

Warsztaty badawcze 2

Praca domowa 5.

Ada Gąssowska, Elżbieta Jowik

Random Search

Baseline model 1.

	Random Search cv AUCROC scores	Original model cv AUCROC scores
Fold 1 of 5.	0.885	0.880
Fold 2 of 5.	0.887	0.889
Fold 3 of 5.	0.878	0.888
Fold 4 of 5.	0.892	0.886
Fold 5 of 5.	0.894	0.897

```
CV best params: {'xgb__subsample': 0.7, 'xgb__reg_lambda': 0.1,
'xgb__n_estimators': 300, 'xgb__min_child_weight': 3.0,
'xgb__max_depth': 8, 'xgb__learning_rate': 0.1, 'xgb__gamma':
0.5, 'xgb__colsample_bytree': 0.6, 'xgb__colsample_bylevel':
0.9}
```

```
Original params: max_depth=3, n_estimators=300,
learning_rate=0.05
```

Conclusion:

- There are reasons to infer that the models had been optimized by the authors before being used.

Roshomon Sets

- We researched 100 models
- Best mean test score (AUROC) - 0.90367
- For theta 0.002 five models were in the Roshomon Set
- For these 5 models we analysed feature importance
- As seen in the table - the models were all quite good, and the scores did not differ much

Theta	Number of models in Roshomon Set
0.001	2
0.002	5
0.0025	8
0.005	19
0.0075	29
0.01	36
0.025	52
0.05	78

5 best models refit

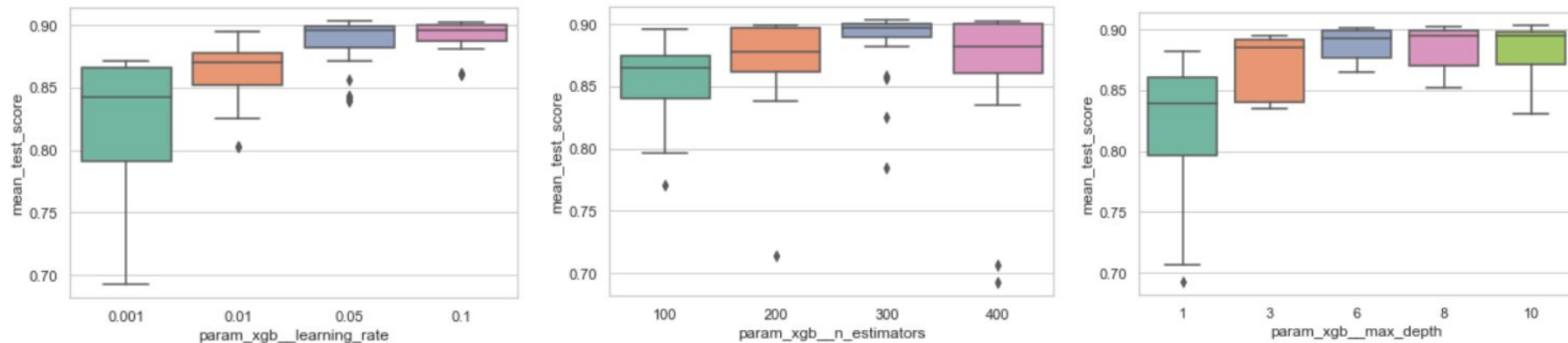
RandomizedSearchCV discards the actual models it trains on each fold after evaluating them, so we were not able to extract the most important features for each model fitted in the process. Instead we decided to pick the parameters that performed best (5 best models) on the validation sets and re-train XGBClassifiers with these parameters.

- As comparing rankings is quite tricky, method used to compare the charts should be fitted to the specific problem. For our models, the adequate method might be to take some top number of variables for each model and see what percent of variables repeats in feature importance of both models.
- The PDP charts might not be the best way to analyse Roshomon Sets on our dataset as they pretty much don't differ for different models.

Results

Model rank	Features with importance >1%
1.	f45, f46, f48, f47, f74, f94, f97
2.	f48, f47, f46, f45, f104, f74, f100, f94
3.	f45, f46, f47, f48, f74, f94
4.	f48, f47, f46, f45, f74, f100, f94, f91
5.	f48, f45, f46, f47, f74, f92, f104, f58

Mean test score depending on parameters values



Conclusion:

The mean AUCROC result on the test set is strongly dependent on the values of the model parameters. The differences in the spreads and interquartile ranges of the obtained results are particularly noticeable