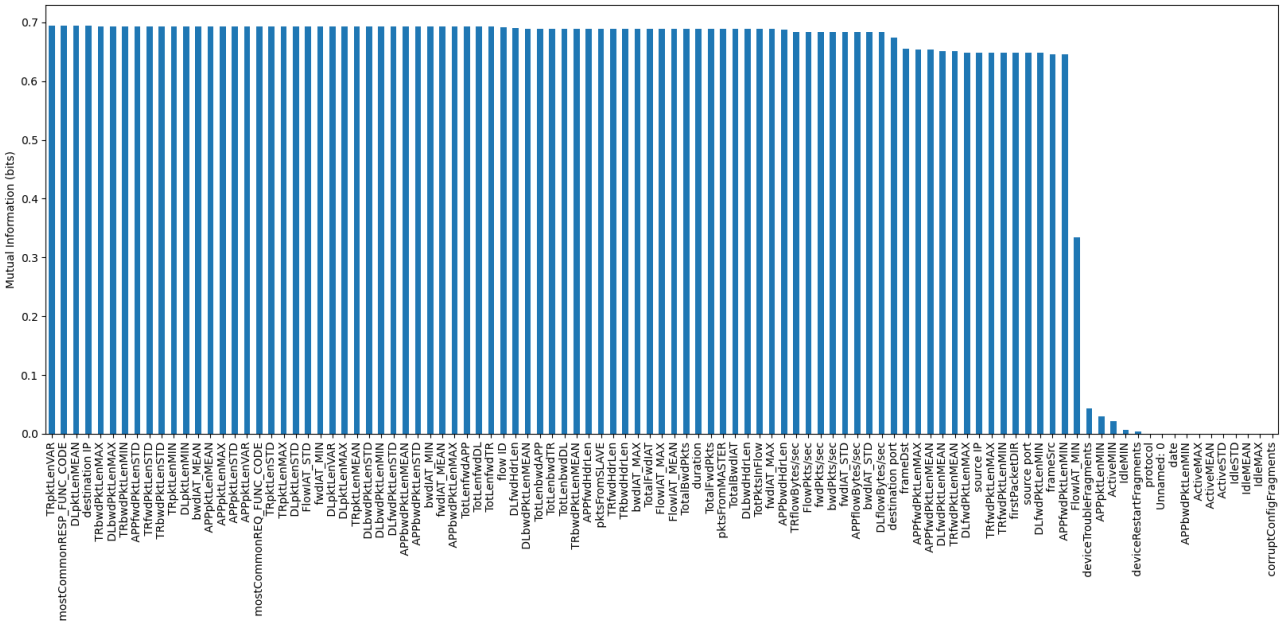
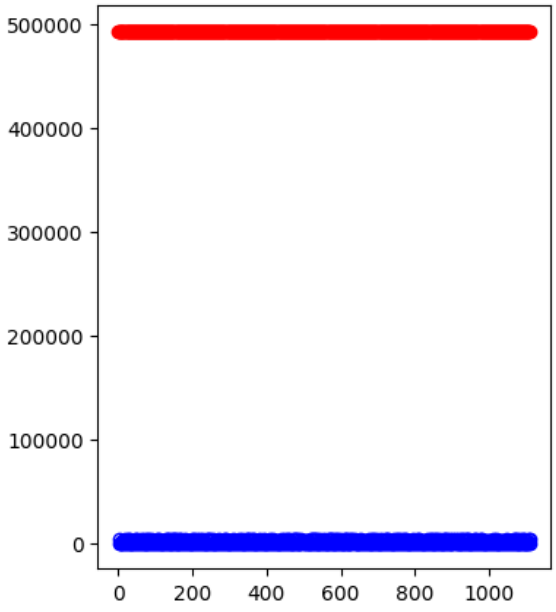
DNP3 Features

The dataset features below were calculated from the "*20200516_DNP3_info_UOWM_DNP3_Dataset_Slave_02_pcap60DNP3_FLOWLABELED.csv*" file in the dataset

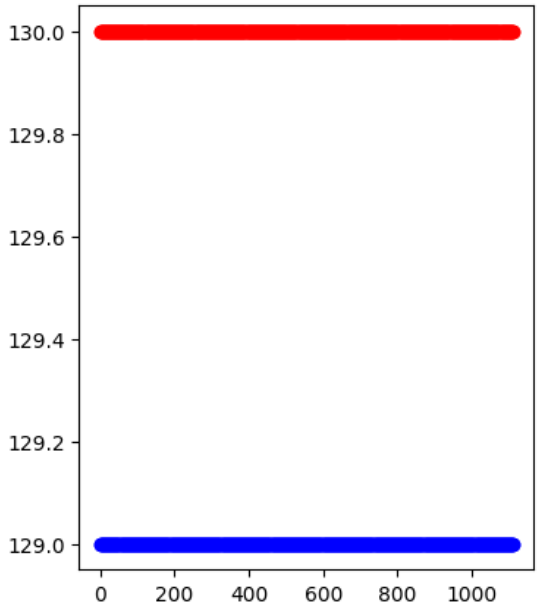
The below Figure is the feature importance for the HDGM data calculated using the sklearn *mutual_info_classif* function. Note that there may be some discrepancy between this and Figure 4 in out paper. This is because we did not force the *random_state* parameter in the *mutual_info_classif* function to *int* so that the function used random number generation for adding small noise to continuous variables in order to remove repeated values. The CSV file contains 102 features 97 of which are numerical. This is the dataset with the lowest complexity and as expected has the highest mutual information across all features.



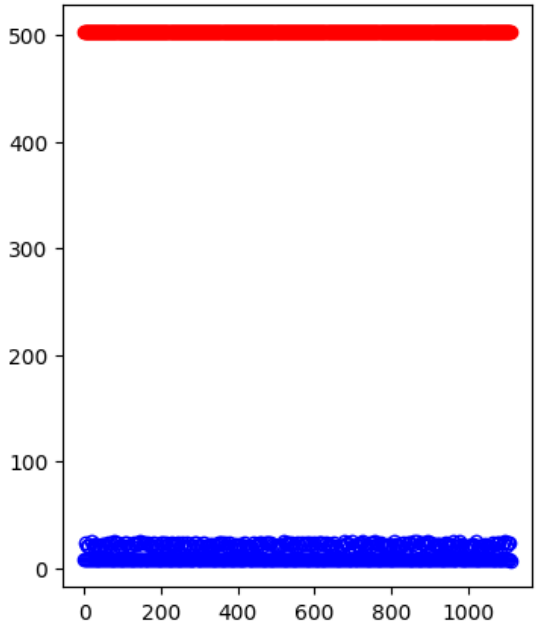
The figures below show the individual feature plots for the 10 most important features.



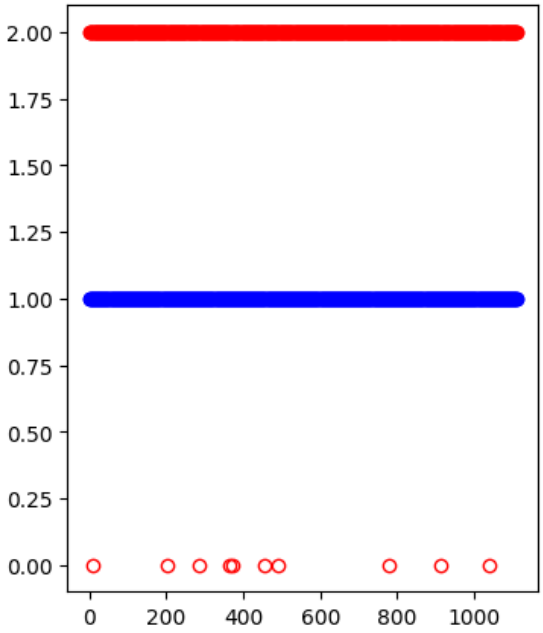
TRpktLenVAR



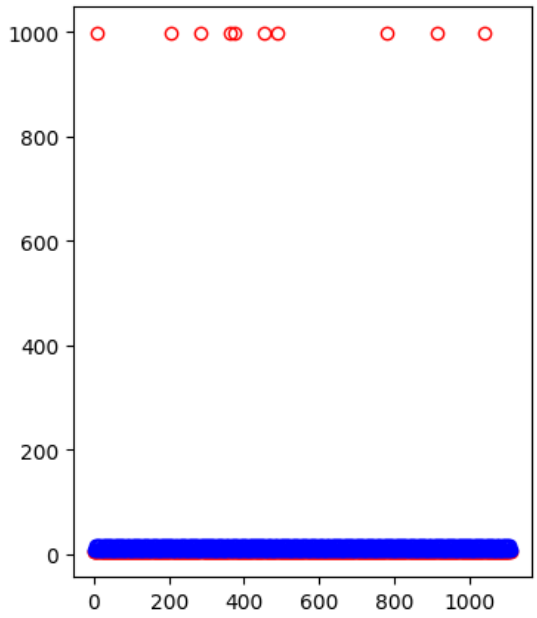
mostCommonRESP_FUNC_CODE



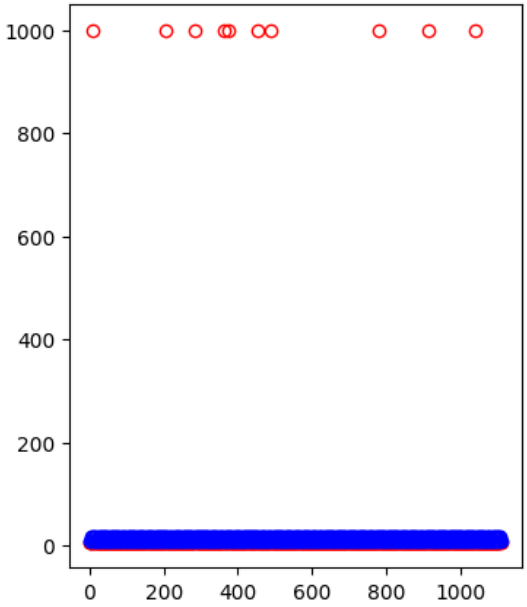
DLpktLenMEAN



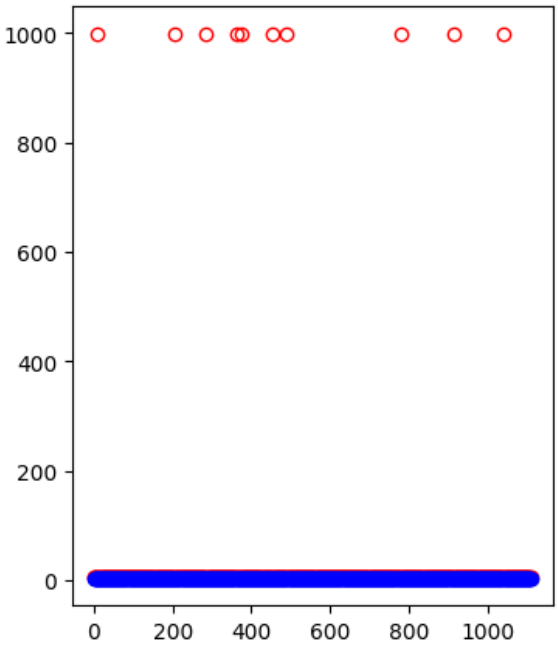
destination IP



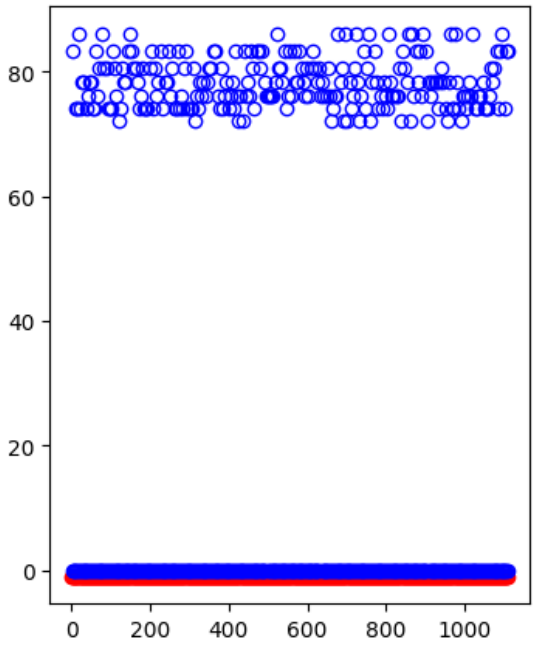
TRbwdPktLenMAX



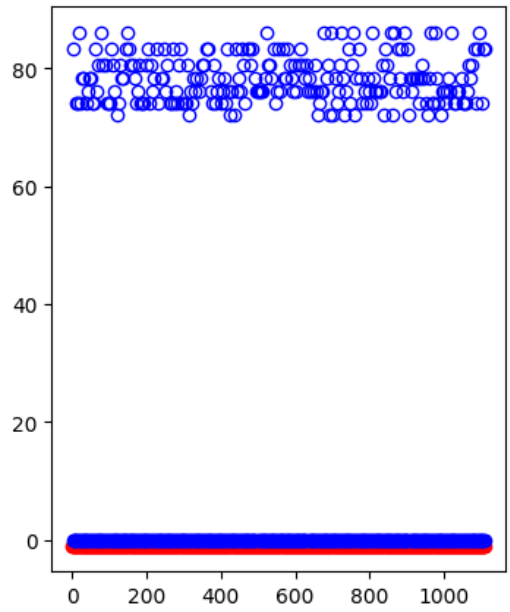
DLbwdPktLenMAX



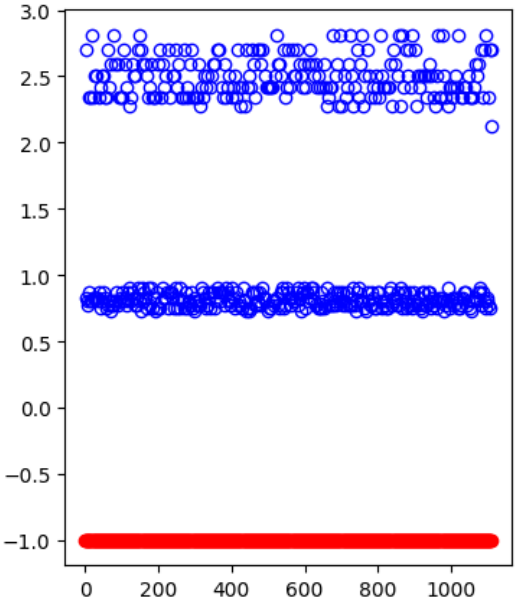
TRbwdPktLenMIN



APPfwdPktLenSTD



TRfwdPktLenSTD



TRbwdPktLenSTD

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