

Wybieranie optymalnego modelu zanonimizowanemu zbiorowi danych

Wojciech Bogucki
Karol Pysiak

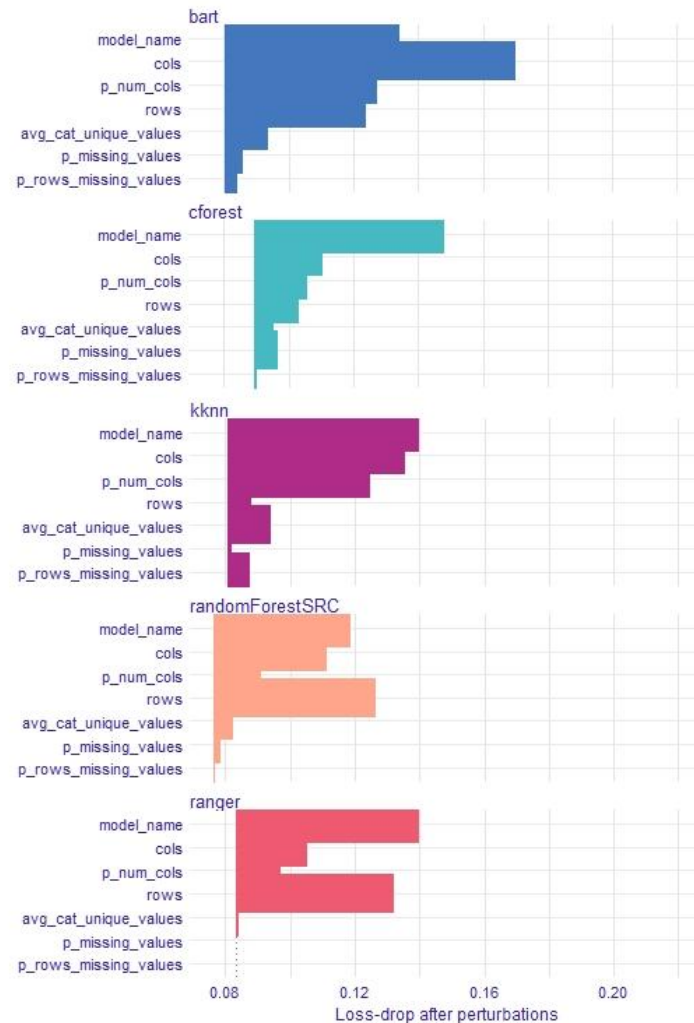
Podejście ogólne

Dowolny zbiór  Propozycja modelu

Scrapowanie informacji o modelach

	Id	rows	cols	p_missing_values	p_rows_missing_values	p_num_cols	avg_cat_unique_values	model_name	mlr	auc
1	kaggle_beauty	1260	10	0.00	0.000	0.600	2.000	classif.binomial	1	0.7409
2	kaggle_heart_disease	100	14	0.00	0.000	1.000	0.000	classif.randomForest	1	NA
3	kaggle_pokemon	800	12	0.04	0.482	0.583	12.667	classif.svm	1	0.9644
4	kaggle_pokemon	800	12	0.04	0.482	0.583	12.667	classif.gbm	1	0.9527
5	kaggle_pokemon	800	12	0.04	0.482	0.583	12.667	classif.rda	1	0.9538
6	kaggle_pokemon	800	12	0.04	0.482	0.583	12.667	classif.mda	1	0.9399

Predykcja AUC modelu dla danych wejściowych



Wybranie optymalnego modelu

	▲ models_names ▼	response ▼
1	classif.rotationForest	0.8828331
2	classif.cforest	0.8731673
3	classif.ranger	0.8724117
4	classif.bartMachine	0.8693313
5	classif.randomForestSRC	0.8686288
6	classif.ada	0.8684296

Wybór prawdopodobnie optymalnych parametrów

	parameters.K	parameters.L
1	2	10
2	2	13
3	3	7
4	3	13
5	5	18
6	5	10

Wybrany przez nas learner

`classif.rotationForest`

$K = 2$

$L = 7$

Dziękujemy za uwagę