# Forester report

version 1.0.0

2022-10-11 10:35:20

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.

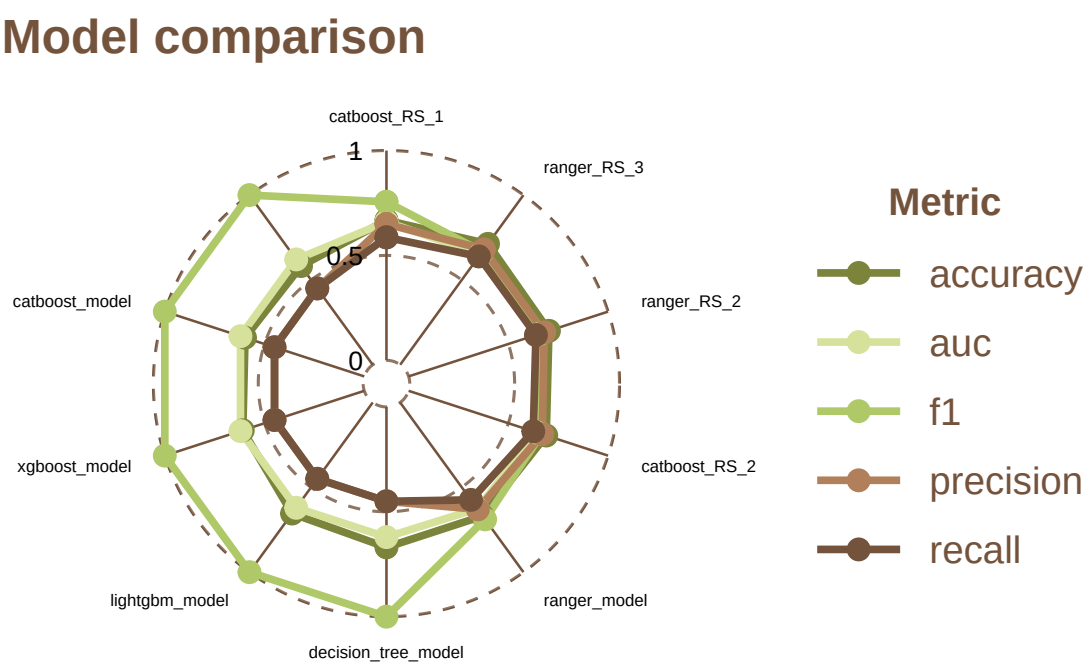
## The best models

This is the **binary_clf** task.
The best model is: **catboost_RS_3**.

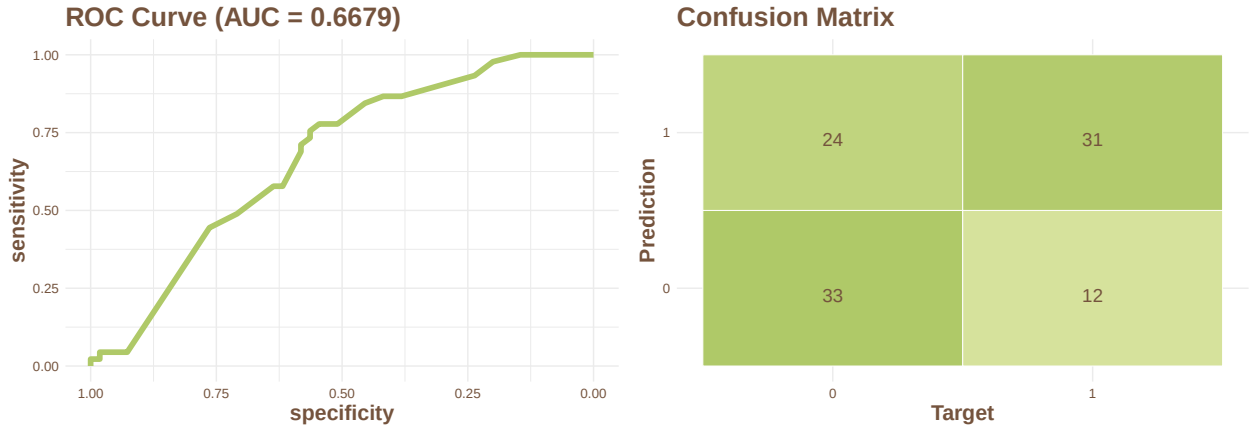
*More details about bests models are present at the end of the report.*

| engine | auc | f1 | recall | precision | accuracy |
|---|---|---|---|---|---|
| catboost_RS_3 | 0.6675 | 0.6602 | 0.7556 | 0.5862 | 0.65 |
| catboost_RS_1 | 0.7125 | 0.6522 | 0.6667 | 0.6383 | 0.68 |
| ranger_RS_3 | 0.7034 | 0.6522 | 0.6667 | 0.6383 | 0.68 |
| ranger_RS_2 | 0.6899 | 0.6452 | 0.6667 | 0.6250 | 0.67 |
| catboost_RS_2 | 0.6691 | 0.6263 | 0.6889 | 0.5741 | 0.63 |
| ranger_model | 0.6697 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| decision_tree_model | 0.6545 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| lightgbm_model | 0.6091 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| xgboost_model | 0.5980 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| catboost_model | 0.5818 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| ranger_bayes | 0.6697 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| decision_tree_bayes | 0.6545 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| lightgbm_bayes | 0.6091 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| xgboost_bayes | 0.5980 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| catboost_bayes | 0.5818 | 0.6207 | 1.0000 | 0.4500 | 0.45 |
| decision_tree_RS_1 | 0.6861 | 0.6136 | 0.6000 | 0.6279 | 0.66 |
| decision_tree_RS_2 | 0.6861 | 0.6136 | 0.6000 | 0.6279 | 0.66 |
| decision_tree_RS_3 | 0.6861 | 0.6136 | 0.6000 | 0.6279 | 0.66 |
| xgboost_RS_3 | 0.6436 | 0.6067 | 0.6000 | 0.6136 | 0.65 |
| lightgbm_RS_3 | 0.6905 | 0.5977 | 0.5778 | 0.6190 | 0.65 |
| ranger_RS_1 | 0.7145 | 0.5909 | 0.5778 | 0.6047 | 0.64 |
| lightgbm_RS_1 | 0.6768 | 0.5806 | 0.6000 | 0.5625 | 0.61 |
| lightgbm_RS_2 | 0.6729 | 0.5806 | 0.6000 | 0.5625 | 0.61 |
| xgboost_RS_1 | 0.6513 | 0.5682 | 0.5556 | 0.5814 | 0.62 |
| xgboost_RS_2 | 0.6493 | 0.5682 | 0.5556 | 0.5814 | 0.62 |

## Plots for all models




## Plots for the best model - catboost_RS_3




## Feature Importance for the best model - catboost_RS_3

```
[1] "Feature importance unavailable for catboost model."
```

## Details about data

CHECK DATA REPORT

**The dataset has 500 observations and 7 columns which names are:**

Two_yr_Recidivism; Number_of_Priors; Age_Above_FourtyFive; Age_Below_TwentyFive; Misdemeanor; Ethnicity; Sex;

**With the target value described by a column:** Two_yr_Recidivism.

**No static columns.**

**No duplicate columns.**

**No target values are missing.**

**No predictor values are missing.**

**No issues with dimensionality.**

**No strongly correlated pairs of numerical values.**

**No strongly correlated pairs of categorical values.**

**These obserwation migth be outliers due to their numerical columns values:**

102 108 157 173 210 273 322 326 34 356 393 424 425 455 ;

**Dataset is balanced.**

**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