

Forester report

version 1.0.0

2022-10-29 20:12:34

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 **regression** task.

The best model is: **catboost_RS_6**.

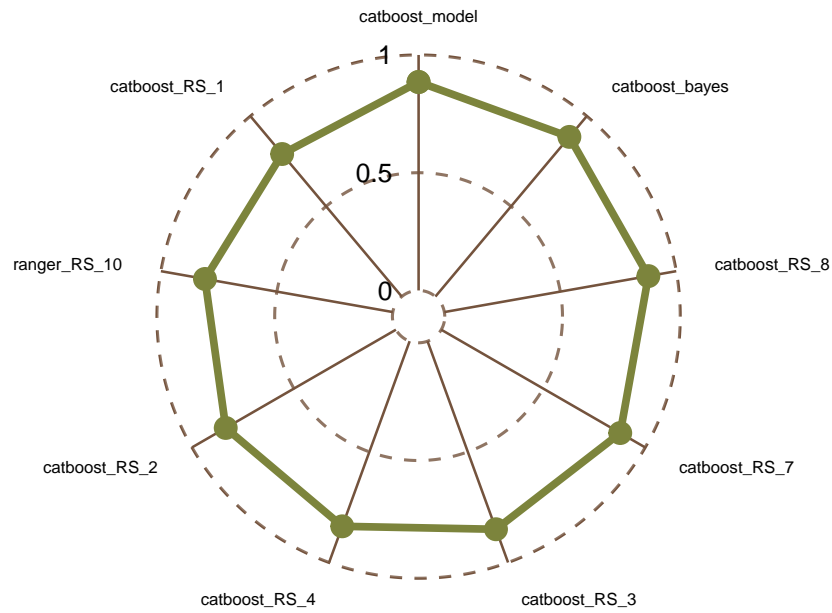
More details about bests models are present at the end of the report.

engine	mse	rmse	r2	mad	mae
catboost_RS_6	13353746751	115558.4	0.8899	49370.13	80144.66
catboost_model	13919850626	117982.4	0.8852	57738.73	82555.68
catboost_bayes	13919850626	117982.4	0.8852	57738.73	82555.68
catboost_RS_8	14579647766	120746.2	0.8798	48674.38	85500.61
catboost_RS_7	14826170636	121762.8	0.8778	48446.54	87269.56
catboost_RS_3	18207571960	134935.4	0.8499	63306.55	94855.34
catboost_RS_4	19821906760	140790.3	0.8366	49112.15	90034.72
catboost_RS_2	20062956150	141643.8	0.8346	48155.55	88362.99
ranger_RS_10	23043645112	151801.3	0.8100	90605.45	111867.15
catboost_RS_1	25445768730	159517.3	0.7902	47093.75	99729.52
ranger_RS_2	29744035701	172464.6	0.7548	78049.08	115261.25
ranger_RS_4	32076673840	179099.6	0.7355	73426.54	116174.61
lightgbm_RS_3	32885542450	181343.7	0.7288	93431.67	132653.46
lightgbm_RS_4	32885542450	181343.7	0.7288	93431.67	132653.46
lightgbm_RS_7	32991540296	181635.7	0.7280	91107.68	131976.35
lightgbm_RS_9	32991540296	181635.7	0.7280	91107.68	131976.35
ranger_RS_6	33177234154	182146.2	0.7264	140032.15	148109.41
ranger_RS_1	33183864286	182164.4	0.7264	86461.39	122950.60
lightgbm_model	33307251362	182502.7	0.7254	98513.85	132299.45
lightgbm_RS_8	33307251362	182502.7	0.7254	98513.85	132299.45
lightgbm_bayes	33307251362	182502.7	0.7254	98513.85	132299.45
ranger_model	34333011362	185291.7	0.7169	62768.43	110160.76
ranger_bayes	34333011362	185291.7	0.7169	62768.43	110160.76
decision_tree_model	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_1	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_2	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_3	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_4	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_5	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_6	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_7	34948901243	186946.3	0.7118	88899.58	126518.67

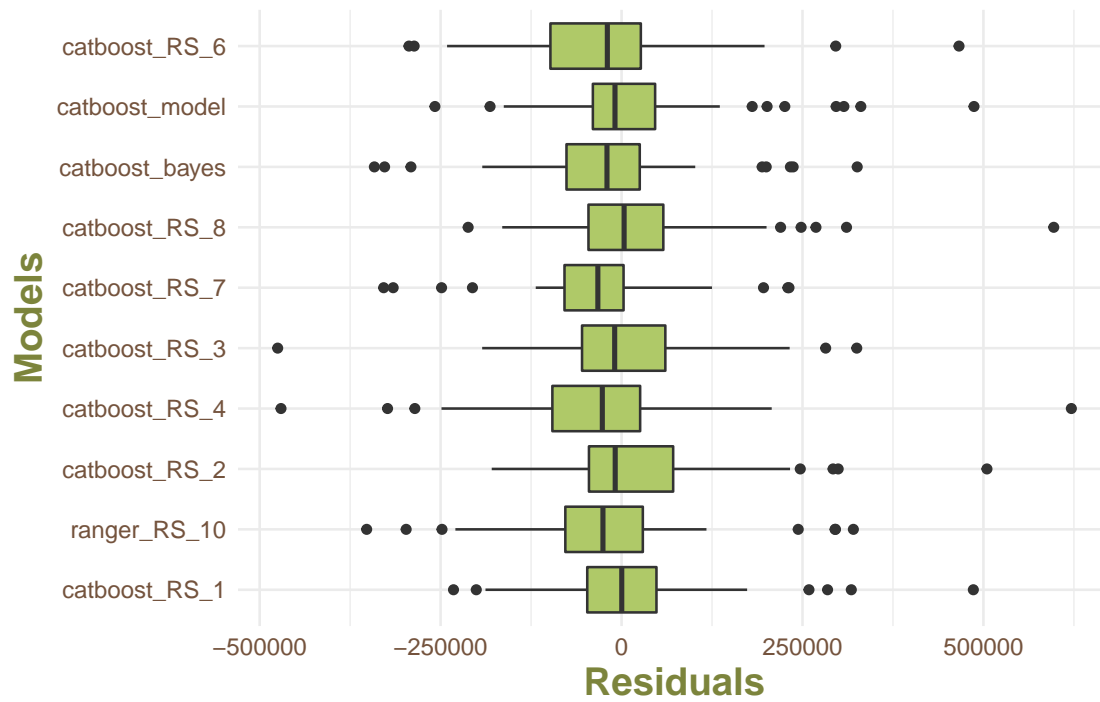
engine	mse	rmse	r2	mad	mae
decision_tree_RS_8	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_9	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_RS_10	34948901243	186946.3	0.7118	88899.58	126518.67
decision_tree_bayes	34948901243	186946.3	0.7118	88899.58	126518.67
ranger_RS_7	35436494675	188245.8	0.7078	56691.75	112922.10
lightgbm_RS_1	35488228313	188383.2	0.7074	86257.99	133377.22
lightgbm_RS_10	36012695793	189770.1	0.7031	119023.09	140972.00
xgboost_RS_4	36843465002	191946.5	0.6962	86301.80	135179.87
xgboost_RS_1	39061180732	197639.0	0.6779	83441.67	140866.94
ranger_RS_5	40238940110	200596.5	0.6682	49075.31	112558.56
ranger_RS_3	40389152160	200970.5	0.6670	137410.81	158465.51
lightgbm_RS_2	41196618898	202969.5	0.6603	87601.78	138300.96
ranger_RS_9	42745646808	206750.2	0.6475	78090.82	126968.65
lightgbm_RS_5	47098267837	217021.4	0.6117	127655.71	162800.76
lightgbm_RS_6	47098267837	217021.4	0.6117	127655.71	162800.76
ranger_RS_8	51767657511	227525.1	0.5732	148442.35	168236.29
catboost_RS_5	64115920475	253211.2	0.4713	100026.78	155761.77
catboost_RS_10	67097523815	259031.9	0.4468	104211.18	163720.22
catboost_RS_9	76416298395	276435.0	0.3699	107169.05	168477.69
xgboost_RS_2	119126284483	345146.8	0.0178	179840.80	264539.20
xgboost_RS_5	119126284483	345146.8	0.0178	179840.80	264539.20
xgboost_RS_8	119126284483	345146.8	0.0178	179840.80	264539.20
xgboost_RS_10	119126284483	345146.8	0.0178	179840.80	264539.20
xgboost_model	173815084745	416911.4	-0.4332	83791.05	160454.41
xgboost_bayes	173815084745	416911.4	-0.4332	83791.05	160454.41
xgboost_RS_3	269033584689	518684.5	-1.2183	353415.49	428241.60
xgboost_RS_6	269033584689	518684.5	-1.2183	353415.49	428241.60
xgboost_RS_7	270719709137	520307.3	-1.2322	356445.80	426719.88
xgboost_RS_9	270719709137	520307.3	-1.2322	356445.80	426719.88

Plots for all models

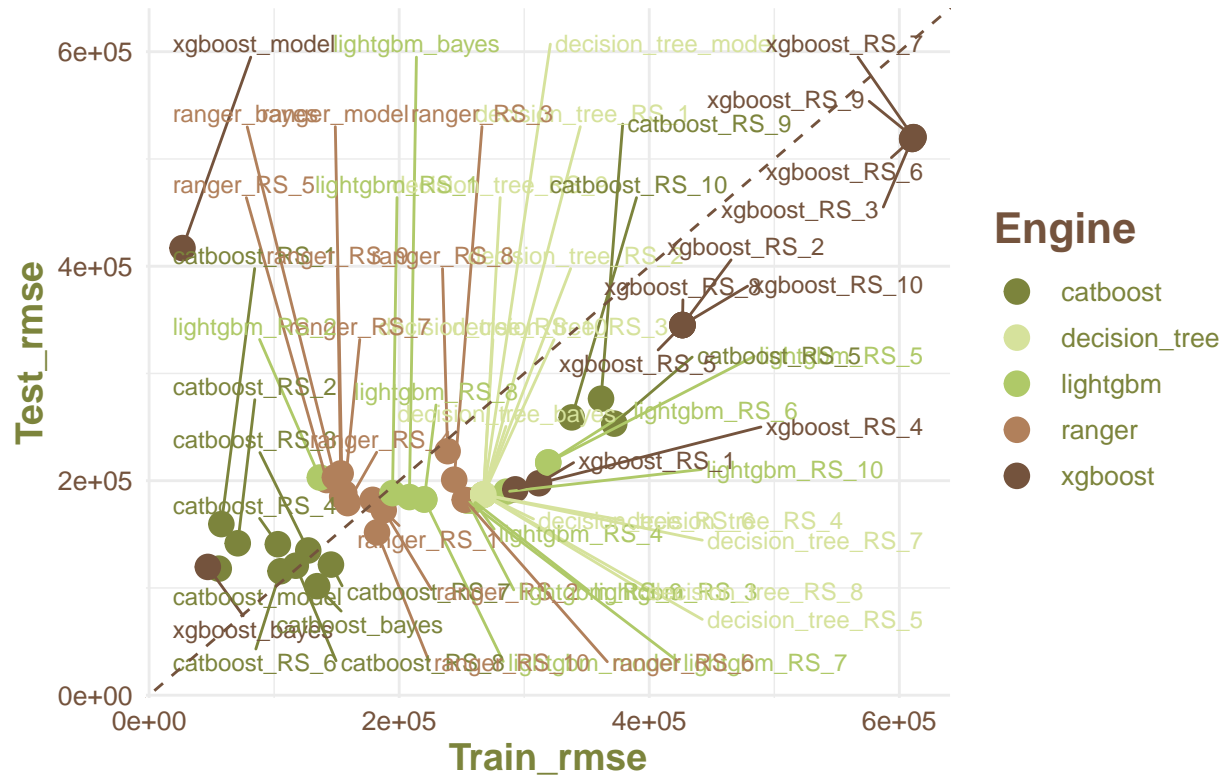
R2 comparison



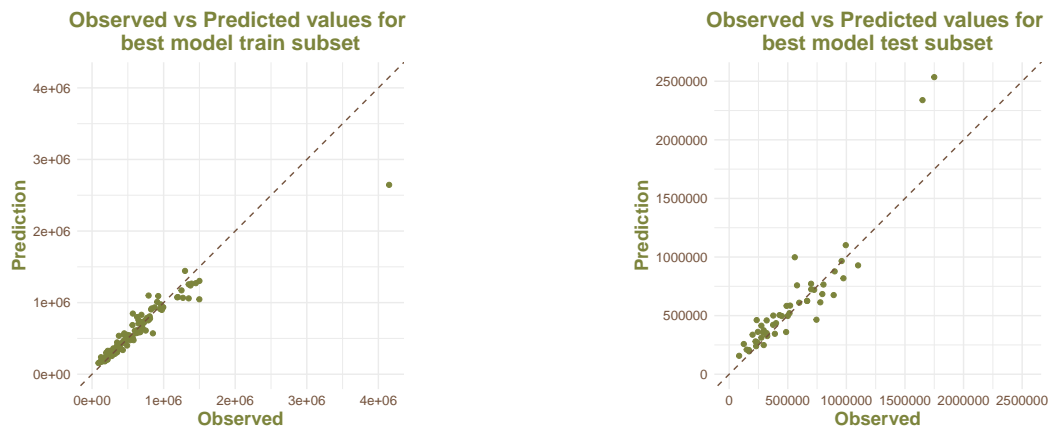
Combined Models Residuals Plot



RMSE Train vs Test plot



Plots for the best model - catboost_RS_6



Feature Importance for the best model - catboost_RS_6

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

Details about data

[CHECK DATA REPORT](#)

The dataset has 246 observations and 17 columns which names are:

Id; Condition; PropertyType; PropertySubType; Bedrooms; Bathrooms; AreaNet; AreaGross; Parking; Latitude; Longitude; Country; District; Municipality; Parish; Price.M2; Price;

With the target value described by a column: Price.

Static columns are: Country; District; Municipality;

With dominating values: Portugal; Lisboa; Lisboa;

These column pairs are duplicate: District - Municipality;

No target values are missing.

No predictor values are missing.

No issues with dimensionality.

Strongly correlated pairs of numerical values are:

Bedrooms - AreaNet: 0.77; Bedrooms - AreaGross: 0.77; Bathrooms - AreaNet: 0.78; Bathrooms - AreaGross: 0.78; AreaNet - AreaGross: 1;

Strongly correlated pairs of categorical values are:

PropertyType - PropertySubType: 1;

These observation might be outliers due to their numerical columns values:

145 146 196 44 5 51 57 58 59 60 61 62 63 64 69 75 76 77 78 ;

Target data is not evenly distributed with quantile bins: 0.25 0.35 0.14 0.26

Columns names suggest that some of them are IDs, removing them can improve the model.
Suspicious columns are:

Id

Columns data suggest that some of them are IDs, removing them can improve the model.
Suspicious columns are:

Id

————— **CHECK DATA REPORT END** —————

The best model details