Forester report

version 1.2.1

2023-05-16 12:55:55

This report contains details about the best trained model, table with metrics for every trained model, scatter plot for chosen metric and info about used data.

The best models

This is the **binary_clf** task.
The best model is: **ranger_model**.

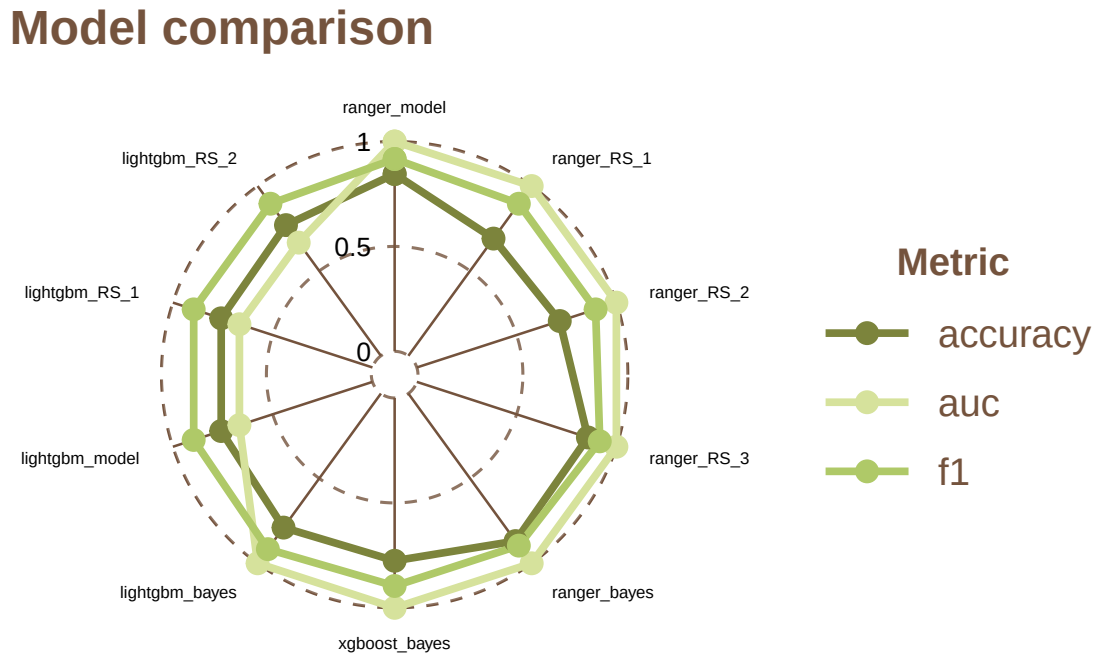
The names of the models were created by a pattern *Engine_TuningMethod_Id*, where:

- Engine describes the engine used for the training (random_forest, xgboost, decision_tree, lightgbm, catboost),
- TuningMethod describes how the model was tuned (basic for basic parameters, RS for random search, bayes for Bayesian optimization),
- Id for separating the random search parameters sets.

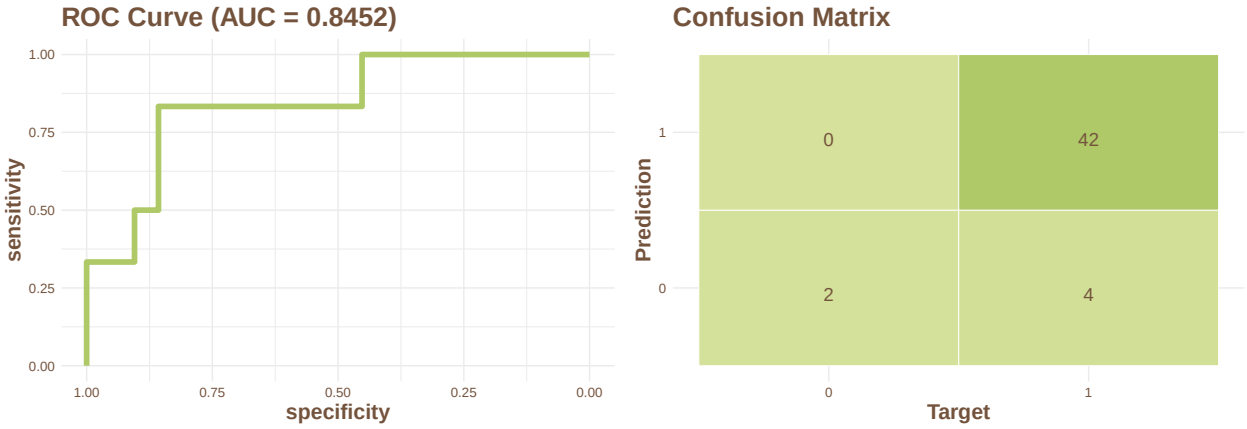
More details about the best model are present at the end of the report.

no.	name	accuracy	auc	f1
1	ranger_model	1.0000	0.9167	0.8452
5	ranger_RS_1	1.0000	0.8958	0.6905
6	ranger_RS_2	1.0000	0.8958	0.7163
7	ranger_RS_3	1.0000	0.9167	0.8571
17	ranger_bayes	1.0000	0.8958	0.8690
18	xgboost_bayes	1.0000	0.8958	0.7758
20	lightgbm_bayes	1.0000	0.9167	0.7897
4	lightgbm_model	0.6667	0.8958	0.7579
14	lightgbm_RS_1	0.6667	0.8958	0.7579
15	lightgbm_RS_2	0.6667	0.8958	0.7698
16	lightgbm_RS_3	0.6667	0.8958	0.7778
8	xgboost_RS_1	0.5000	0.8750	0.6389
9	xgboost_RS_2	0.5000	0.8750	0.6468
10	xgboost_RS_3	0.5000	0.8750	0.6468
2	xgboost_model	0.0000	0.8542	0.7976
3	decision_tree_model	NaN	0.8750	0.5000
11	decision_tree_RS_1	NaN	0.8750	0.5000
12	decision_tree_RS_2	NaN	0.8750	0.5000
13	decision_tree_RS_3	NaN	0.8750	0.5000
19	decision_tree_bayes	NaN	0.8750	0.5000

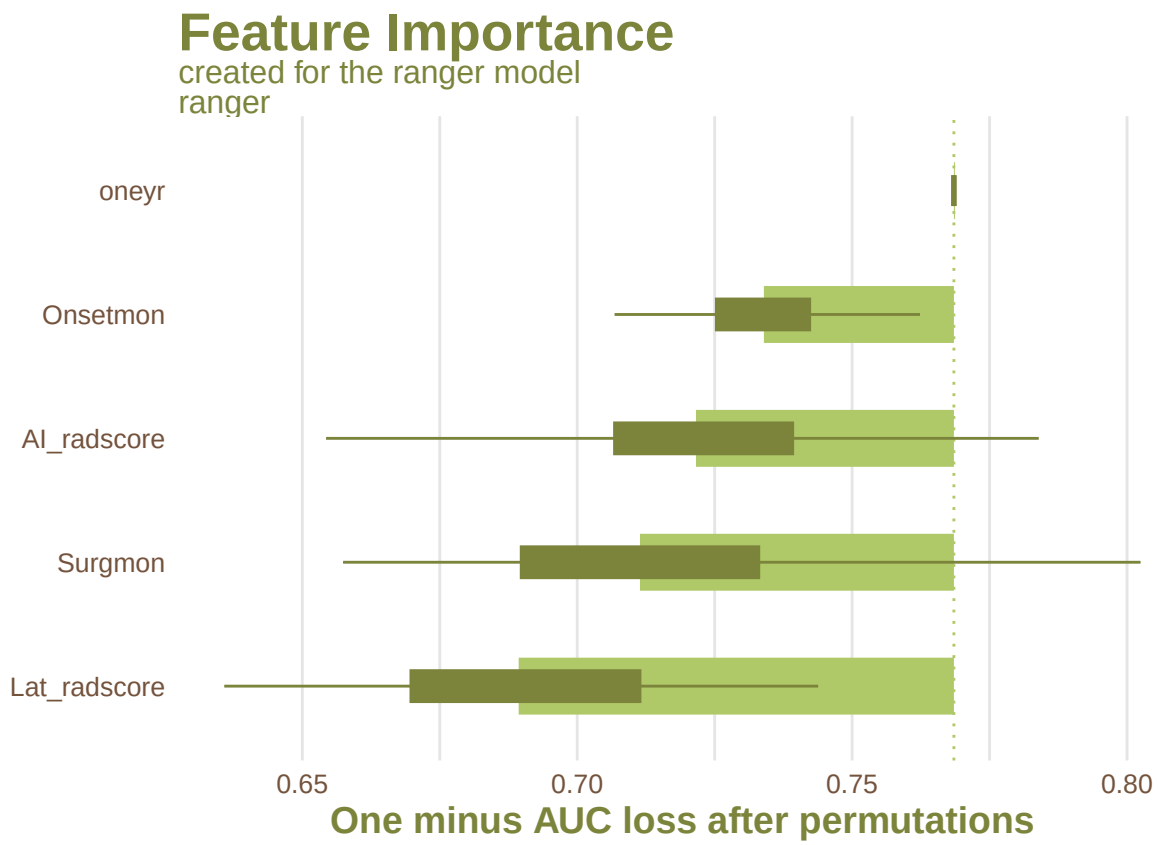
Plots for all models



Plots for the best model - ranger_model



Feature Importance for the best model - ranger_model



Details about data

CHECK DATA REPORT

The dataset has 234 observations and 18 columns which names are:

oneyr; AI_radscore; Lat_radscore; side; Sex; Surgmon; Onsetmon; Durmon; Freq; SE; SGS; early_brain_injury; familial_epilepsy; brain_hypoxia; Central_Nervous_System_Infections; traumatic_brain_injury; history_of_previous_surgery; MRI;

With the target value described by a column: oneyr.

No static columns.

No duplicate columns.

No target values are missing.

No predictor values are missing.

No issues with dimensionality.

No strongly correlated, by Spearman rank, pairs of numerical values.

There are more than 50 possible outliers in the data set, so we are not printing them. They are returned in the output as a vector.

Dataset is unbalanced with: 7.357143 proportion with 0 being a dominating class.

Columns names suggest that none of them are IDs.

Columns data suggest that none of them are IDs.

CHECK DATA REPORT END

The best model details

```
------------ Ranger model ------------

Parameters
  num.trees: 500
  ntry: 2
  num.samples: 138
  min.node.size: 10
```