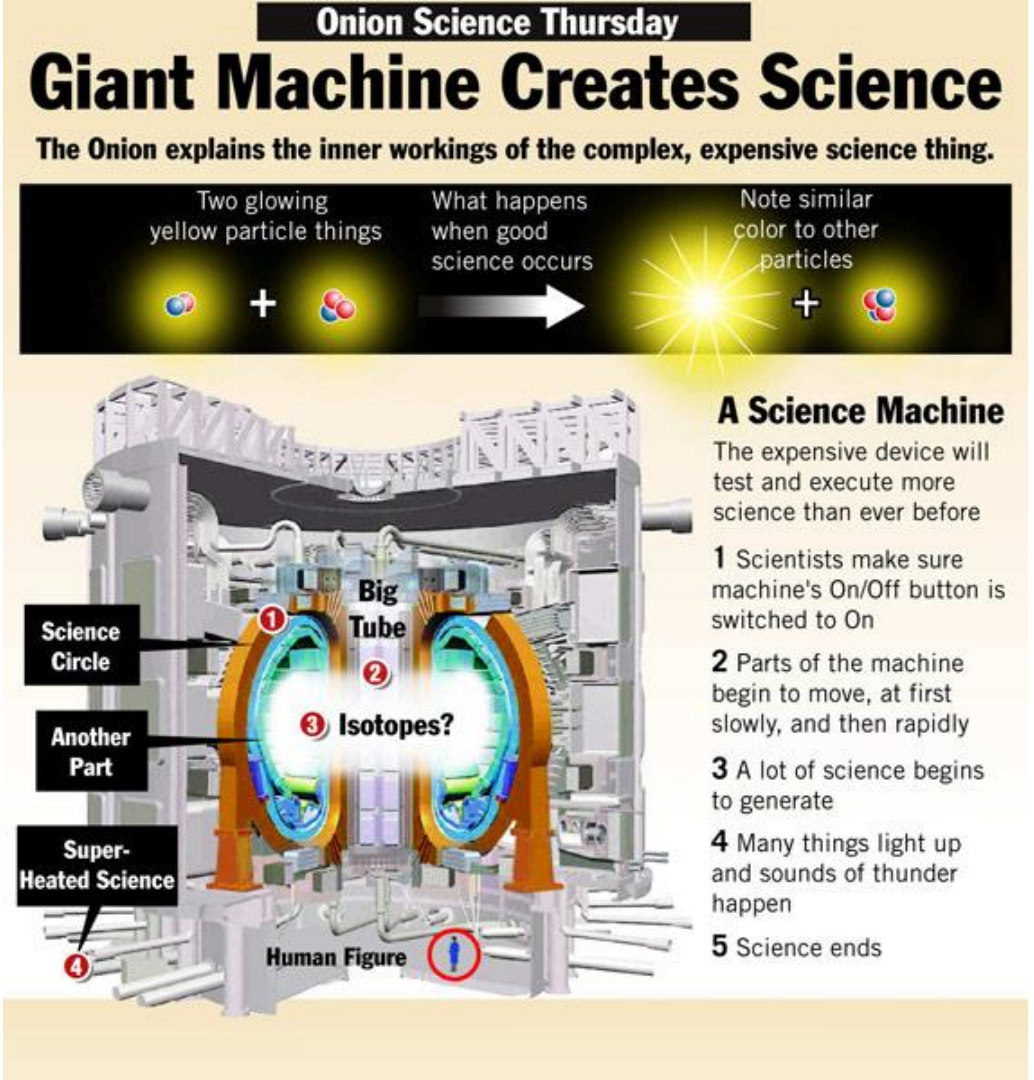
Detecting JET Outliers in the Tokamak Database

Dan Segal
@djsegal

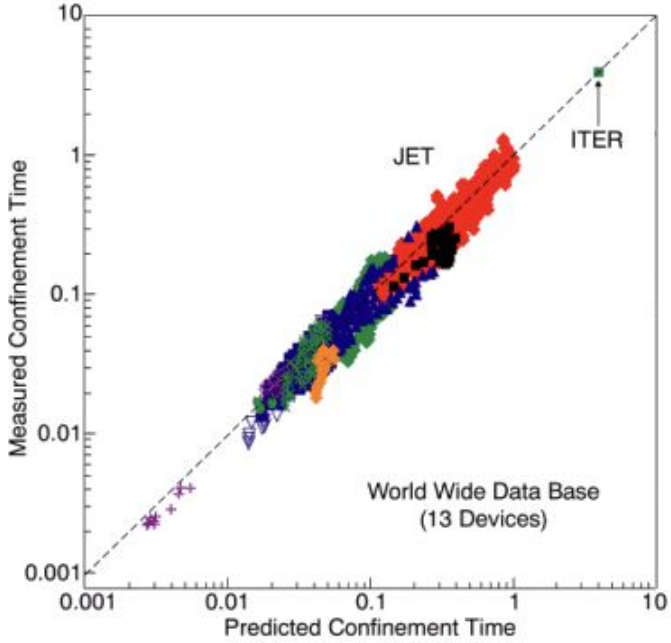
Tokamaks will be fusion reactors that compete with baseload power sources – like coal

ITER, at $20B, is the second most costly experiment in the world



ITER was built using a linear regression on confinement:

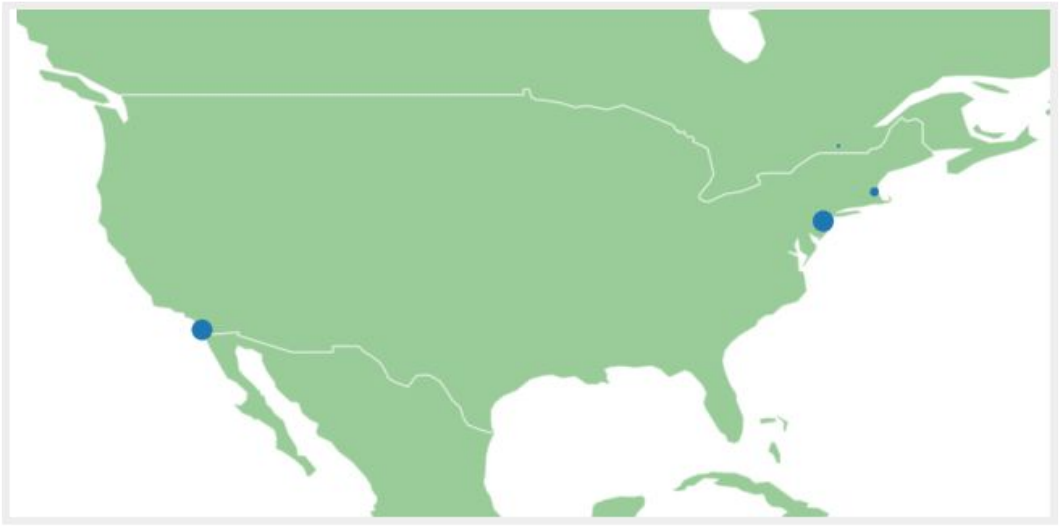
$$\tau_E^H = 0.145\, H\, \frac{I_P^{0.93}\, R_0^{1.39}\, a^{0.58}\, \kappa^{0.78}\, \overline{n}^{0.41}\, B_0^{0.15}\, A^{0.19}}{P_{src}^{0.69}}$$



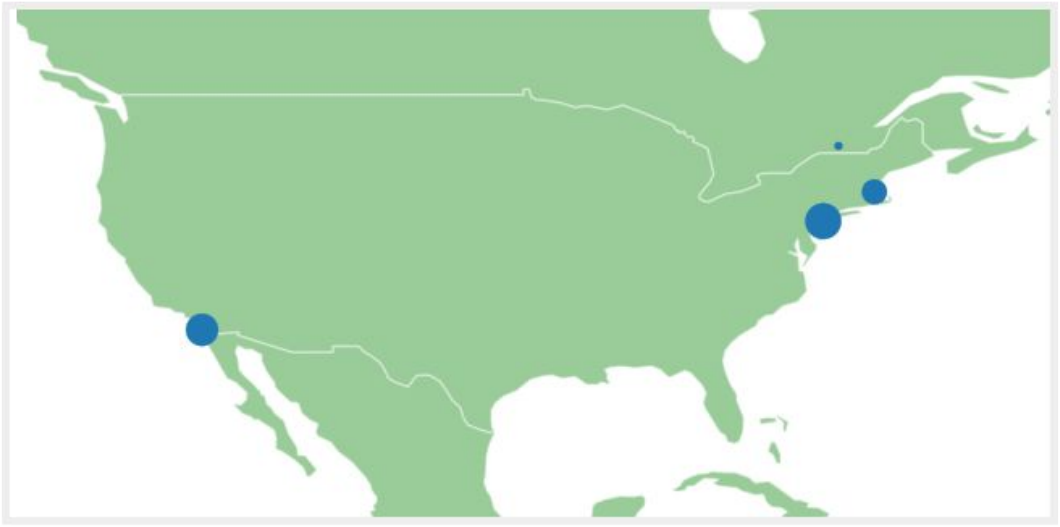
Our goal is to detect JET outliers

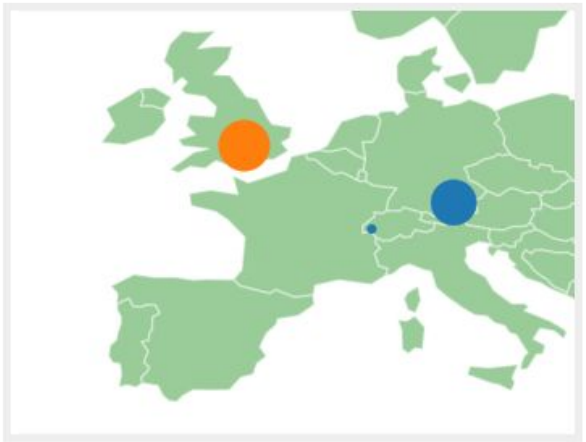
	Good	Bad
JET	**1434**	1710
Else	2097	913

The first step is balancing shots from all the tokamaks

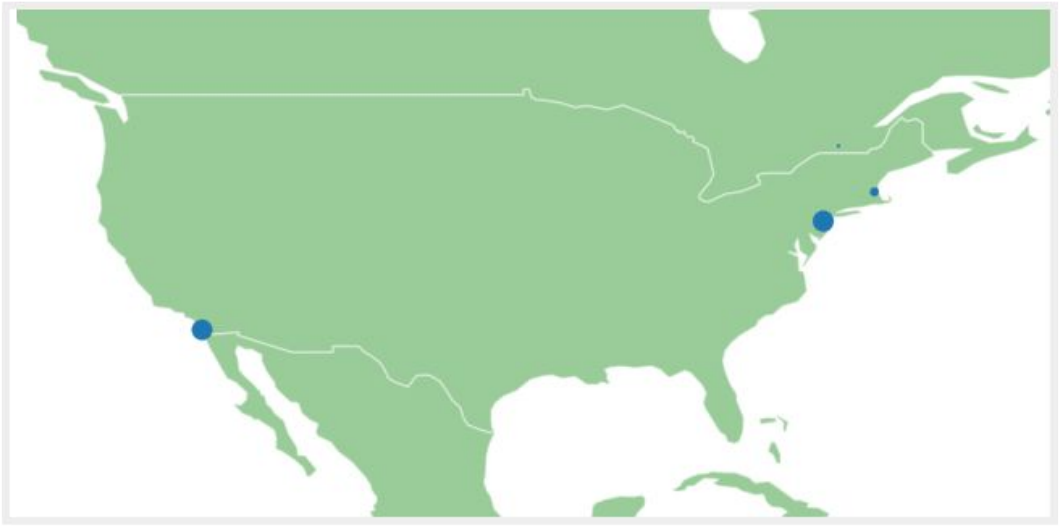


The first step is balancing shots from all the tokamaks



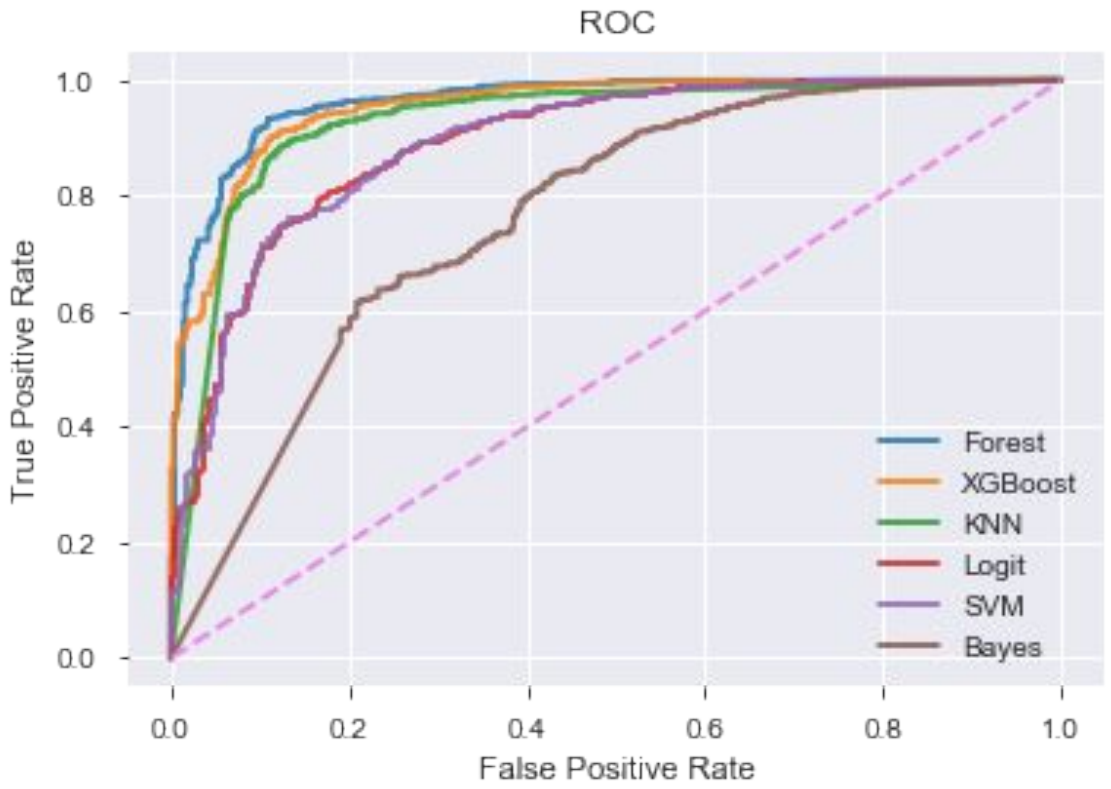


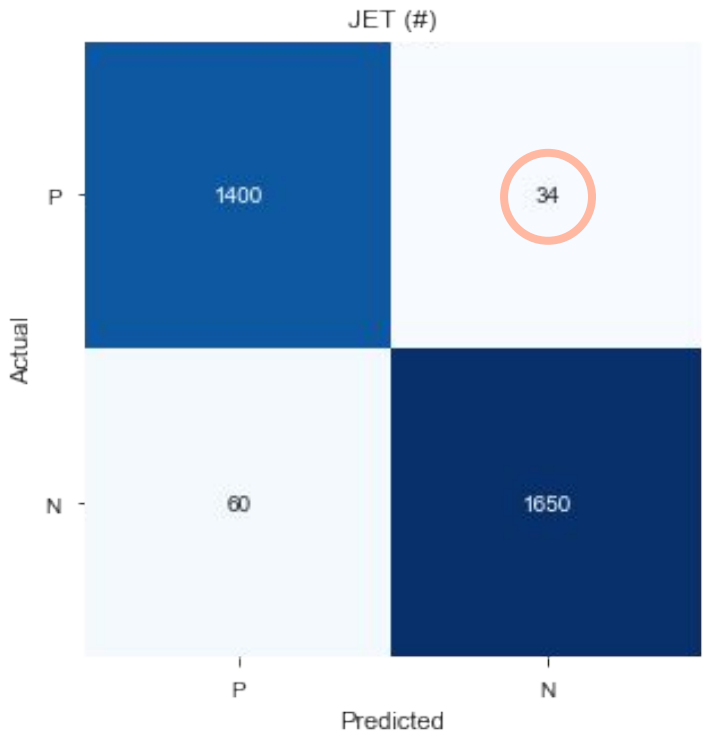
The first step is balancing shots from all the tokamaks

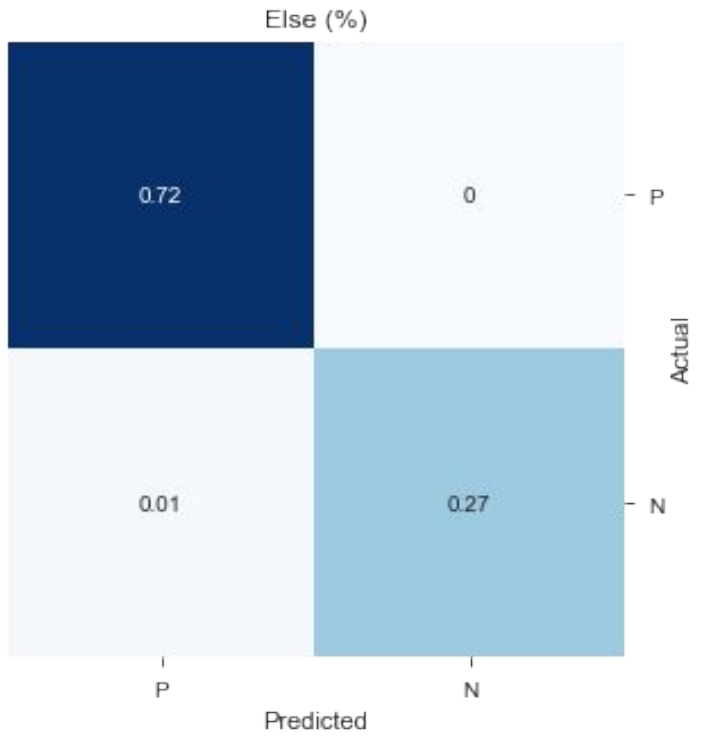


Next we will plug 10k rows into 6 classifiers using the 50 strongest features

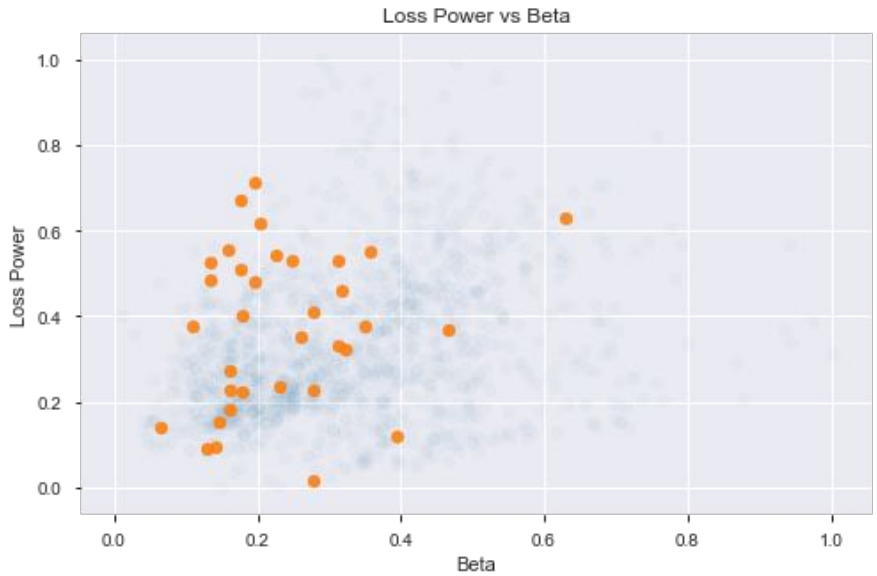
Model	AUC (%)
Forest	96
XGBoost	95
KNN	93
Logit	89
SVM	89
Bayes	*77*




JET (#)
P
N
1400
34
60
1650
Actual
P
N
Predicted

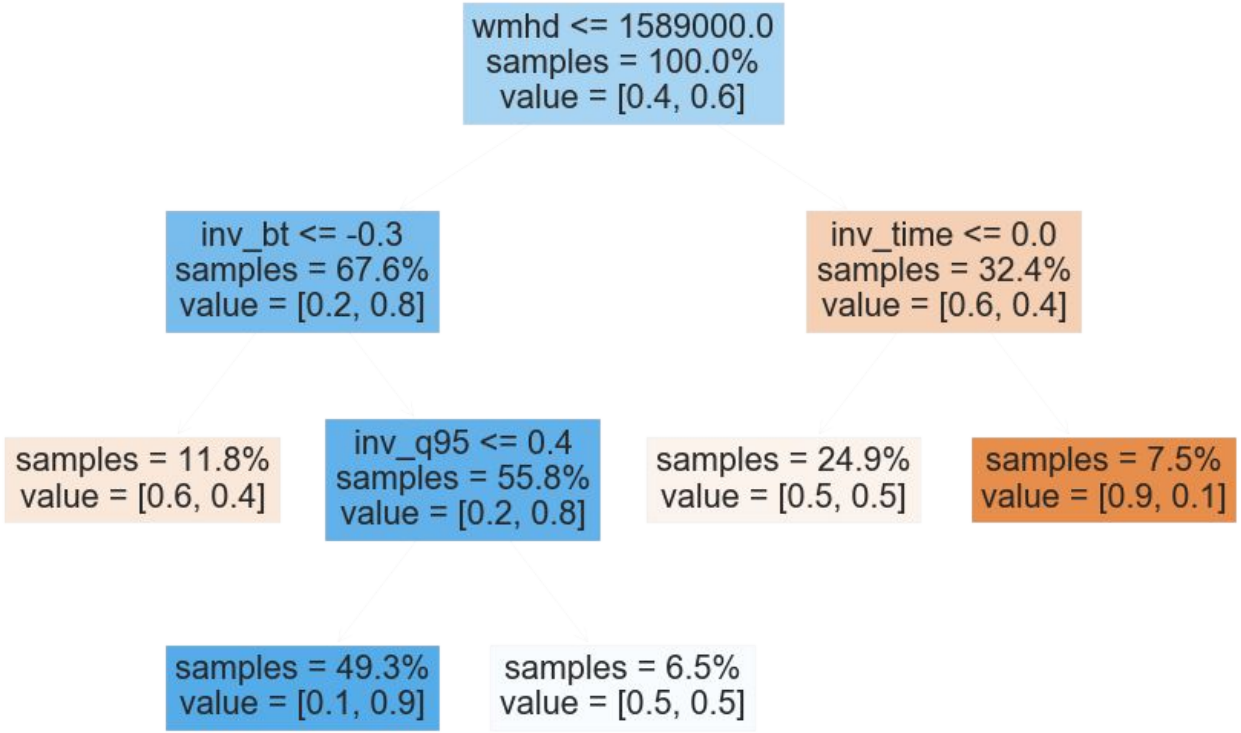

Else (%)
0.72
0
0.01
0.27
P
N
Actual
P
N
Predicted

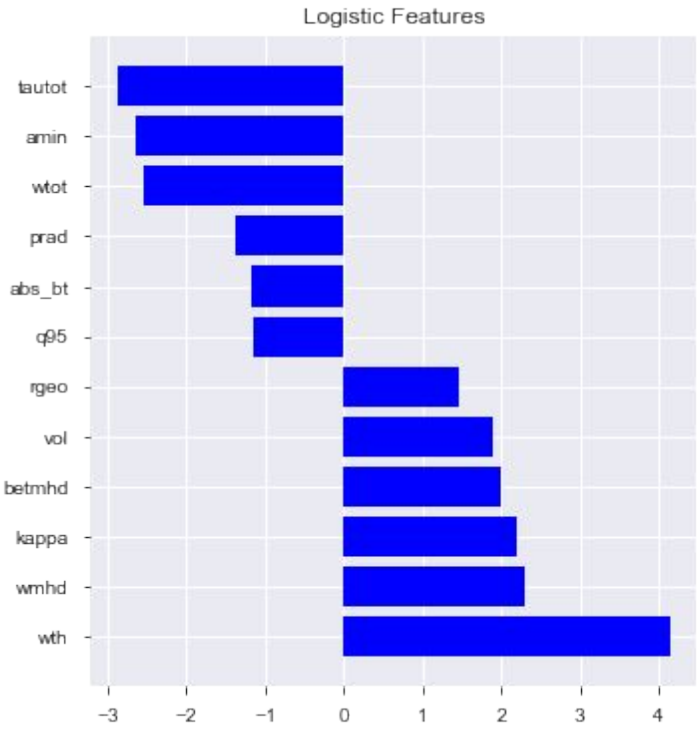
The 34 outliers tend to have low betas and high loss powers

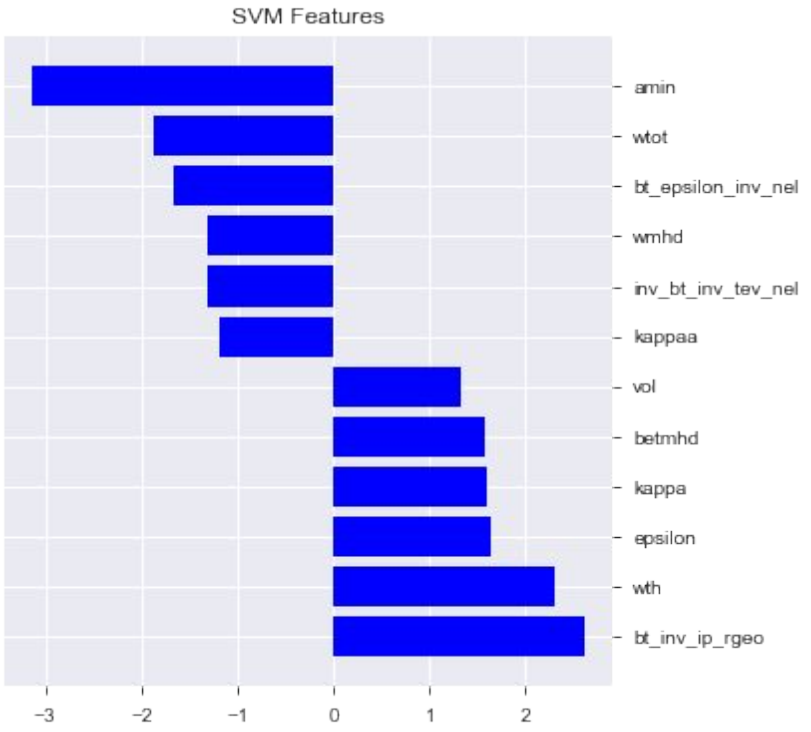


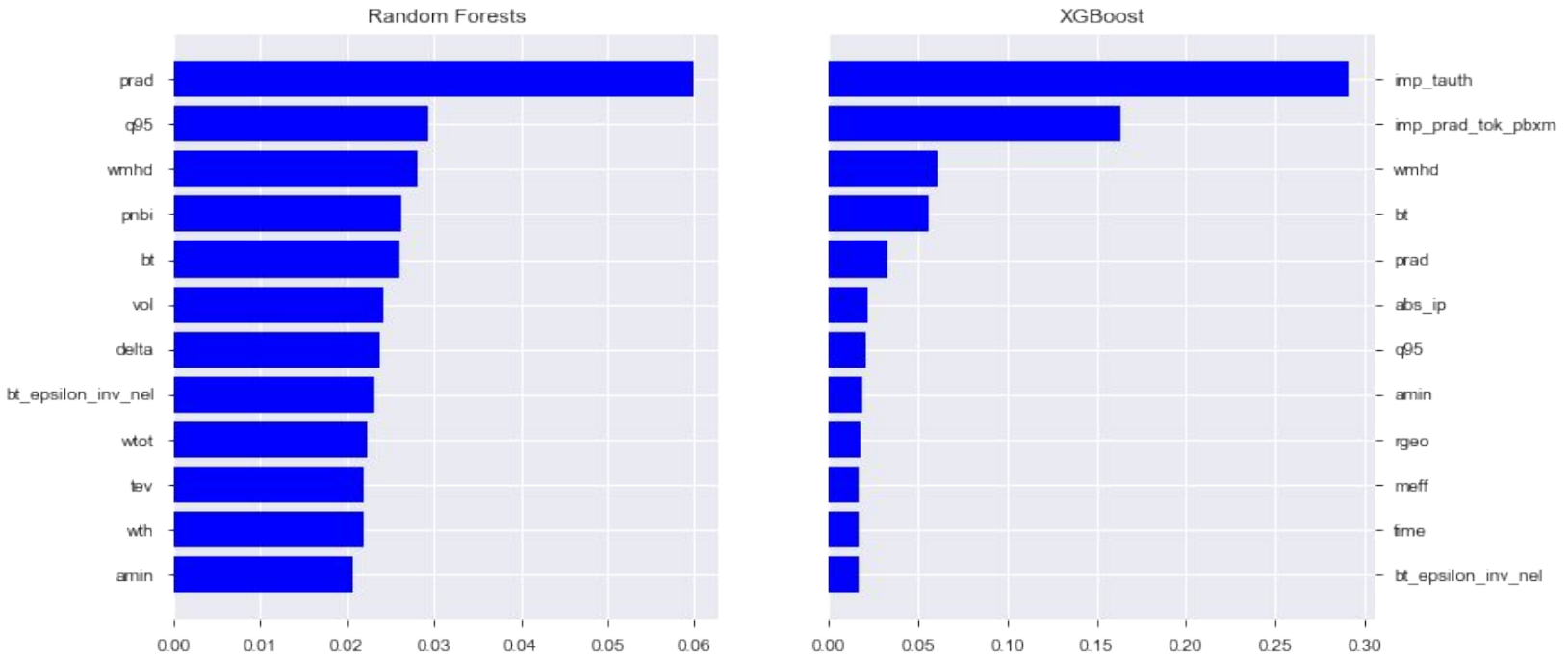
Features

- Started with ~100 numeric parameters
- Imputed NaNs with median or predicted
- Expanded to ~1000 parameters w/ polynomials
- Removed ~500 highly correlated variables
- Eliminated ~200 variables with random forests
- Dwindled to ~50 variables with L1 classifiers


wmhd <= 1589000.0
samples = 100.0%
value = [0.4, 0.6]
inv_bt <= -0.3
samples = 67.6%
value = [0.2, 0.8]
inv_time <= 0.0
samples = 32.4%
value = [0.6, 0.4]
samples = 11.8%
value = [0.6, 0.4]
inv_q95 <= 0.4
samples = 55.8%
value = [0.2, 0.8]
samples = 24.9%
value = [0.5, 0.5]
samples = 7.5%
value = [0.9, 0.1]
samples = 49.3%
value = [0.1, 0.9]
samples = 6.5%
value = [0.5, 0.5]


Logistic Features
tautot
amin
wtot
prad
abs_bt
q95
rgeo
vol
betmhd
kappa
wmhd
wth
−3
−2
−1
0
1
2
3
4


SVM Features
amin
wtot
bt_epsilon_inv_nel
wmhd
inv_bt_inv_tev_nel
kappaa
vol
betmhd
kappa
epsilon
wth
bt_inv_ip_rgeo
−3
−2
−1
0
1
2


Random Forests
prad
q95
wmhd
pnbi
bt
vol
delta
bt_epsilon_inv_nel
wtot
tev
wth
amin
0.00
0.01
0.02
0.03
0.04
0.05
0.06
XGBoost
imp_tauth
imp_prad_tok_pbxm
wmhd
bt
prad
abs_ip
q95
amin
rgeo
meff
time
bt_epsilon_inv_nel
0.00
0.05
0.10
0.15
0.20
0.25
0.30