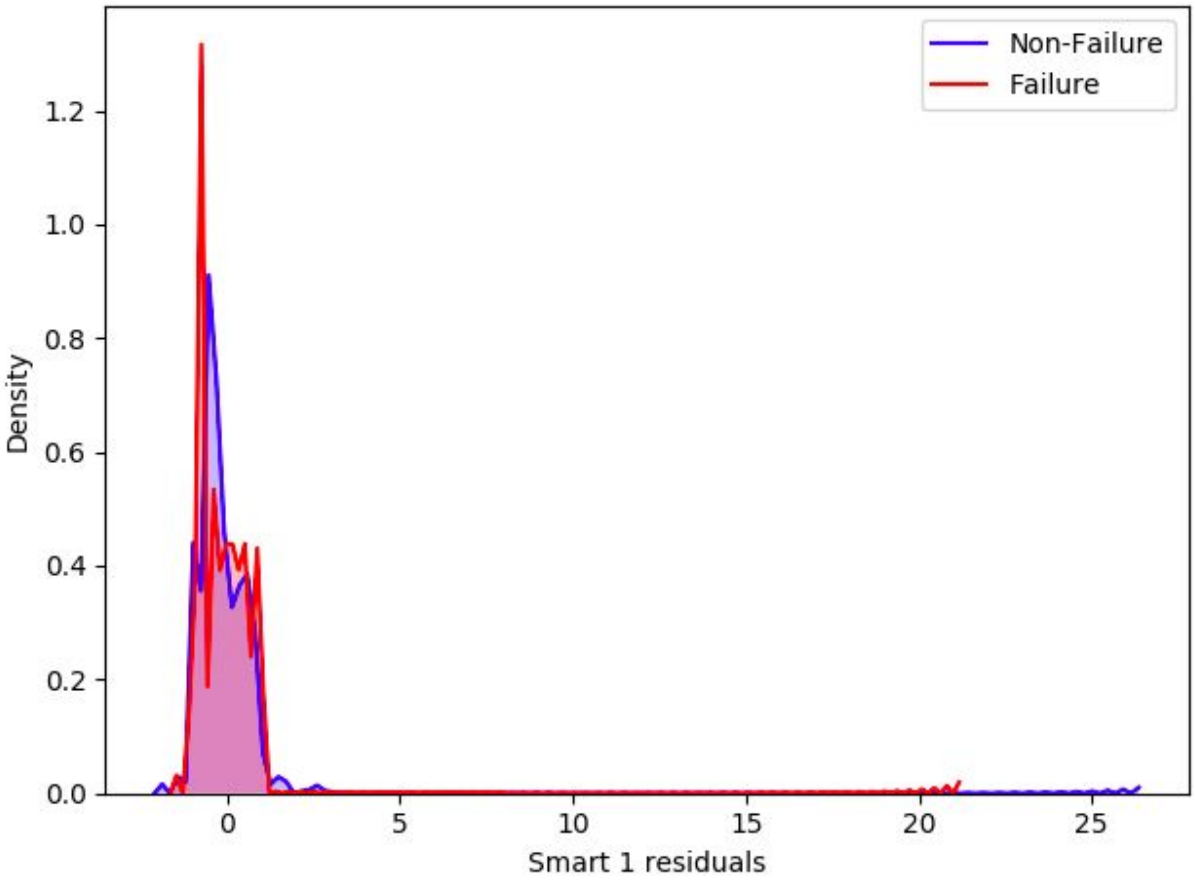


Exploring estimated features from Autoencoders

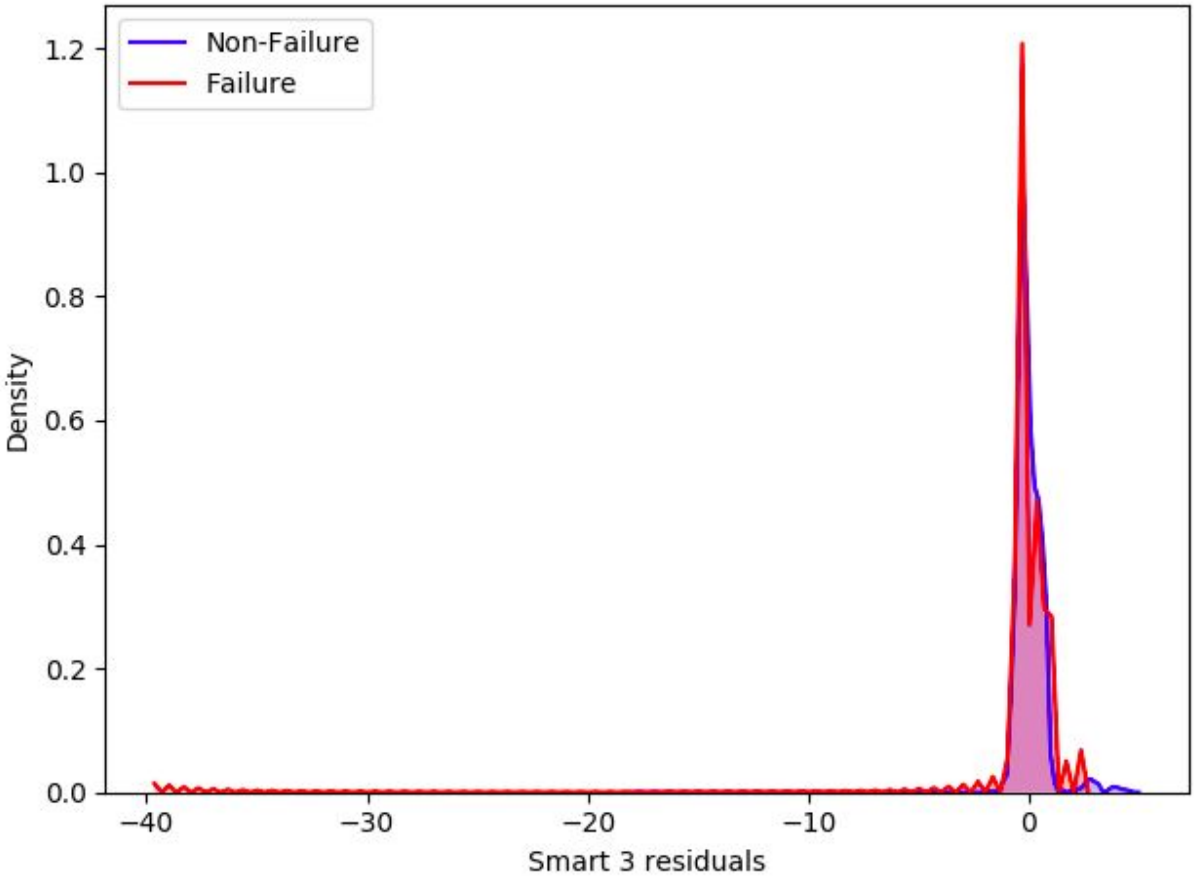


Exploring estimated features from Autoencoders

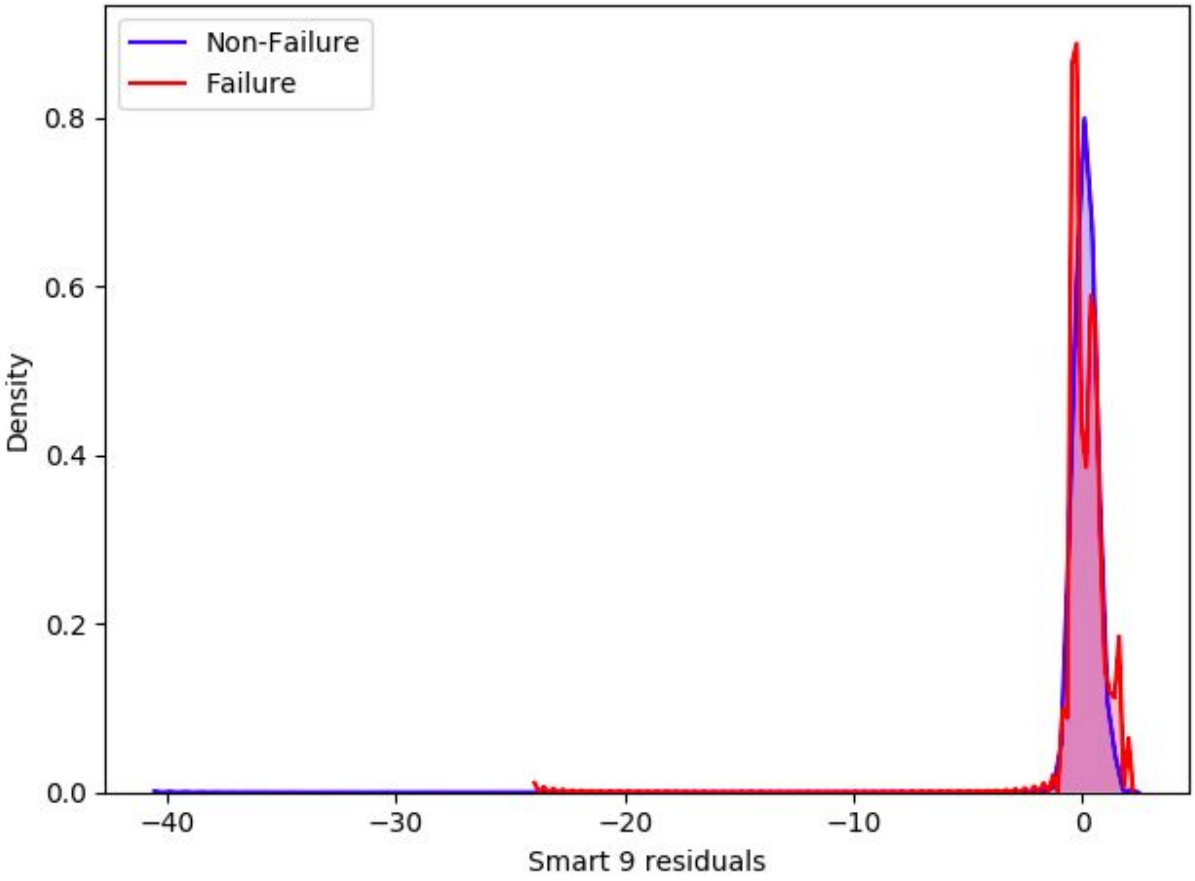
Residual Plots after SMOTE



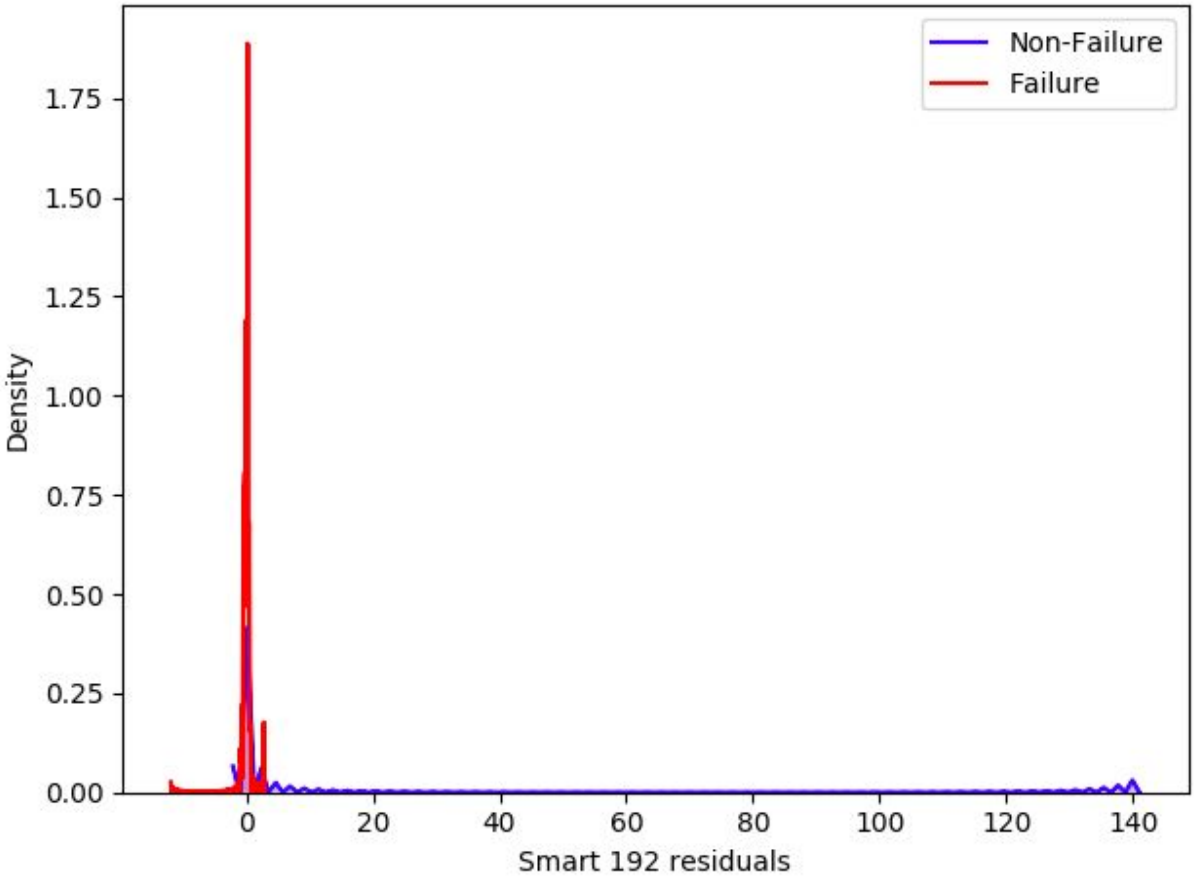
Exploring estimated features from Autoencoders



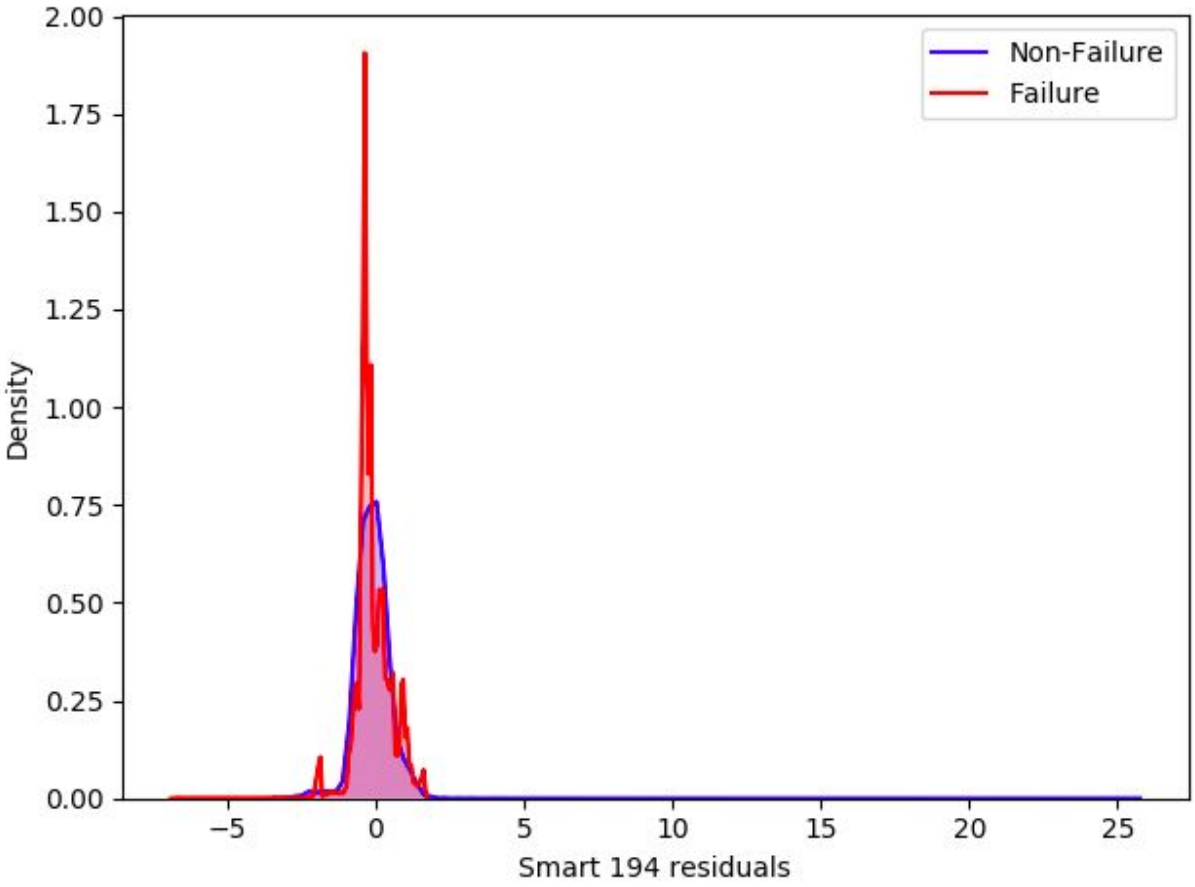
Exploring estimated features from Autoencoders



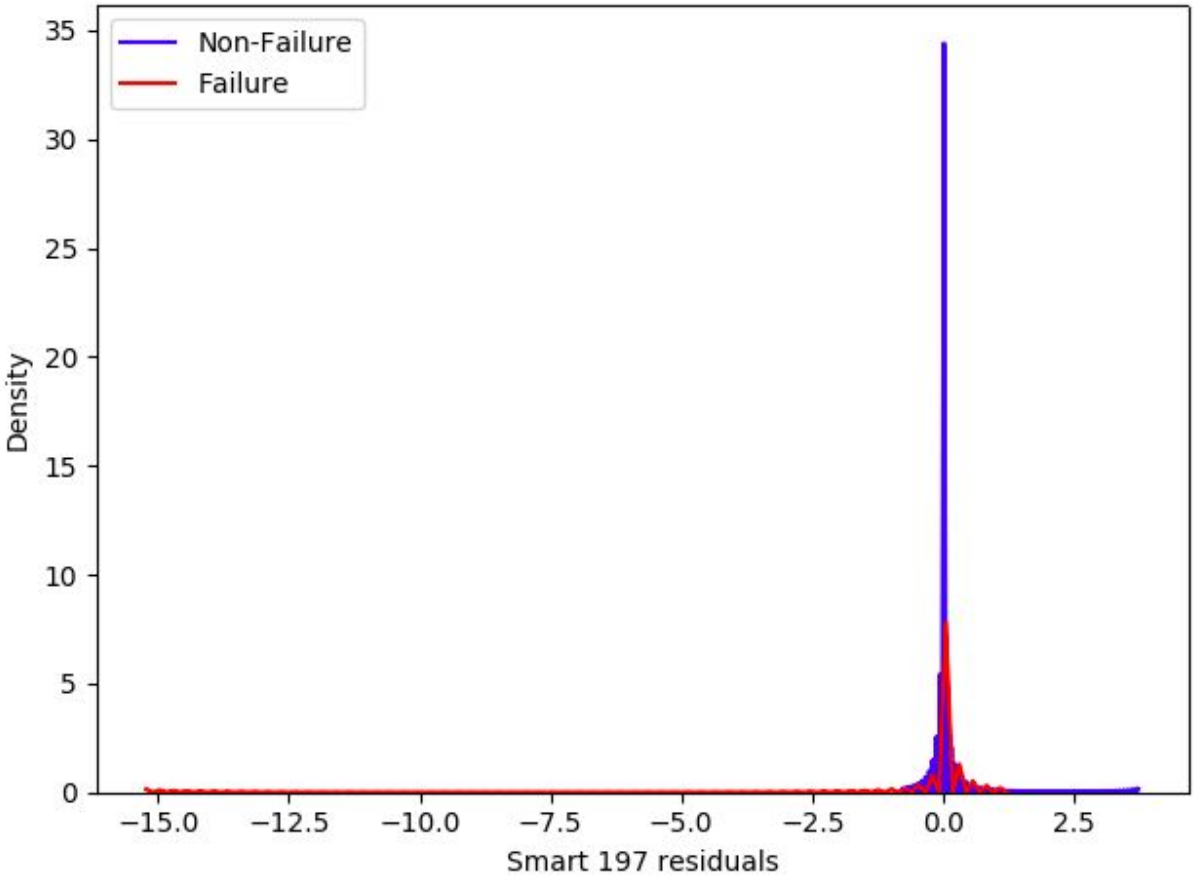
Exploring estimated features from Autoencoders



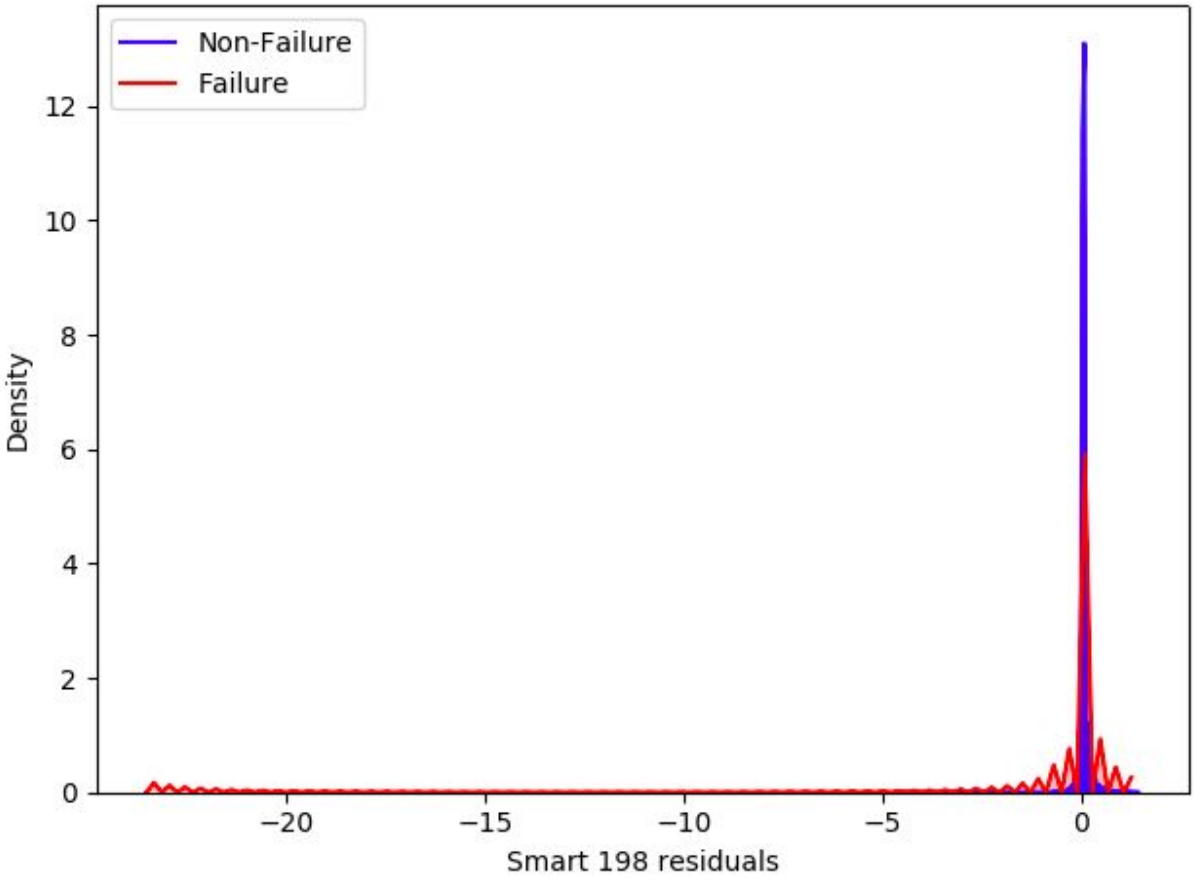
Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders



Random Forest Results

Random Forest without residuals has an accuracy of 88.42% for test data.

Exploring estimated features from Autoencoders

Actual **Predicted**	Class-0/ Negative	Class-1/ Positive (Failure)
Class-0/ Negative	134184 (TN)	21291 (FN / Missed Alarms)
Class-1/ Positive (Failure)	12772 (FP/ False Alarms)	125967 (TP)

Precision = TP/(TP+FP) = 125967/(125967+12772) = 0.9079

Recall = TP/(TP+FN) = 125967/(125967+21291) =0.8554

Random Forest Results with residuals

RF with residuals has accuracy of 94.25% for test data.

Actual **Predicted**	Class-0/ Negative	Class-1/ Positive (Failure)
Class-0/ Negative	139134 (TN)	9097(FN / Missed Alarms)
Class-1/ Positive (Failure)	7822 (FP/ False Alarms)	138161 (TP)

Precision = TP/(TP+FP) = 138161/(138161+7822) = 0.9464

Recall = TP/(TP+FN) = 138161/(138161+9097) = 0.9382

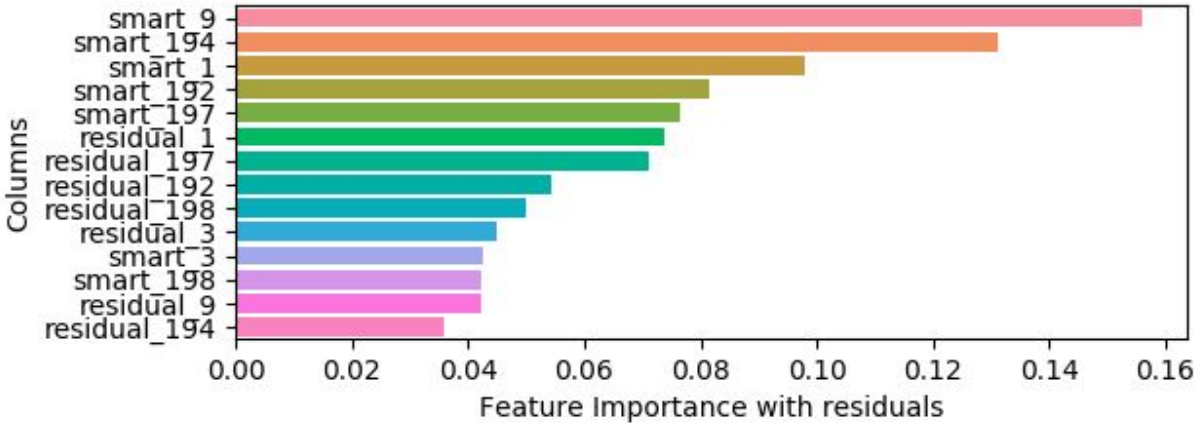
From these two confusion matrices we can notice that after feeding residuals to our classifier

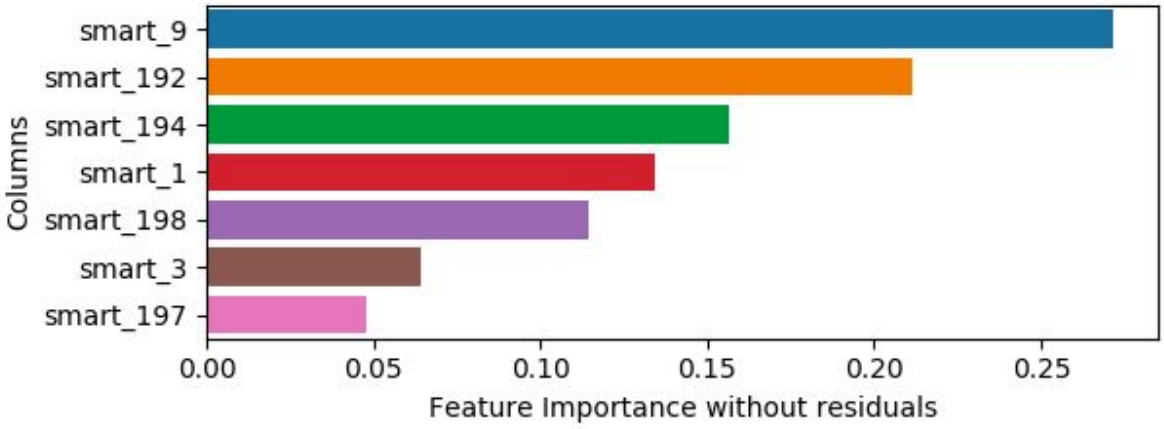
Exploring estimated features from Autoencoders

the Precision, Recall and True Positive counts have increased and False Negative (missed alarms) which are very important to be less to save clients from monetary loss has also decreased

Feature Importance Plot

We can notice that for feature importance few of residuals have even surpassed actual features for example residual 3 and residual 198, and others are also proving to be significant for classification with improved results.





Exploring estimated features from Autoencoders

ROC Curves

Below plot is ROC curve of Random Forest before and after feeding residuals which also shows that for ROC curve with residuals in green has True Positive Rate is at higher than ROC curve without residuals in blue at lower value of False Positive rate.