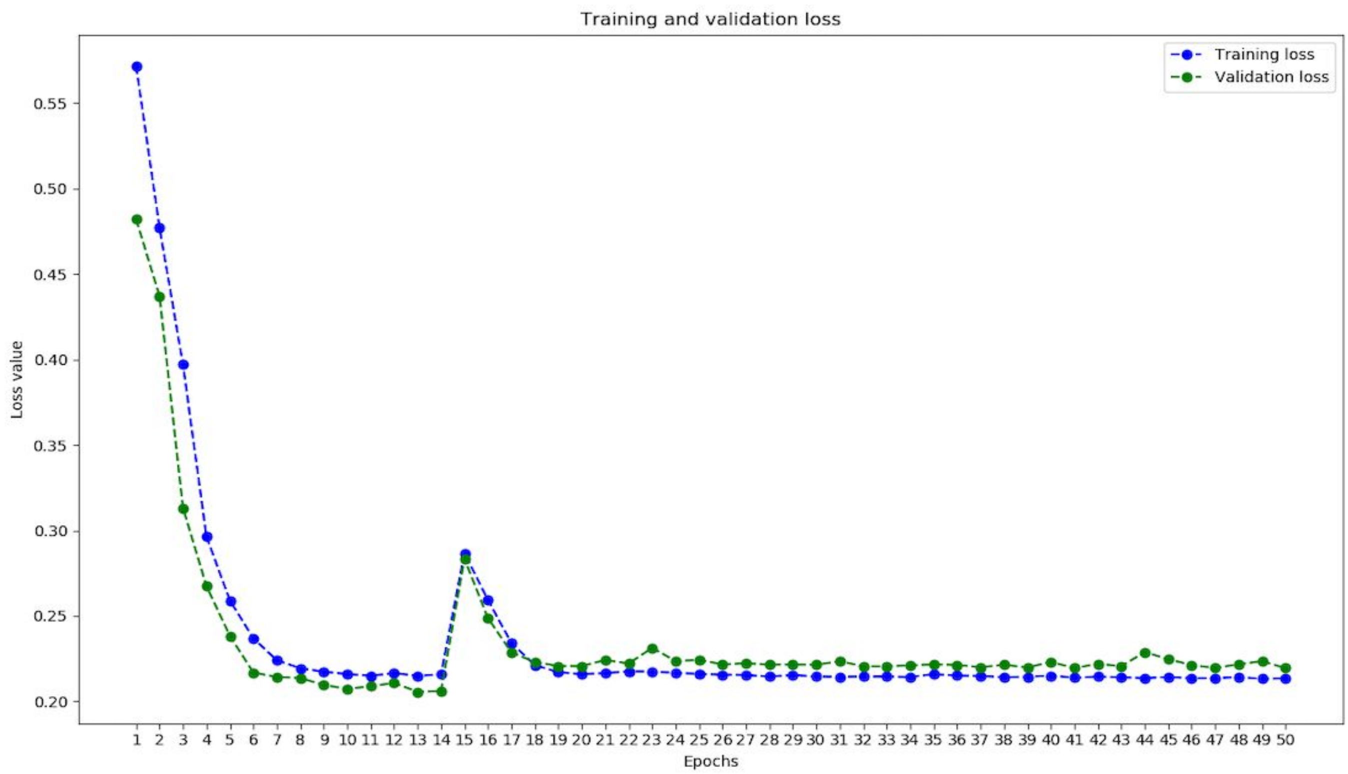
Exploring estimated features from Autoencoders

Autoencoder Loss value Plot

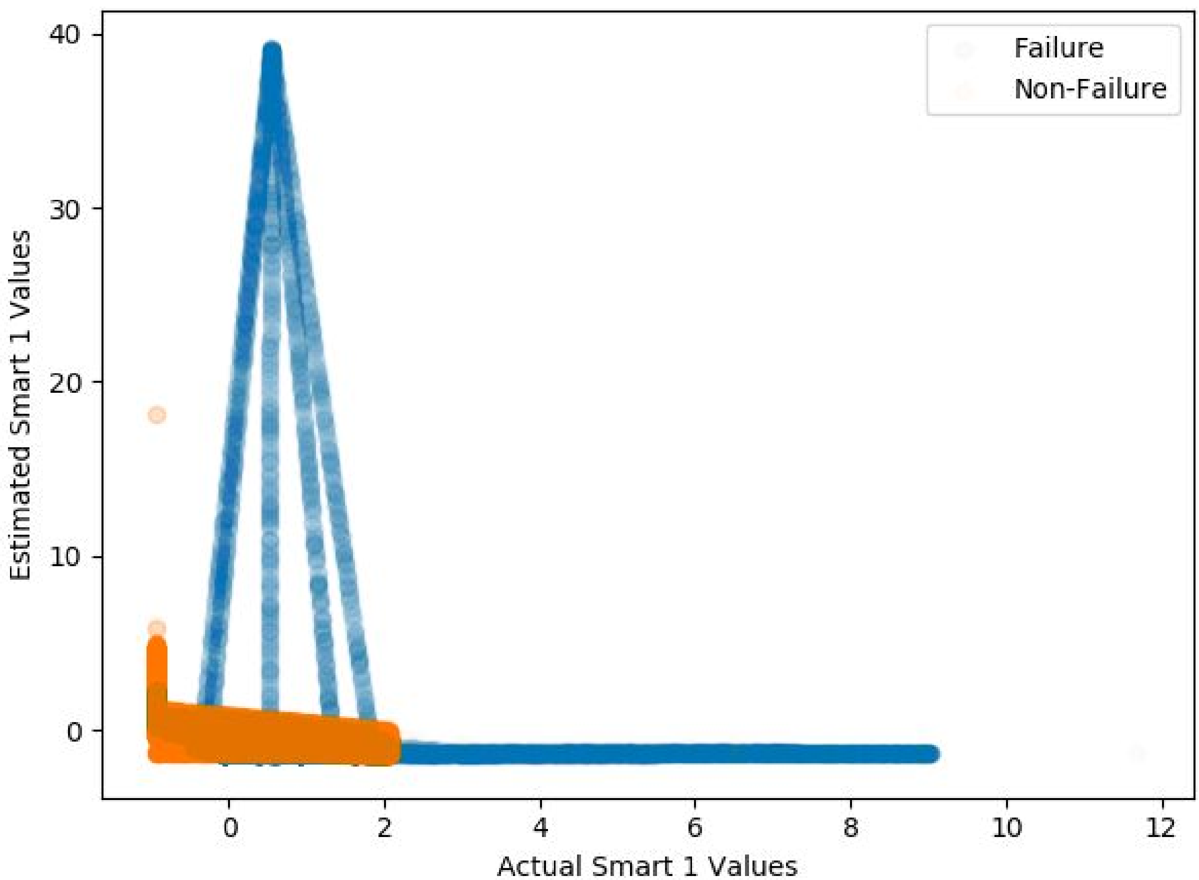
Plot after cross validation over autoencoder, we can see initially validation error is even lesser than training loss but as epoch value increases the autoencoder learns better and hence validation loss becomes a bit more than training loss.
The average loss for training data is - 0.3164
The average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



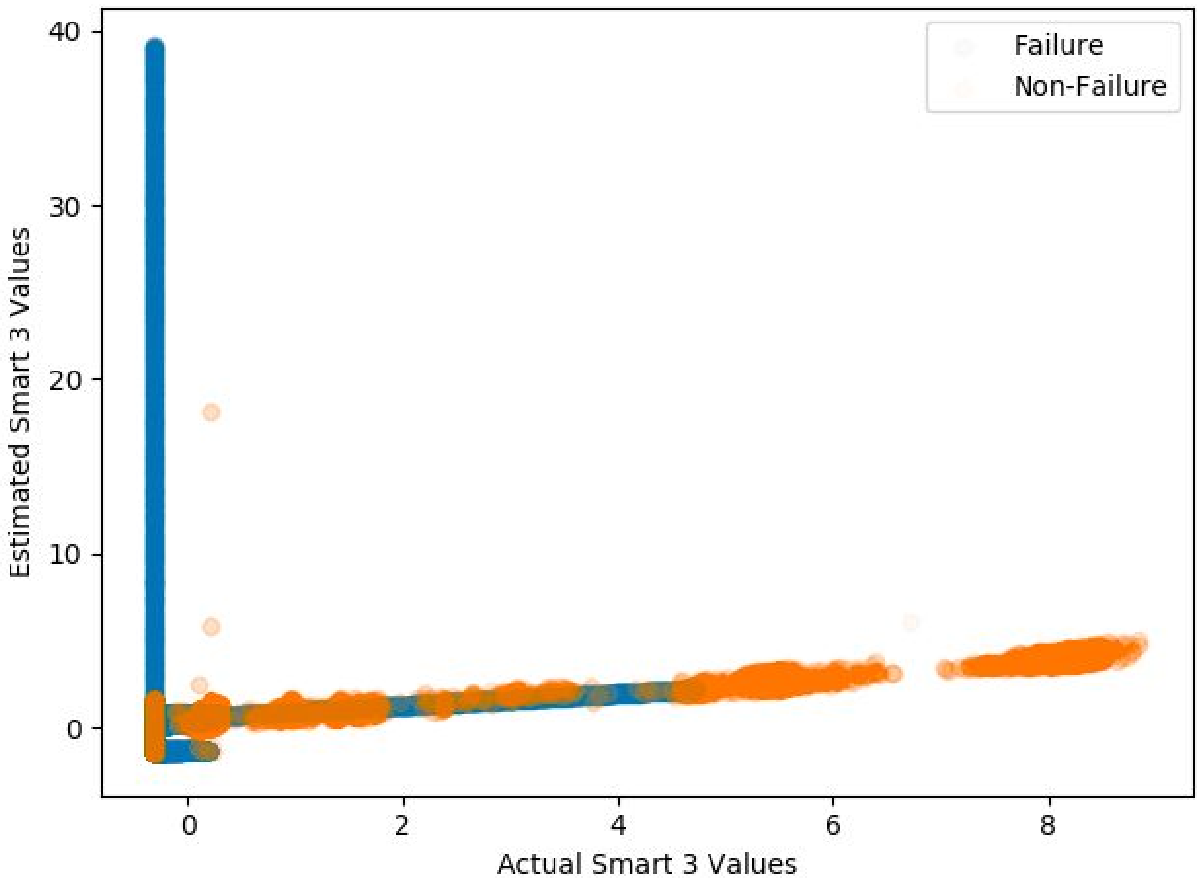
Exploring estimated features from Autoencoders

Estimated Versus Actual Feature Plots

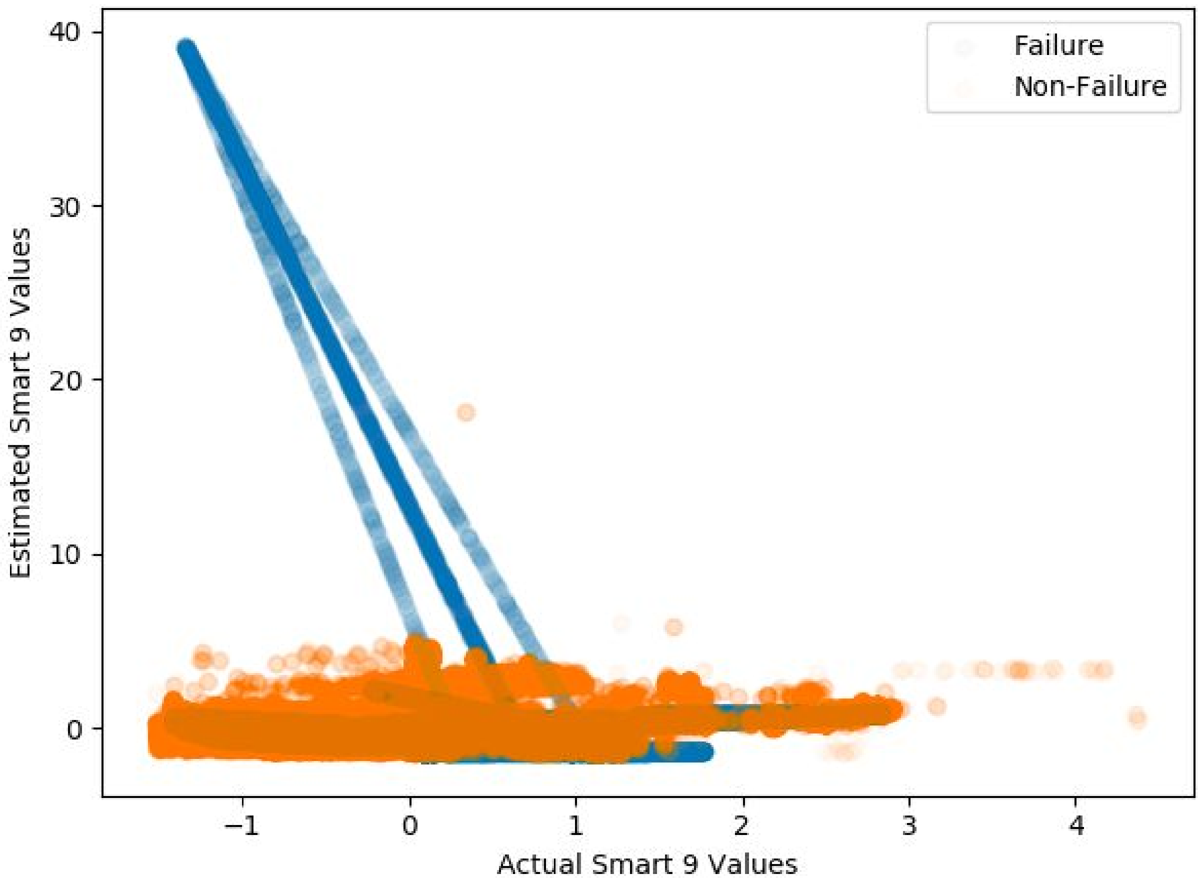
The plots for input and output features estimated by trained autoencoder for new data, used 1 million, 400 data points obtained after performing oversampling using SMOTE



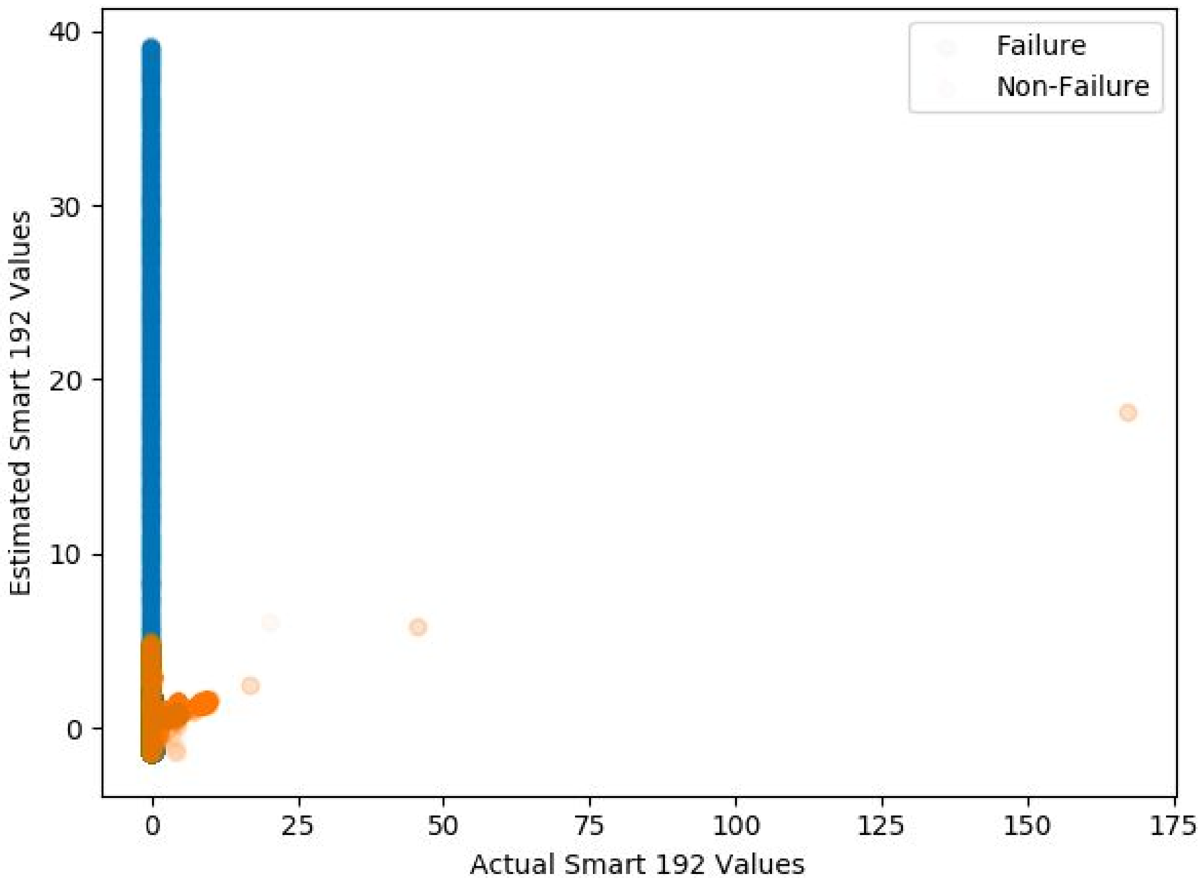
Exploring estimated features from Autoencoders



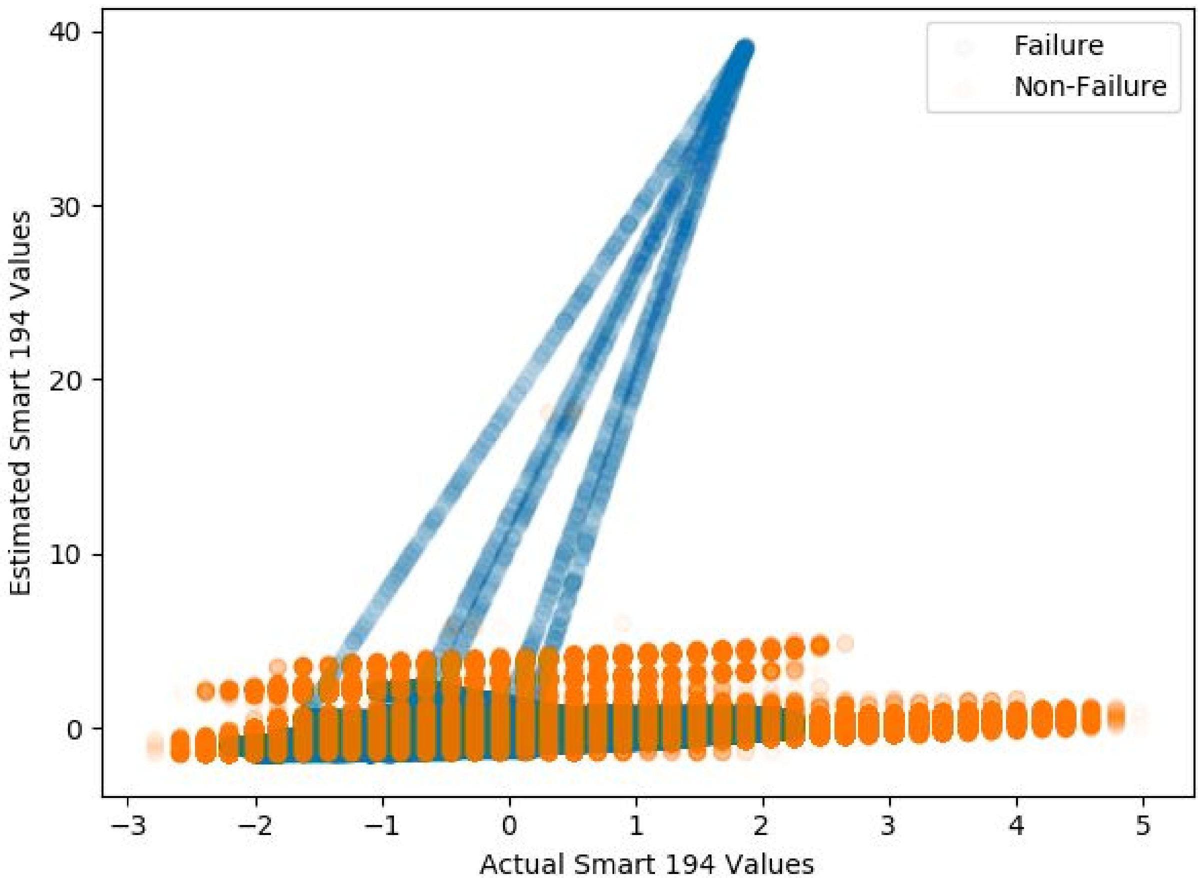
Exploring estimated features from Autoencoders



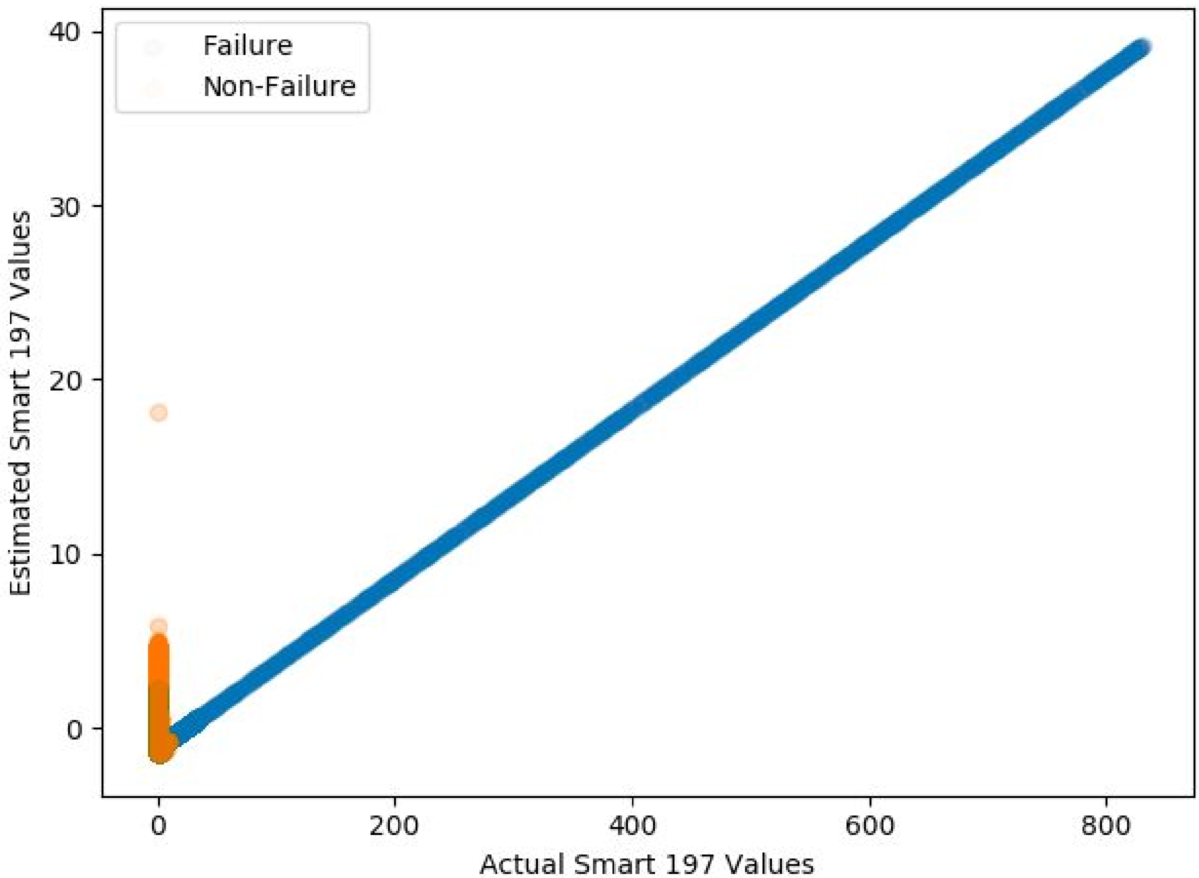
Exploring estimated features from Autoencoders



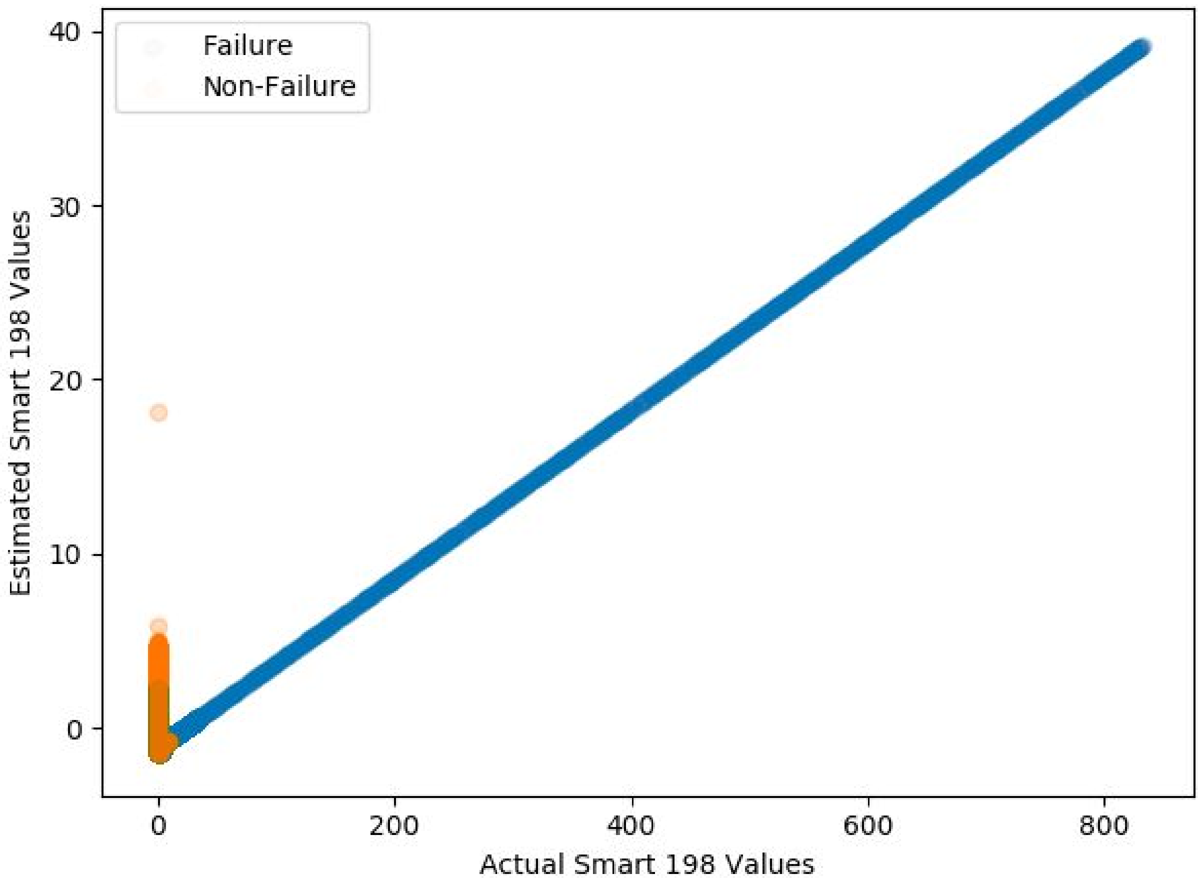
Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders



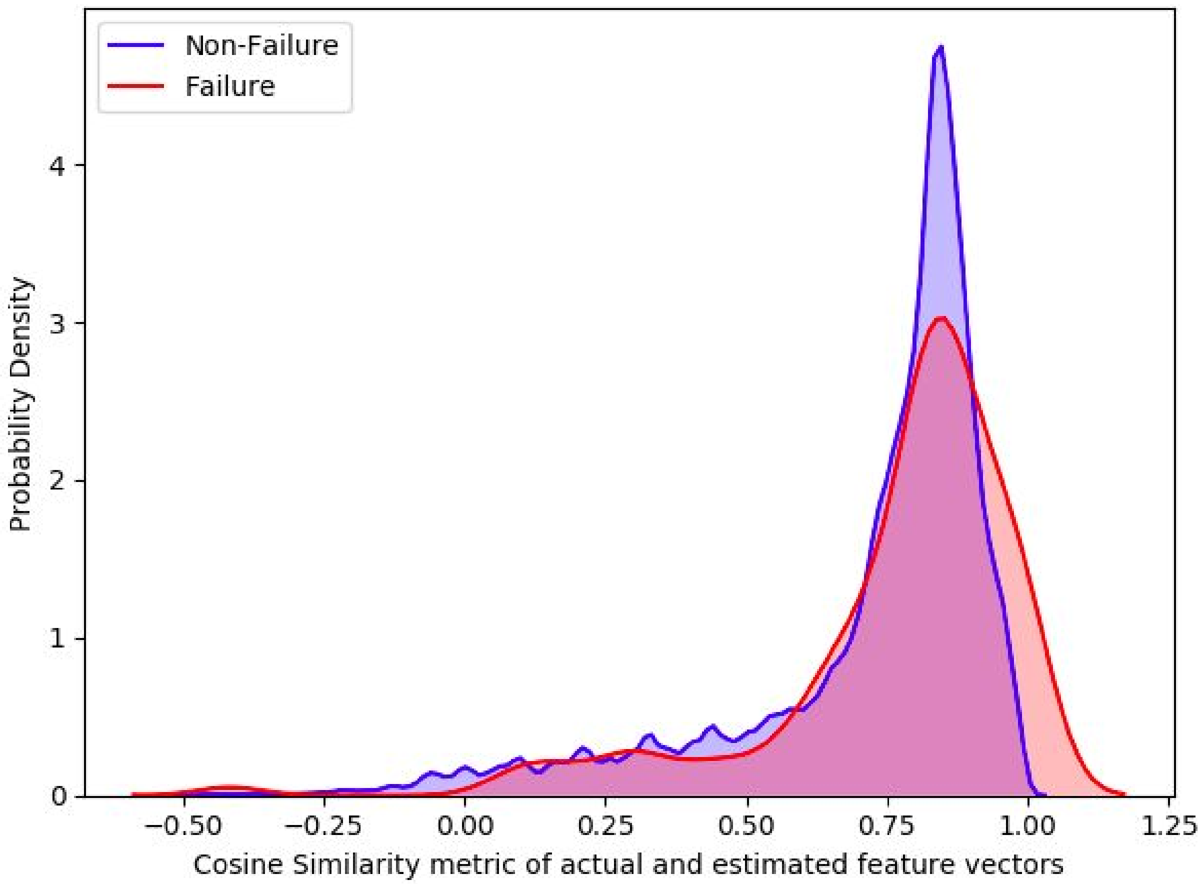
Exploring estimated features from Autoencoders



Cosine Similarity Plot

Density plot for COSINE SIMILARITY of estimated and actual features

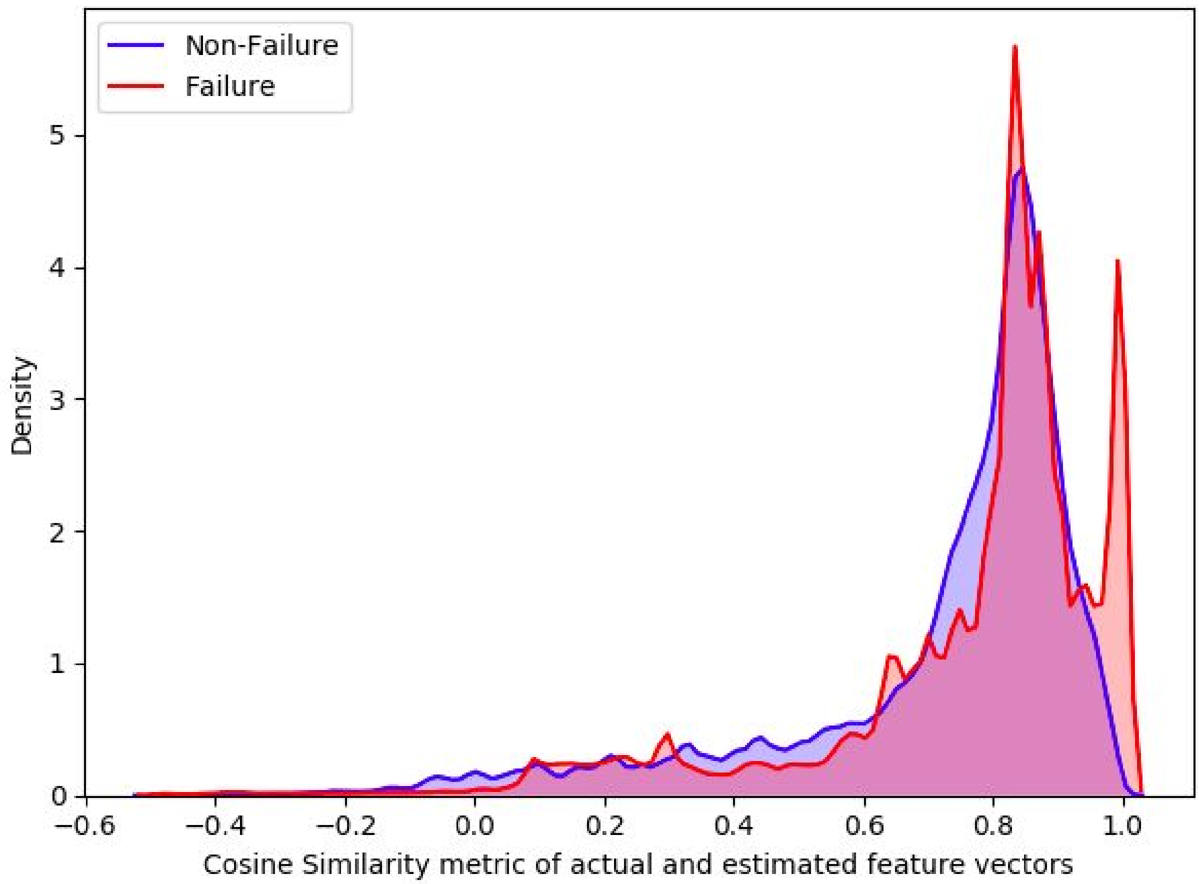
Exploring estimated features from Autoencoders



Cosine Sim Plot after Oversampling

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

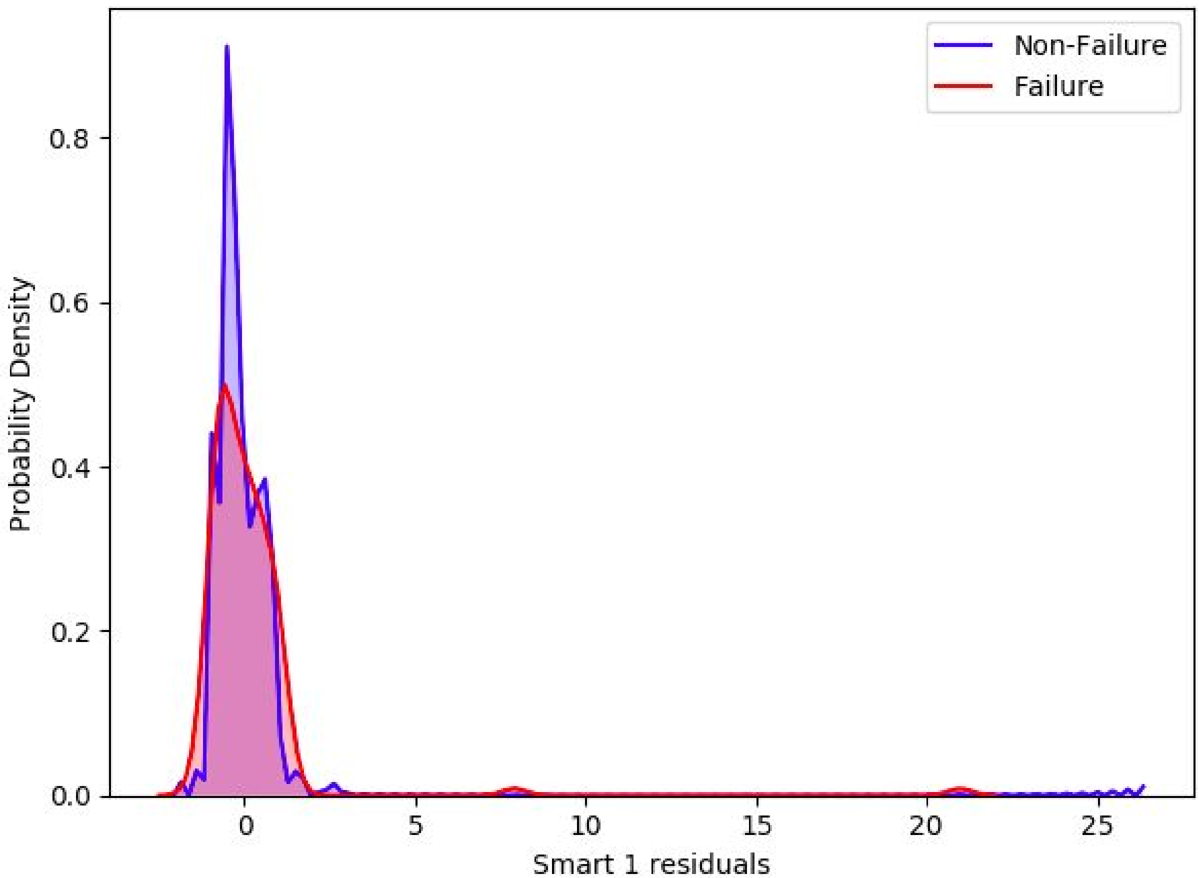
Exploring estimated features from Autoencoders



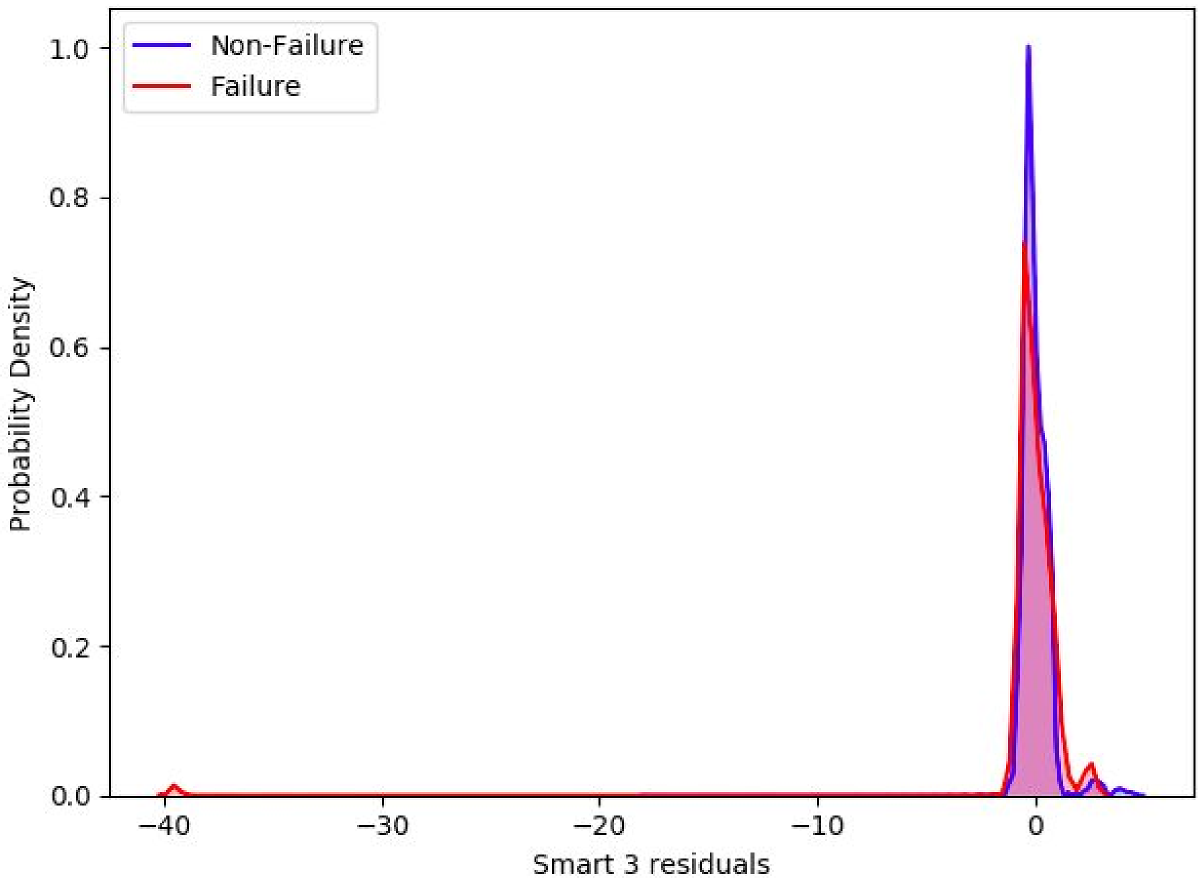
Residuals Plots

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.

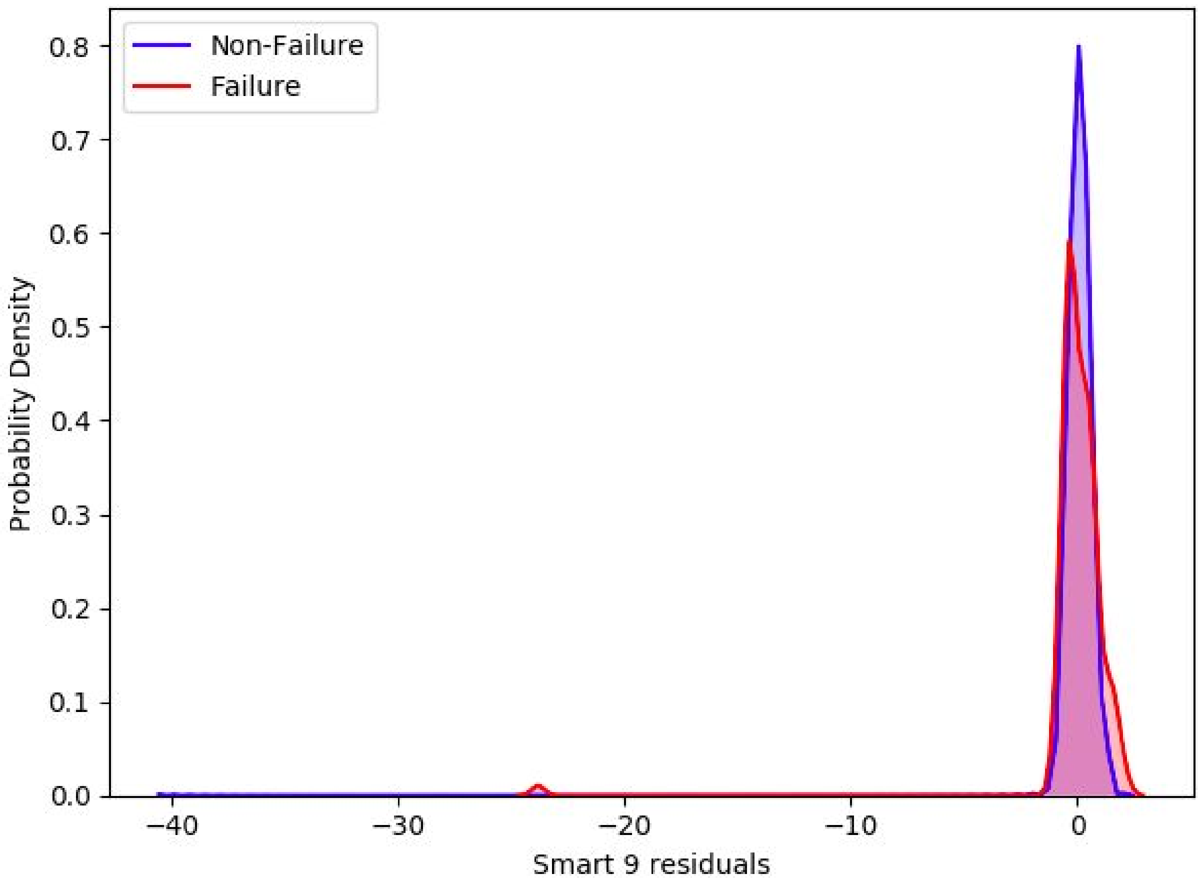
Exploring estimated features from Autoencoders



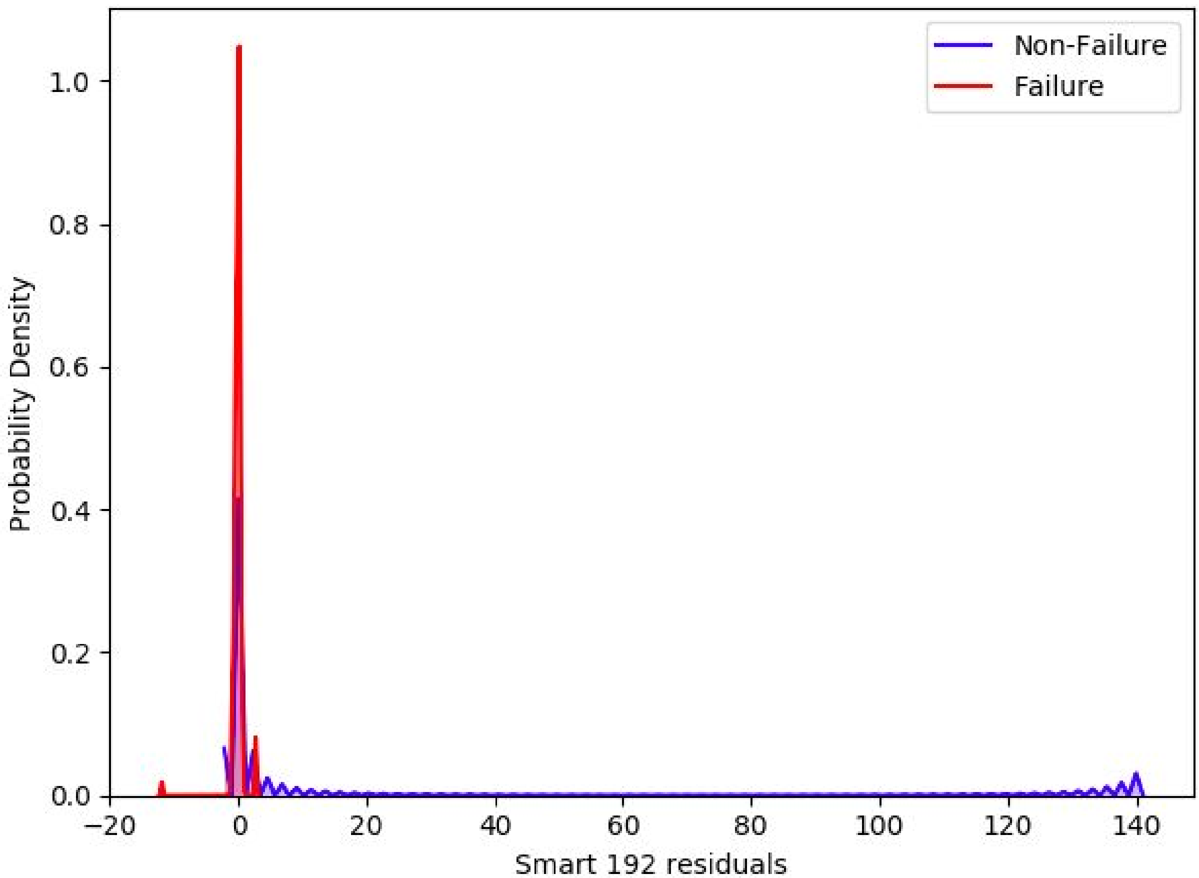
Exploring estimated features from Autoencoders



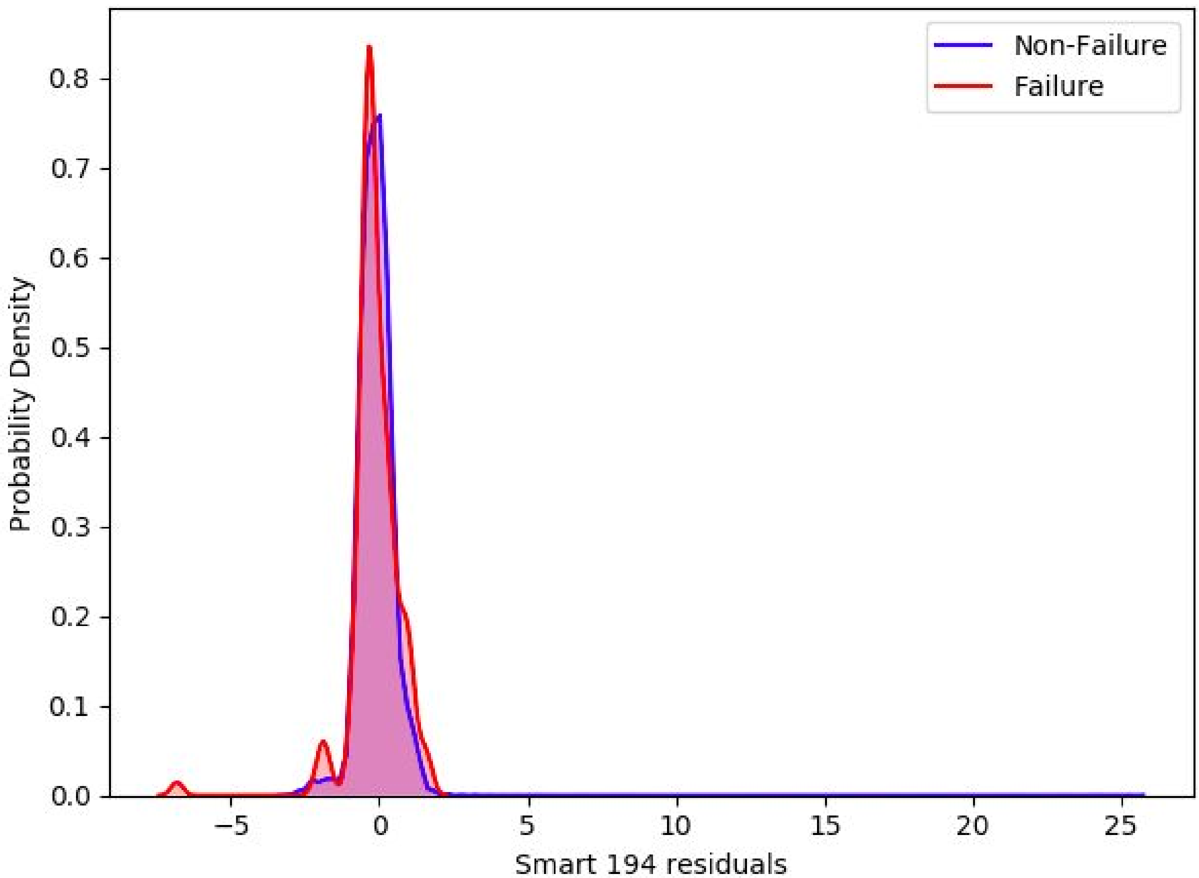
Exploring estimated features from Autoencoders



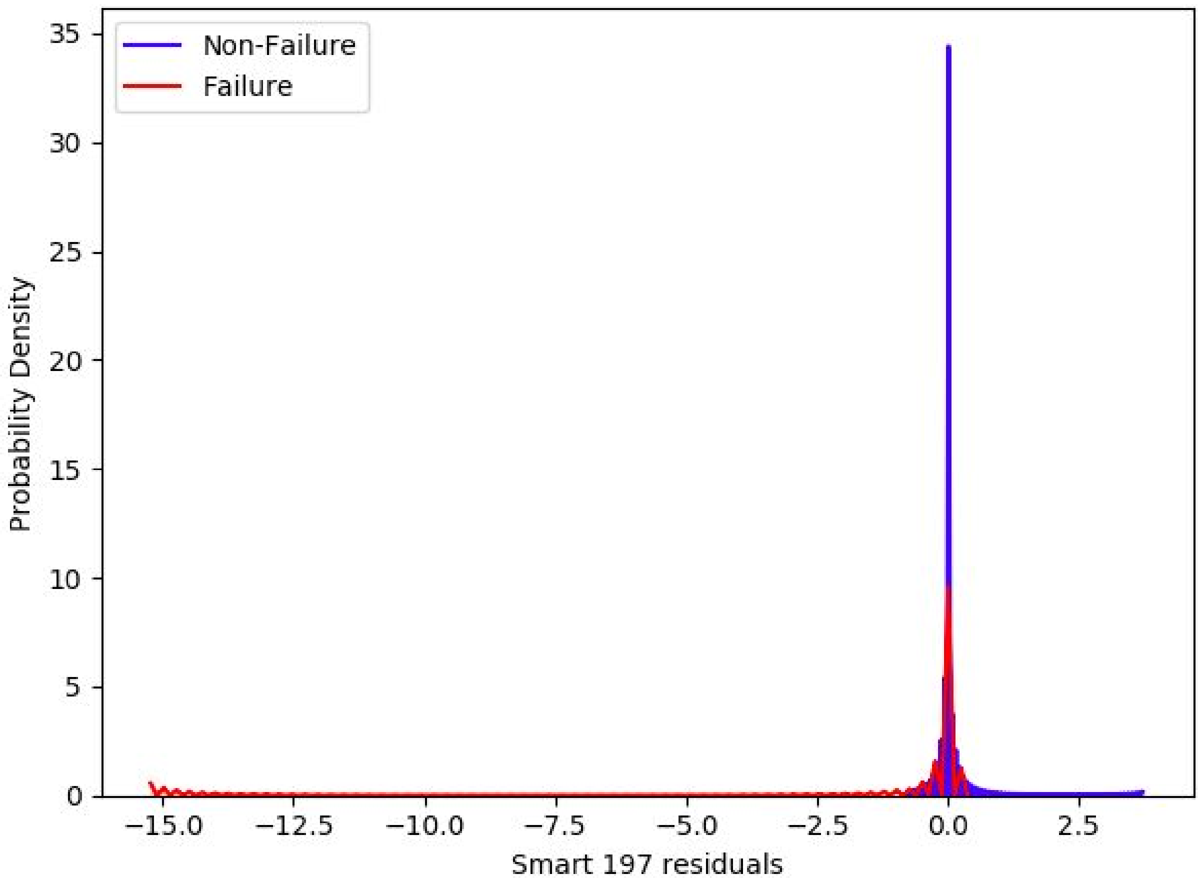
Exploring estimated features from Autoencoders



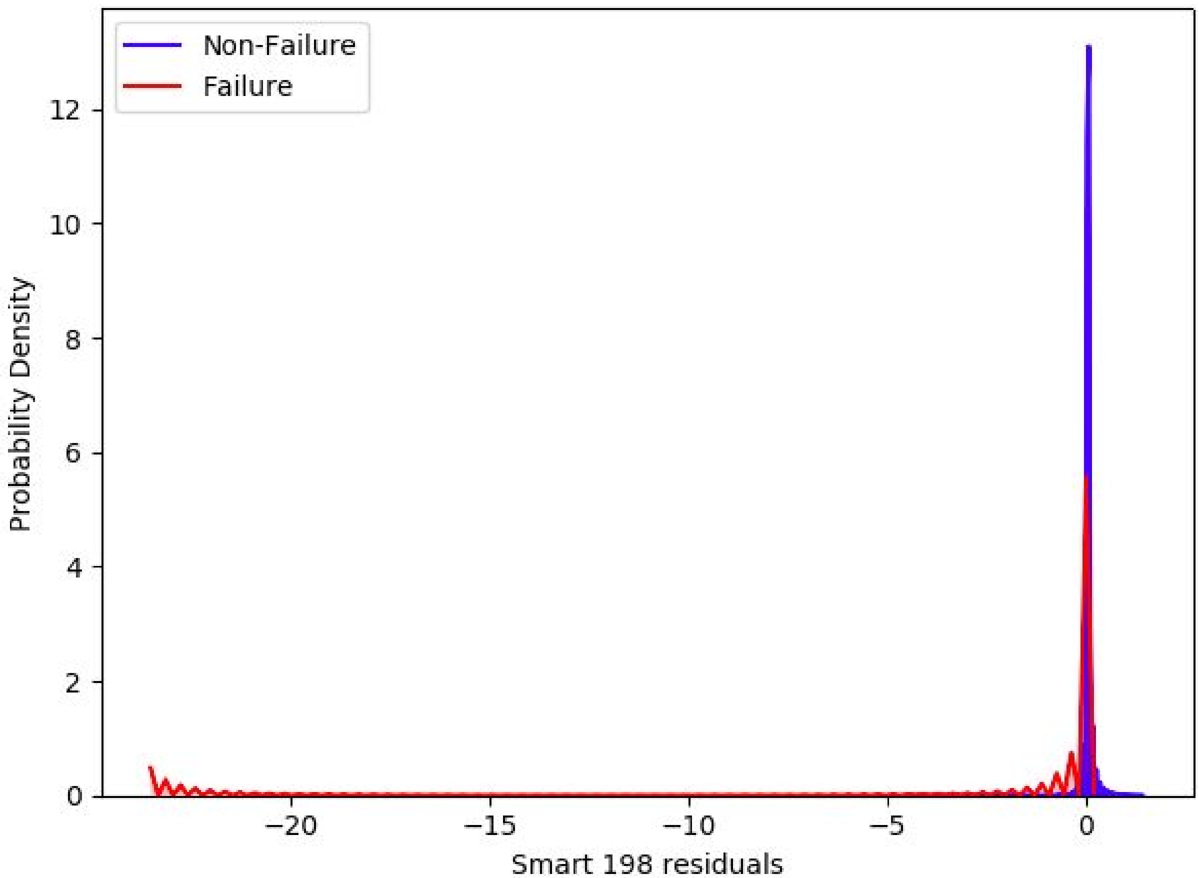
Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders

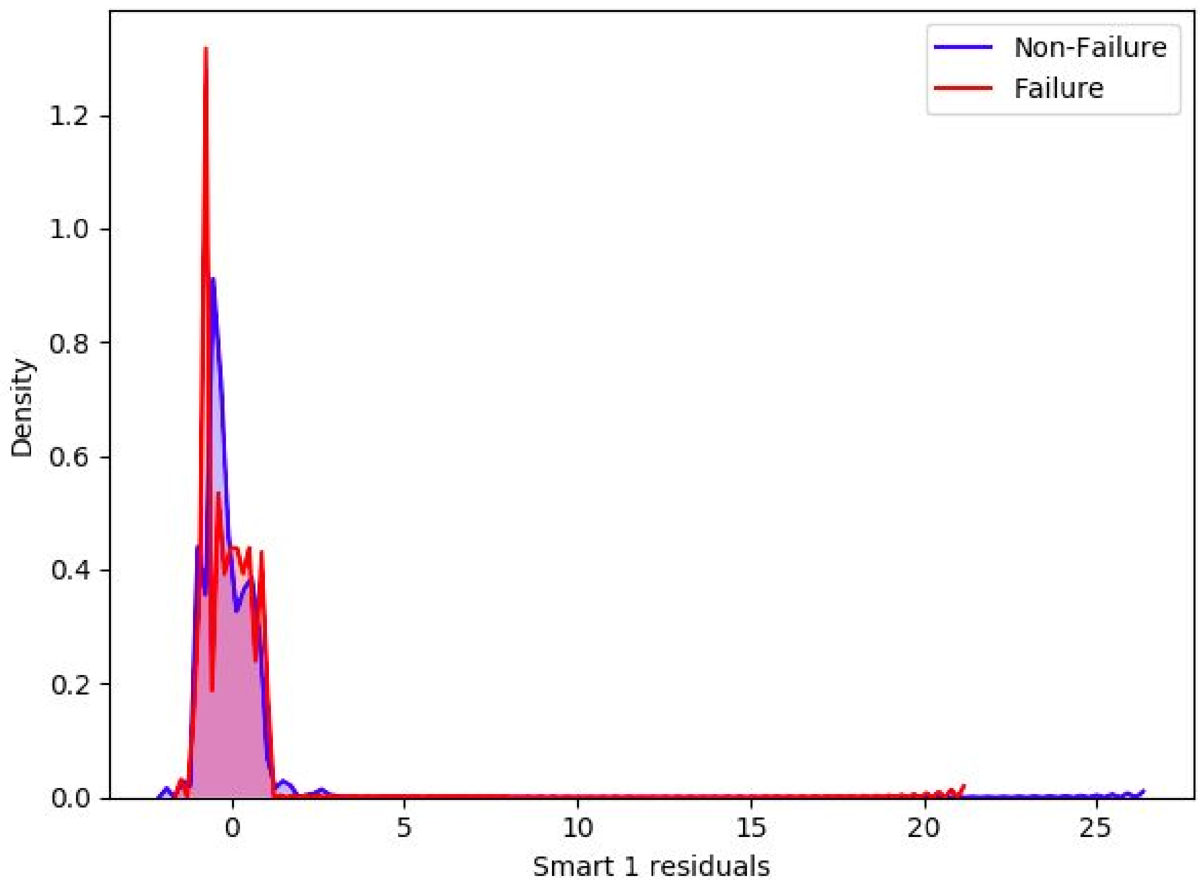


Exploring estimated features from Autoencoders

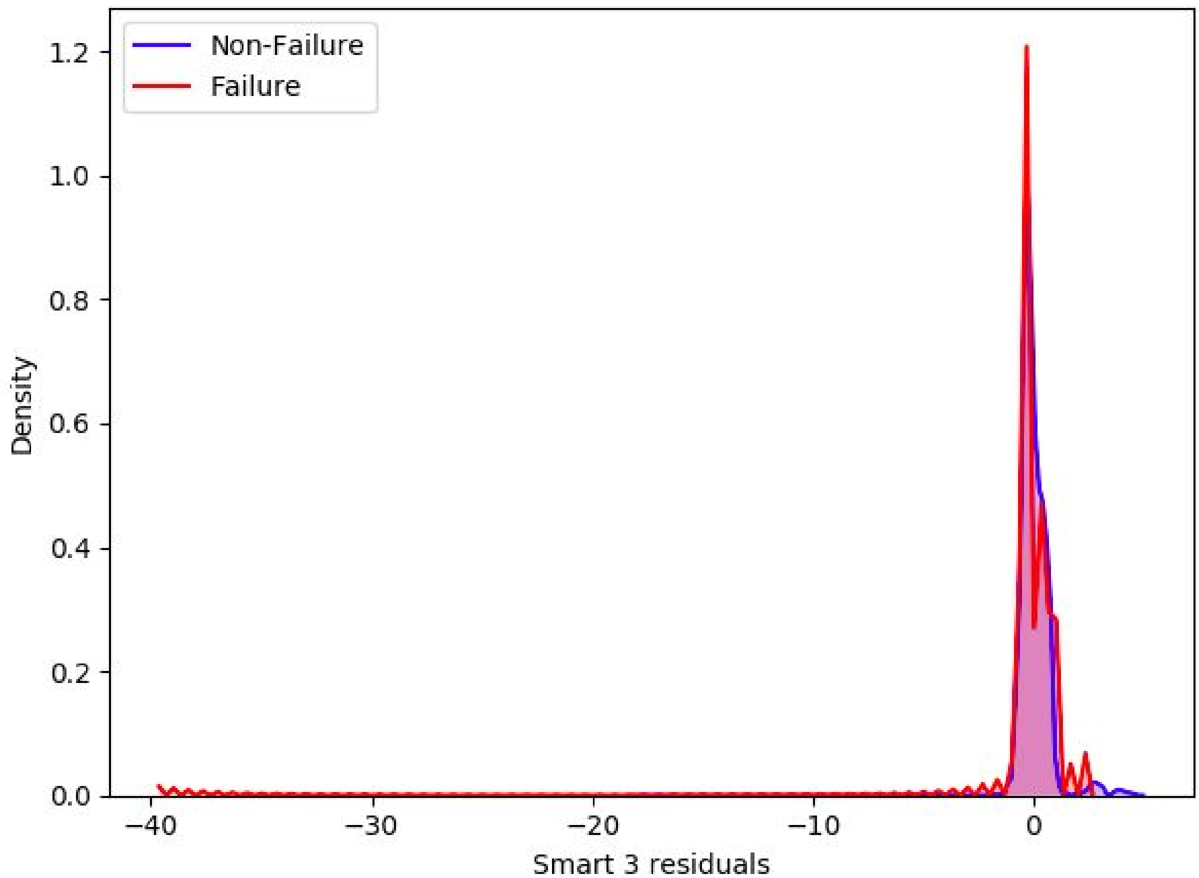


Exploring estimated features from Autoencoders

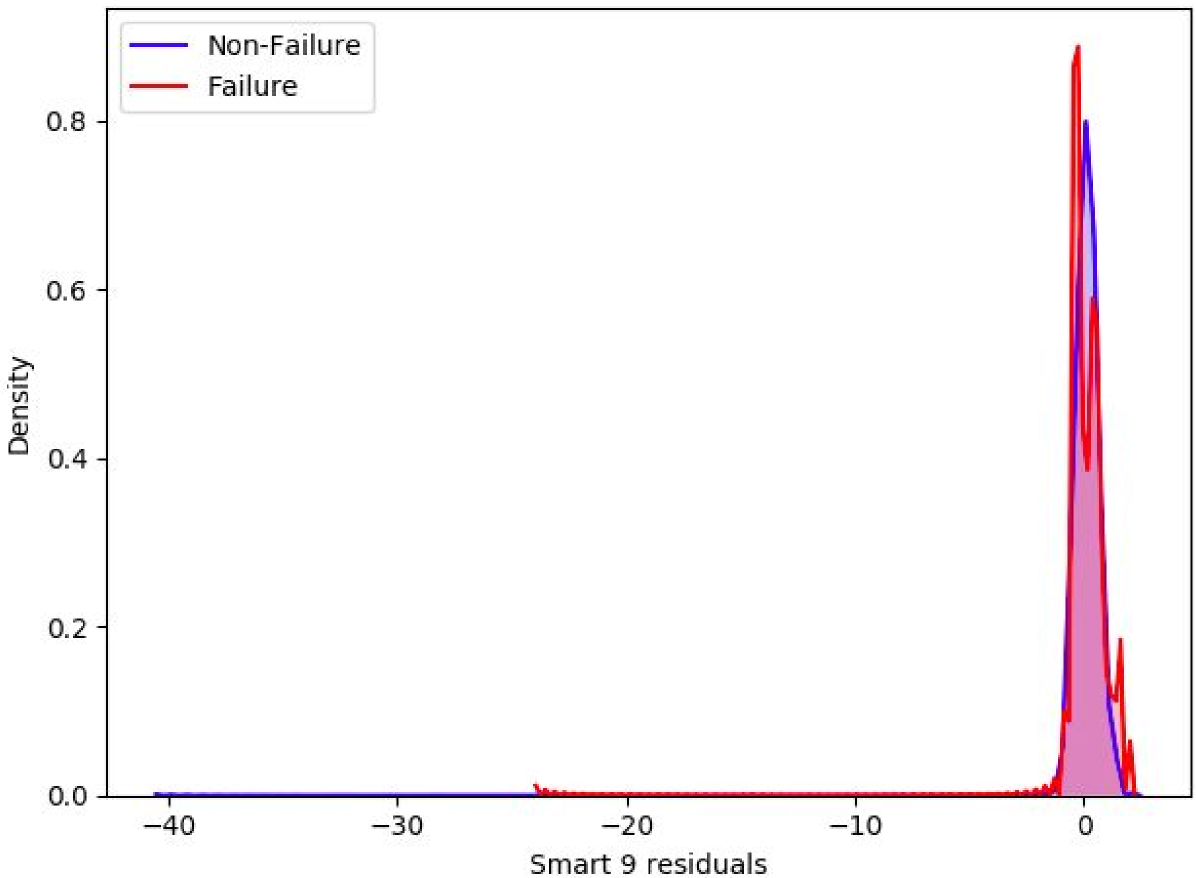
Residual Plots after SMOTE



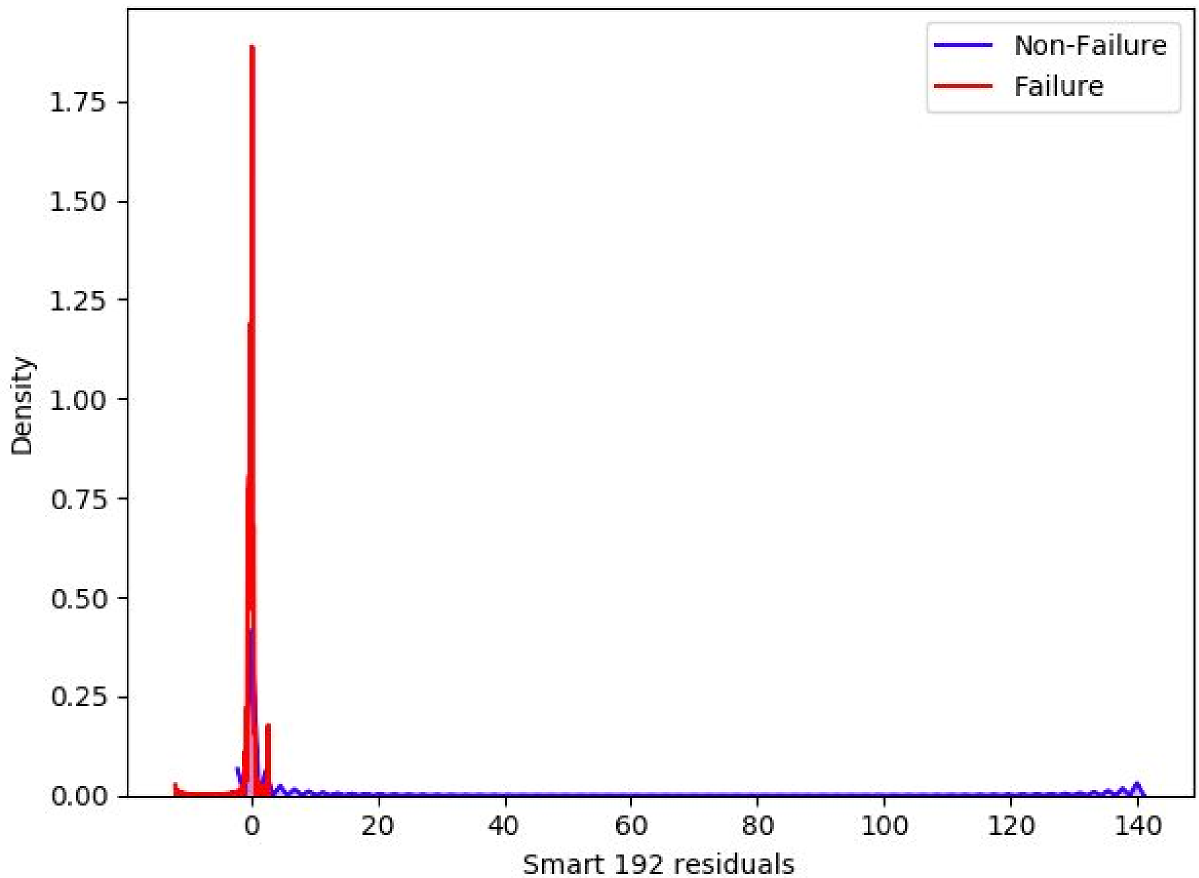
Exploring estimated features from Autoencoders



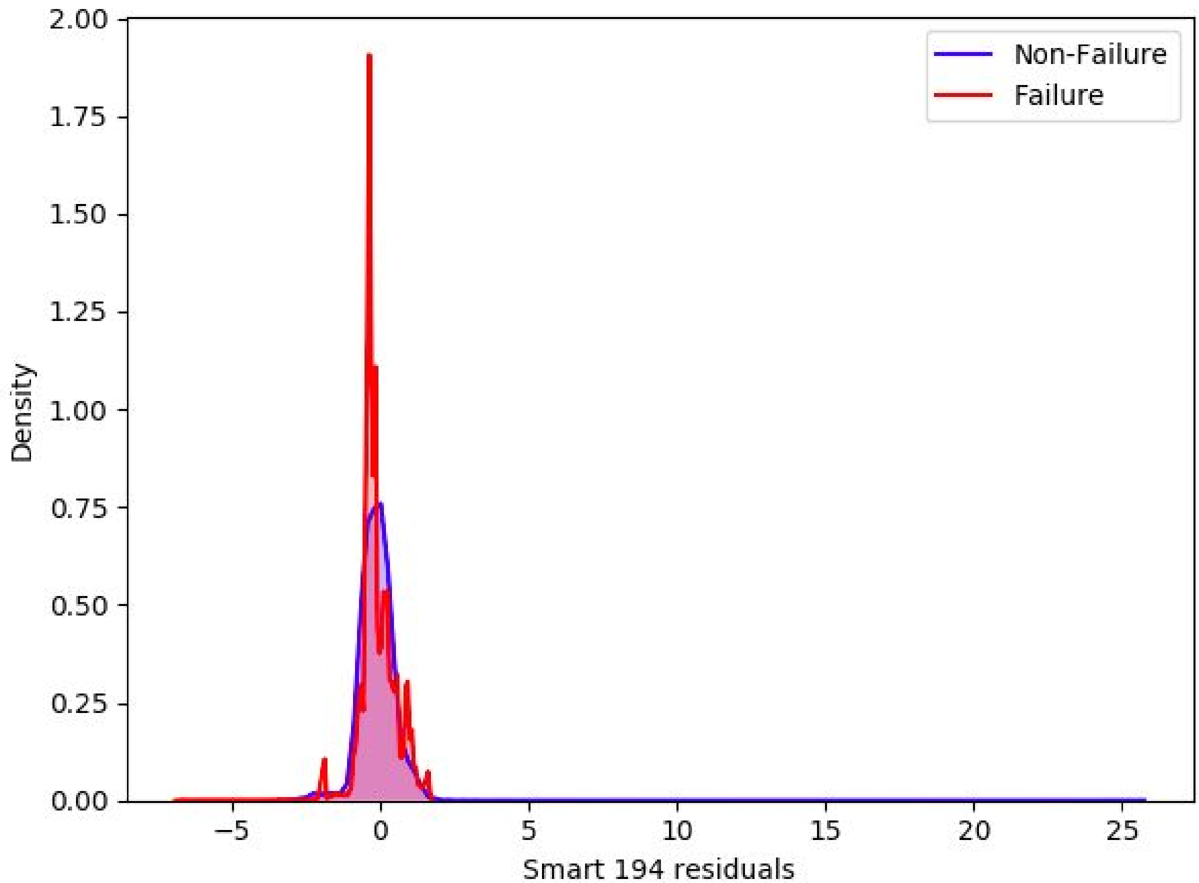
Exploring estimated features from Autoencoders



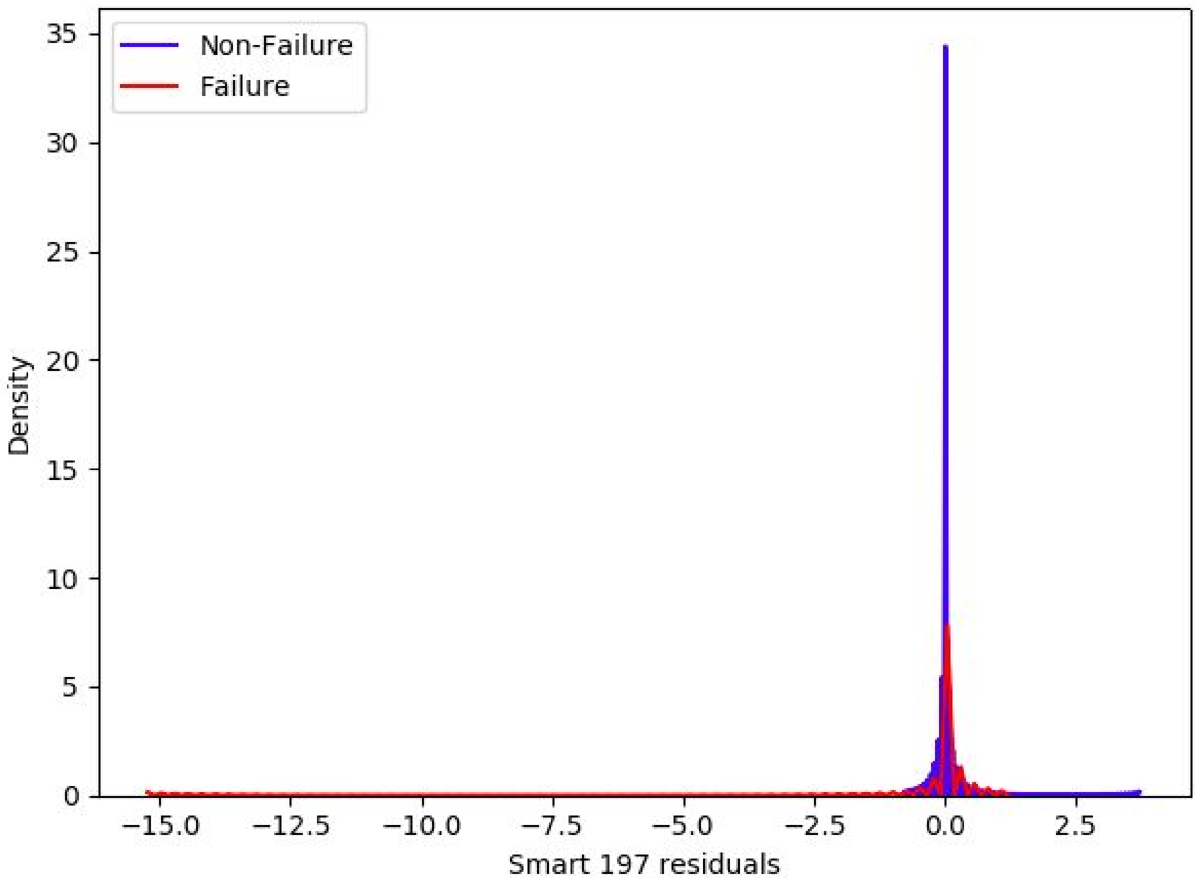
Exploring estimated features from Autoencoders



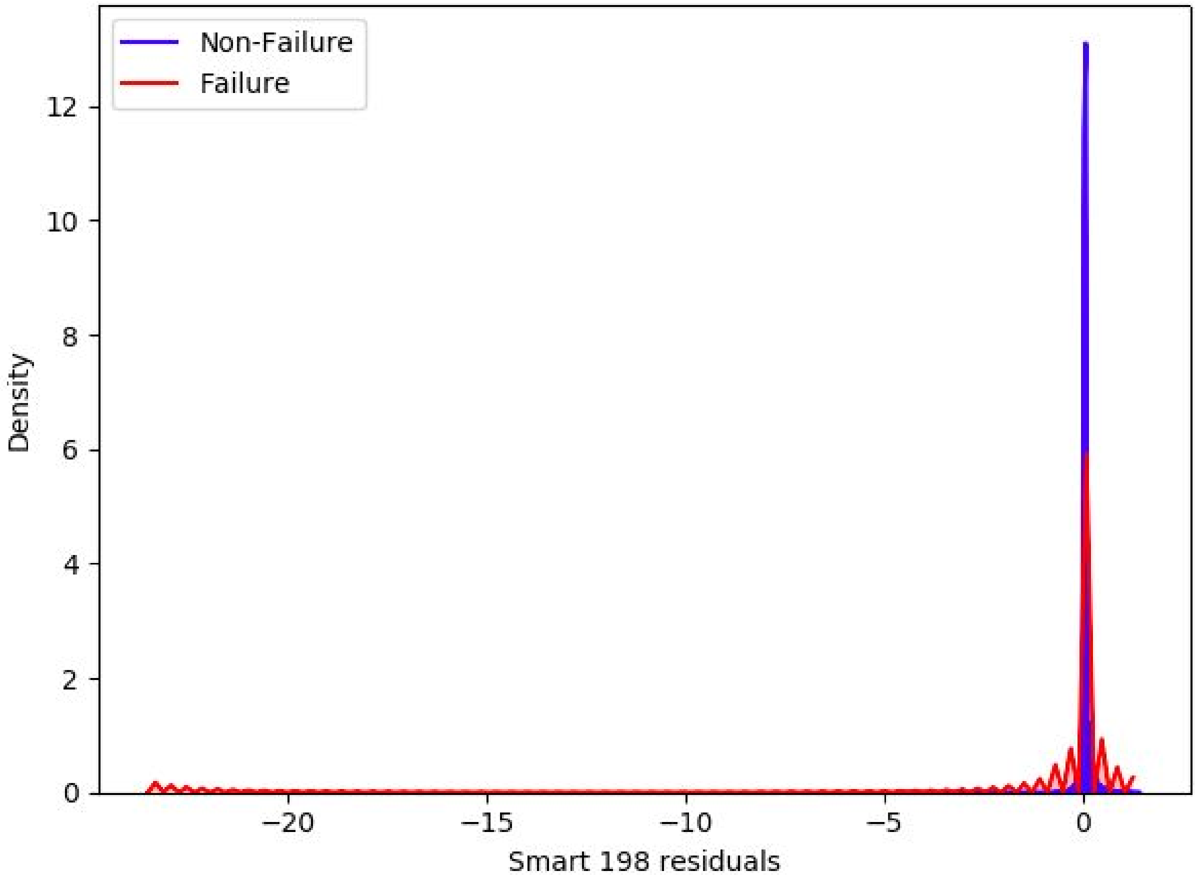
Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders



Random Forest Results

Random Forest without residuals

Exploring estimated features from Autoencoders

Actual **Predicted**	Class-0/ Negative	Class-1/ Positive (Failure)
Class-0/ Negative	135092 (TN)	17224 (FN / Missed Alarms)
Class-1/ Positive (Failure)	11864 (FP/ False Alarms)	130034 (TP)

Precision = TP/(TP+FP) = 130034/(130034+11864) = 0.9164

Recall = TP/(TP+FN) = 130034/(130034+17224) = 0.8830

Random Forest Results with residuals

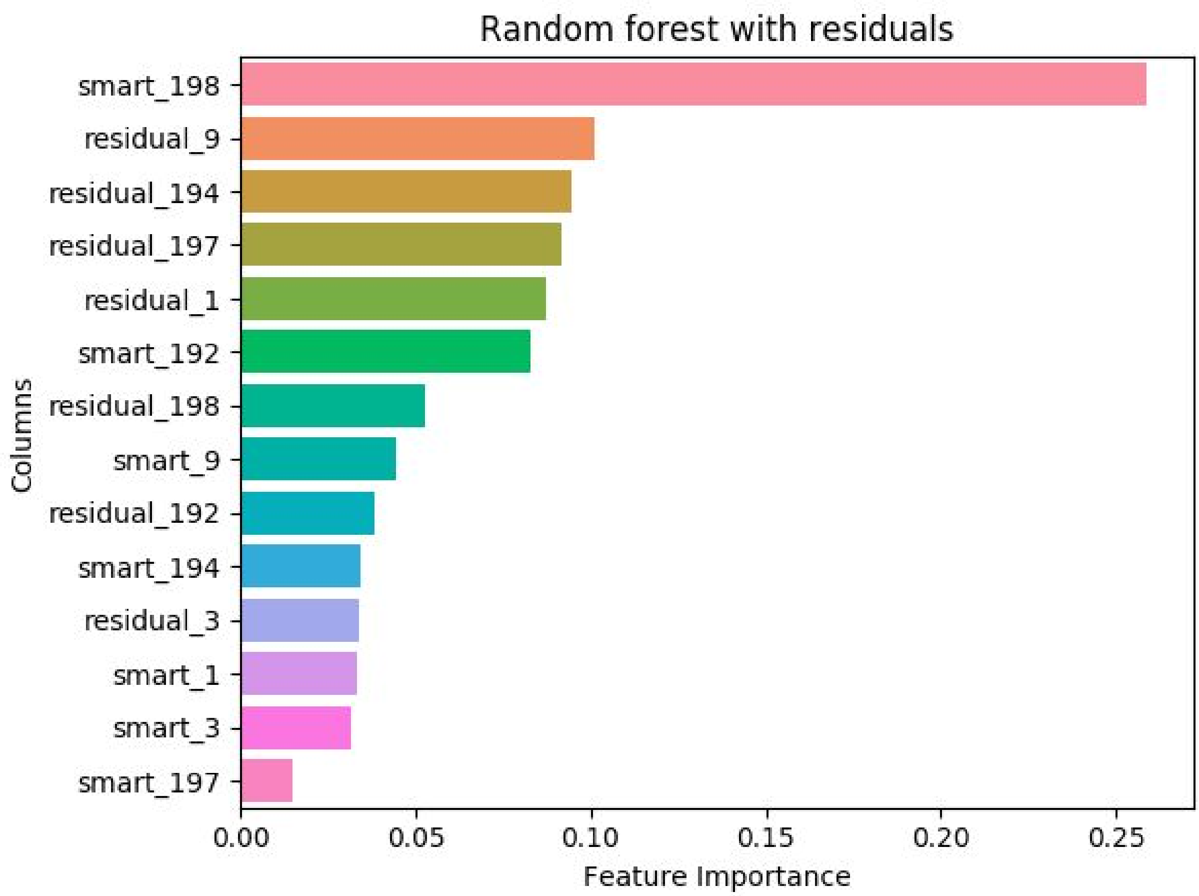
Actual **Predicted**	Class-0/ Negative	Class-1/ Positive (Failure)
Class-0/ Negative	131674 (TN)	13868(FN / Missed Alarms)
Class-1/ Positive (Failure)	15282 (FP/ False Alarms)	133390 (TP)

Precision = TP/(TP+FP) = 133390/(133390+15282) = 0.8972

Recall = TP/(TP+FN) = 133390/(133390+13868) = 0.9058

Again, we can observe that recall has increased by 3% at the cost of precision. Moreover, missed alarms have decreased which is very important.

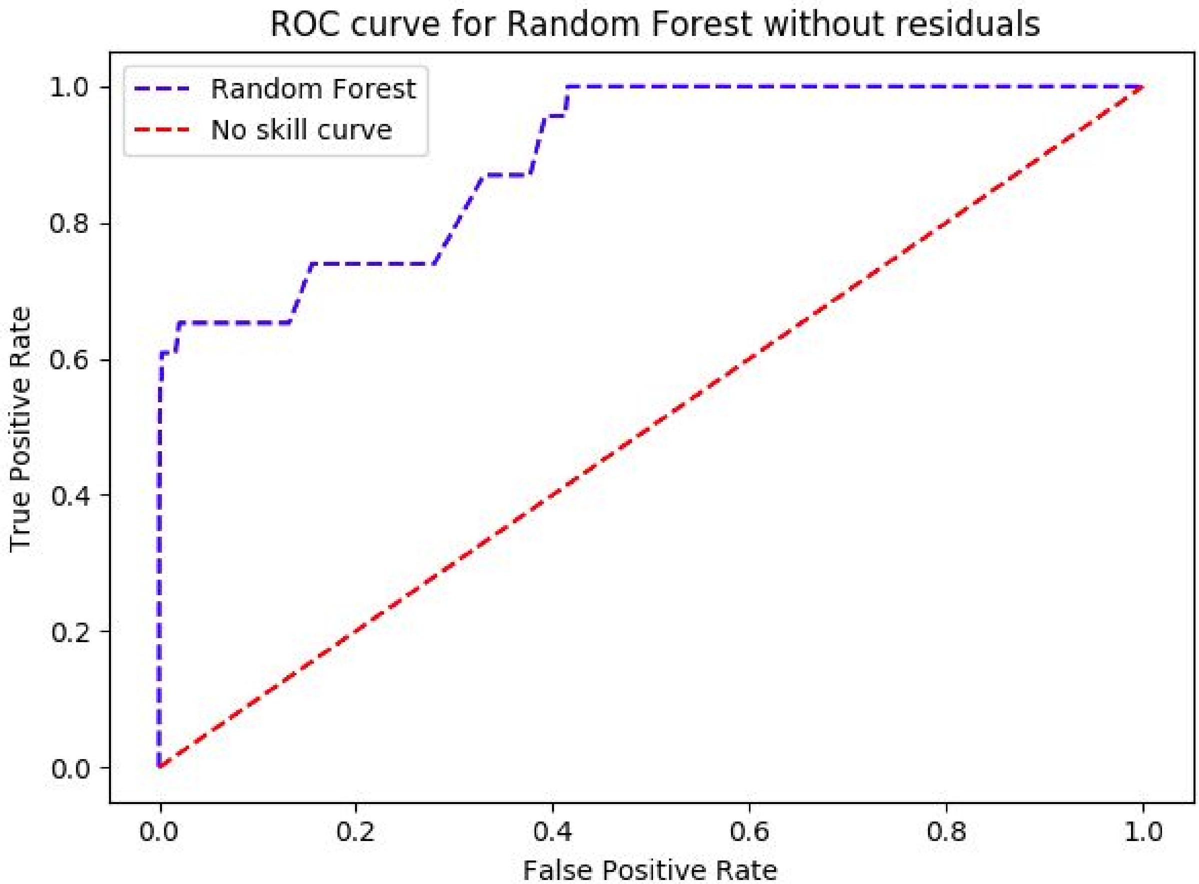
Feature Importance Plot



ROC Curves

Below plot is ROC curve of Random Forest after feeding residuals

Exploring estimated features from Autoencoders



Exploring estimated features from Autoencoders