Data Cleaning

We first ignored all the samples for shots that were intentionally disrupted. Next, we dropped the following features from the train dataset due to large number of missing values (more than 30%):
mirnov
te_width
z_error
z_times_v_z
zcur
v_z

To deal with the remaining missing values, we dropped the corresponding samples from the data. We also dropped the time and intentional_disrupt columns as they are not relevant to measuring the predictability.

Two-fold Model

We used an LSTM with a lookback window of 10 timesteps to classify the samples as disrupted or non-disrupted. Then, we use a random forest regressor to predict the time_until_disrupt for the samples that are classified as disrupted by the LSTM model.