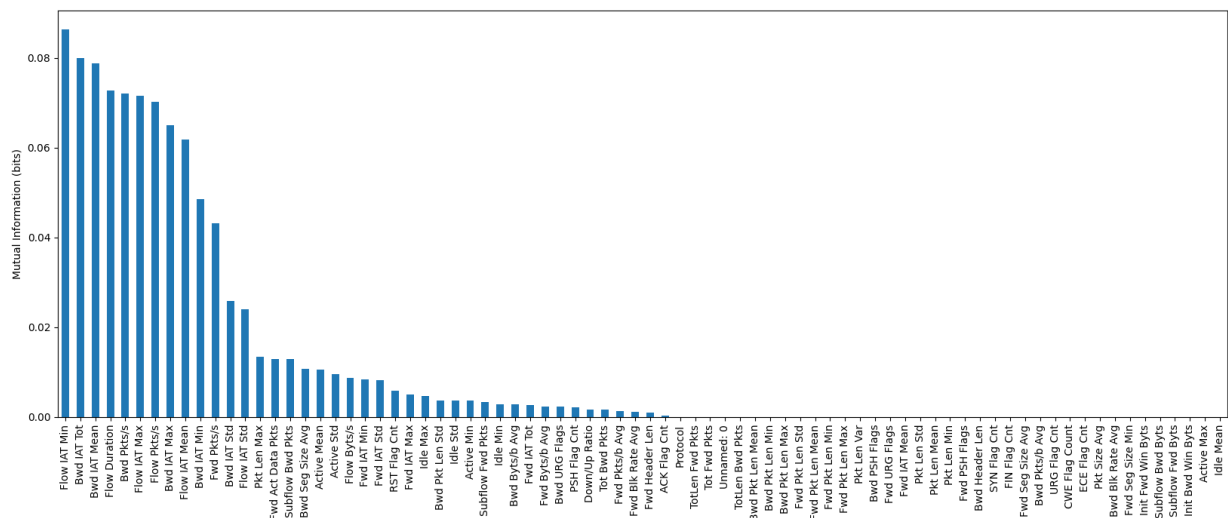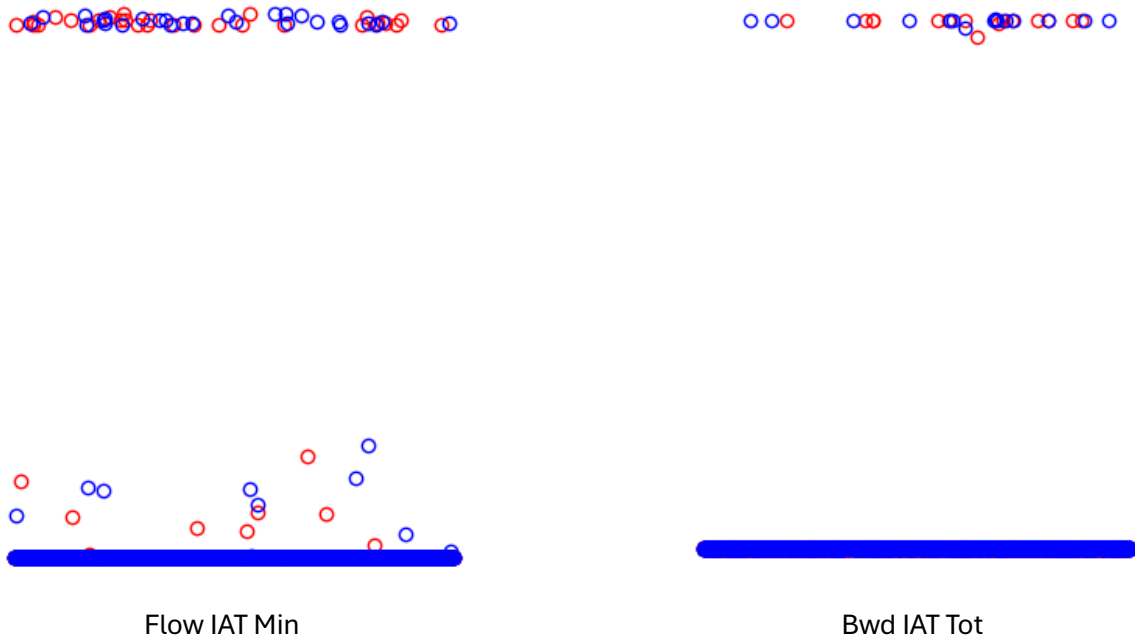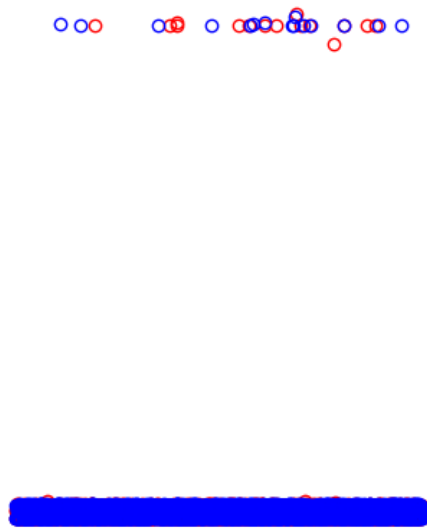# HDGM Features

The below Figure is the feature importance for the HDGM data calculated using the sklearn *mutual_info_classif* function. The CSV file contains 78 features all of which are numerical. This is the dataset with the highest complexity and as expected has the lowest mutual information across all features.
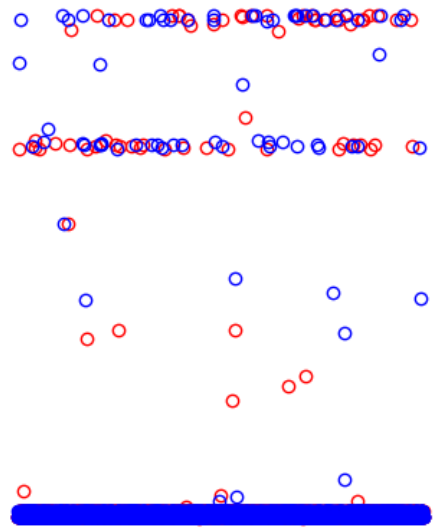

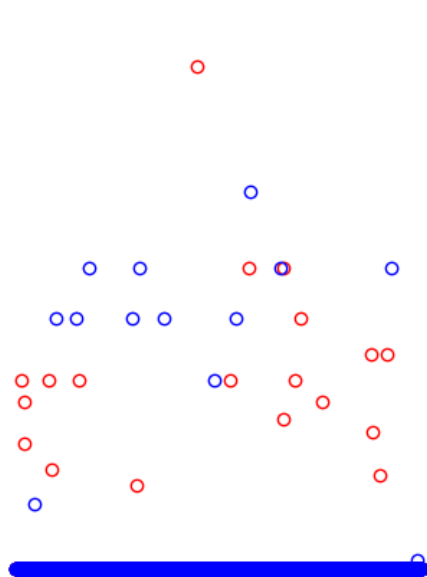The figures below show the individual feature plots for the 10 most important features.
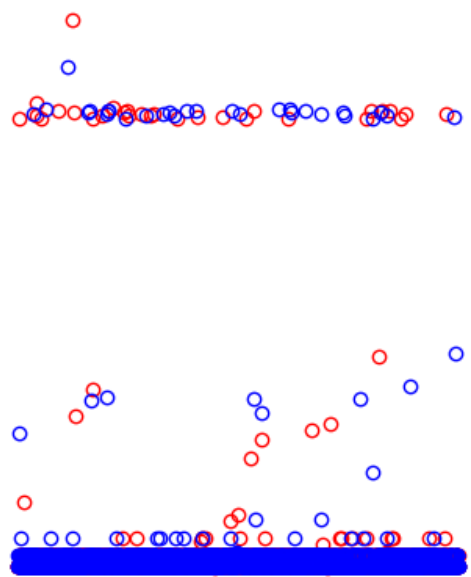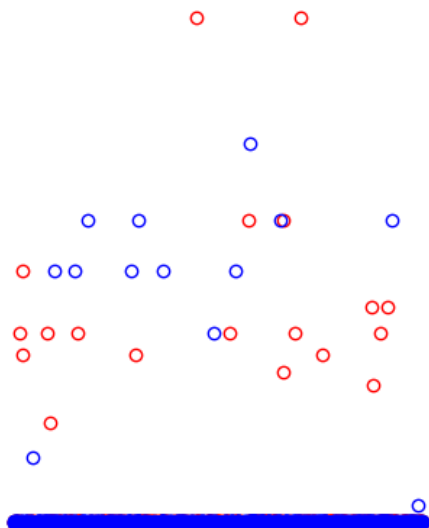

Flow IAT Min
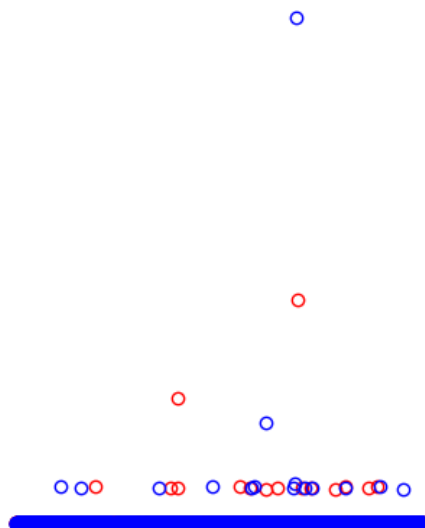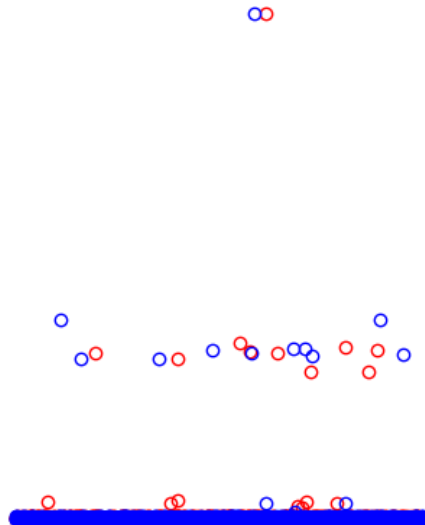
Bwd IAT Tot

Bwd IAT Mean

Flow Duration

Bwd Pkts/s

Flow IAT Max

Flow Pkts/s

Bwd IAT Max

Flow IAT Mean

Bwd IAT Min

The Figure below shows the cross correlation heatmap. We have dropped columns where the feature has a constant value throughout the data