

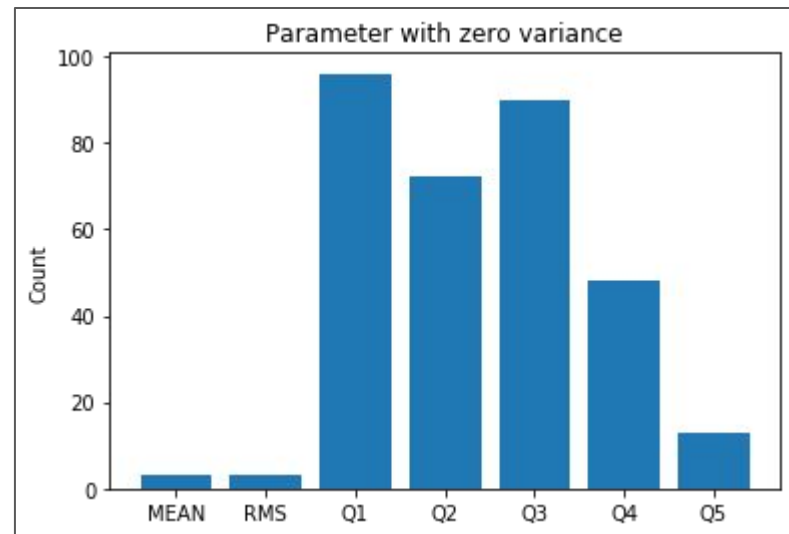
Feature Analysis

Mantas Stankevičius

Feature variance

Zero variance

325 features (~11%)



https://github.com/mantydze/ml4dc/blob/master/src/analysis/feature_variance.ipynb

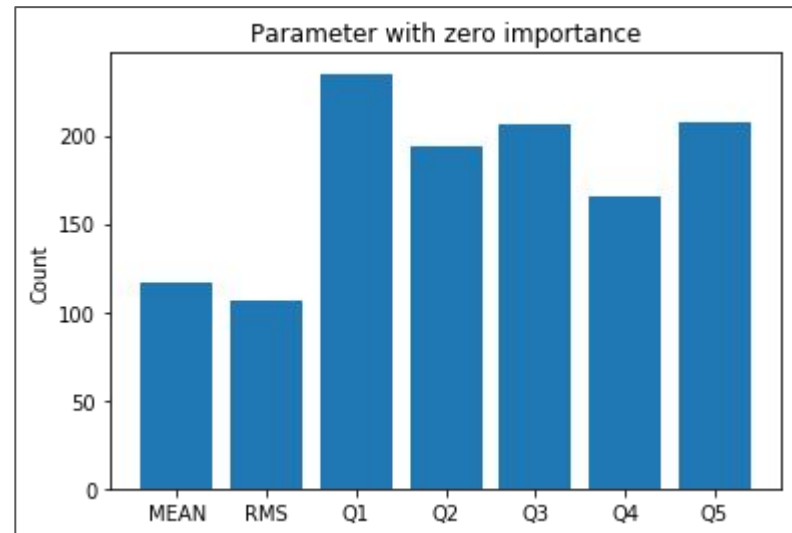
Feature importance (Random Forest)

Random forest model

Mean ROC AUC 0.971

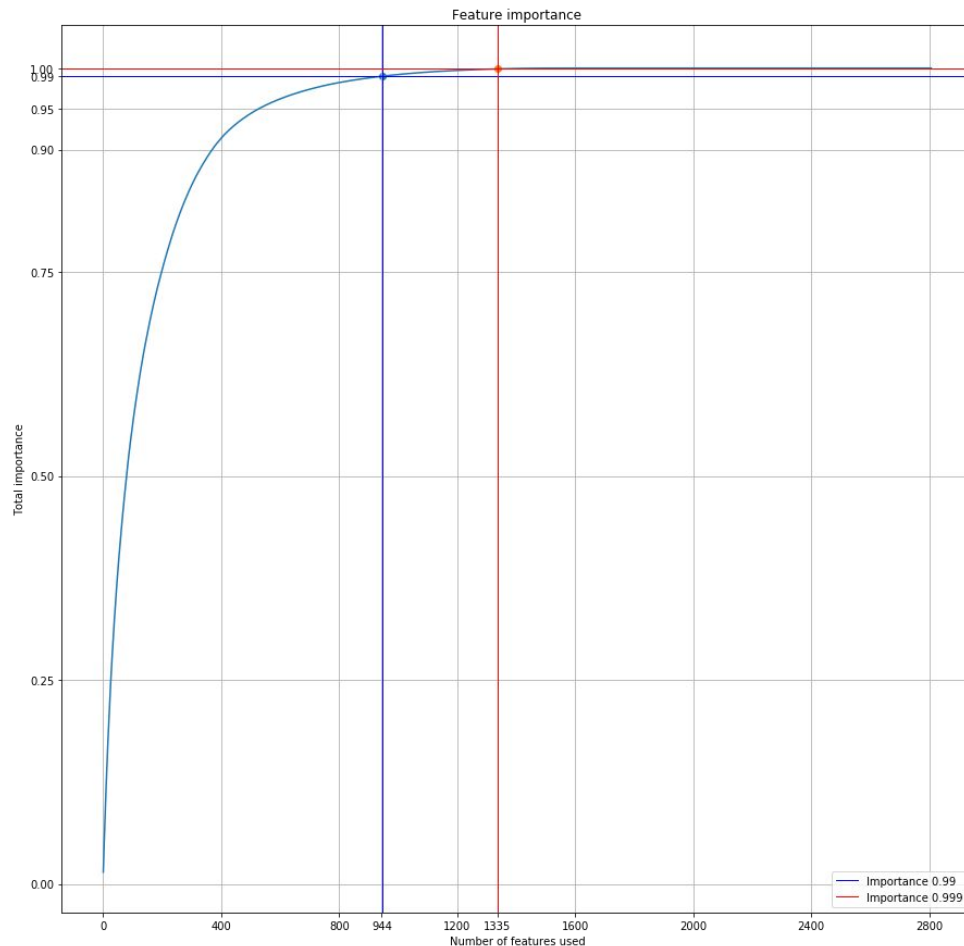
Zero importance

1232 features (~43%)



Feature importance (random forest)

Total importance %	# of features
0.99	945
0.999	1335



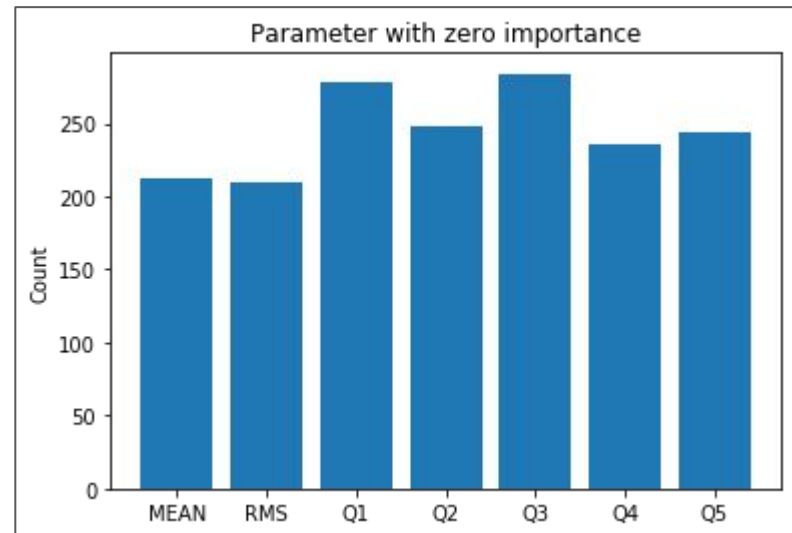
Feature importance (XGBoost)

Random forest model

Mean ROC AUC 0.987

Zero importance

1712 features (~61%)



https://github.com/mantydze/ml4dc/blob/master/src/analysis/xgb_feature_importance.ipynb

Feature importance (XGBoost)

Total importance %	# of features
0.99	1065
0.999	1092
1.0	1095

