Forester report

version 1.0.0

2022-11-29 00:17:11

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

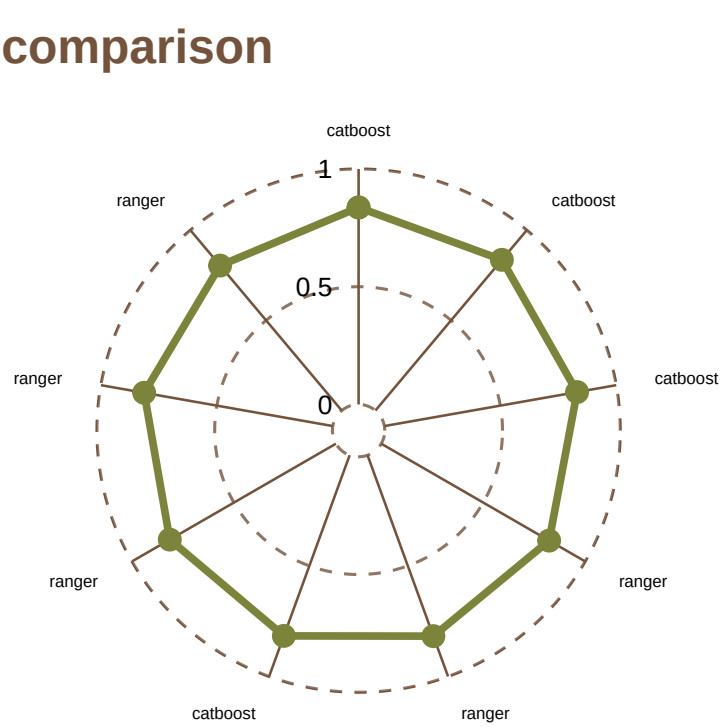
This is the **regression** task.
The best model is: **ranger_bayes**.

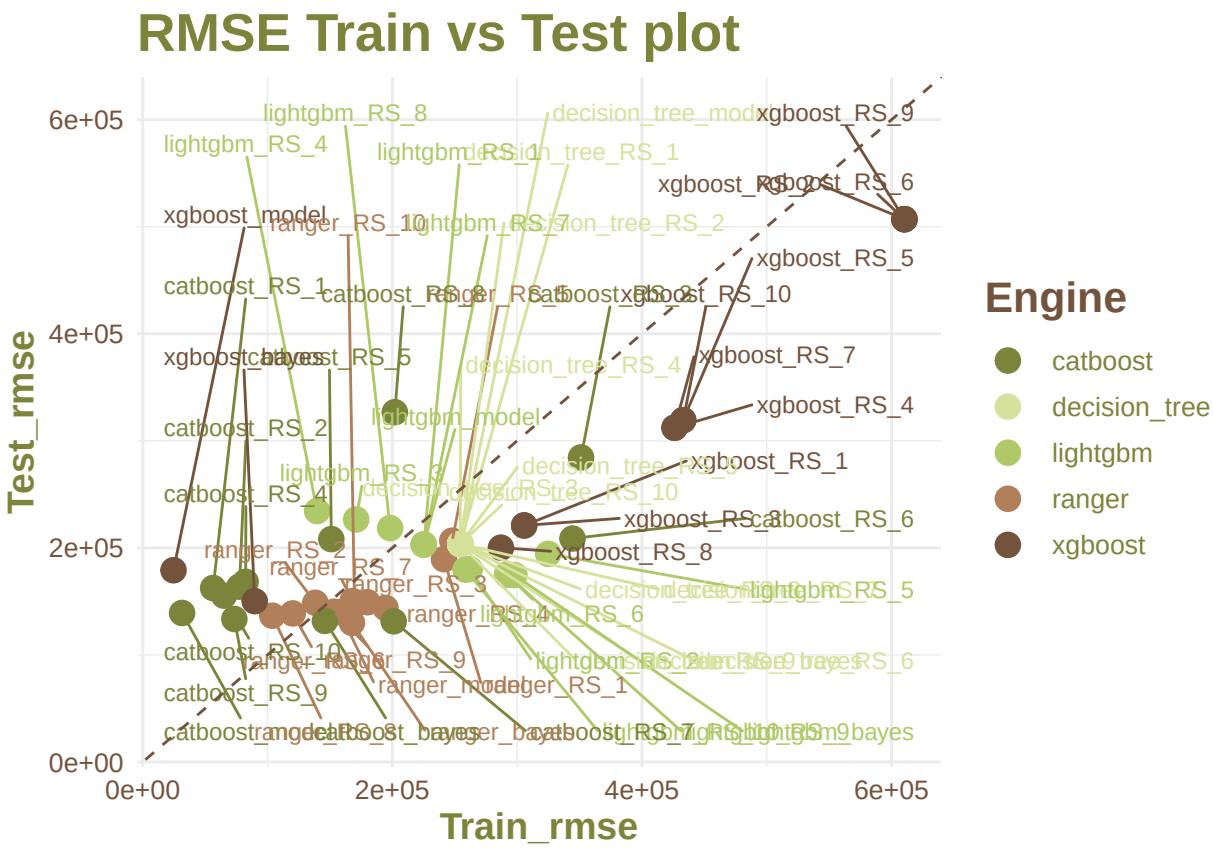
More details about bests models are present at the end of the report.

no.	name	engine	tuning	mse	r2	mae
56	ranger_bayes	ranger	bayes_opt	16883884274	0.8401	98562.63
52	catboost_RS_7	catboost	reandom_search	17260072600	0.8366	101693.45
60	catboost_bayes	catboost	bayes_opt	17353571236	0.8357	91591.98
54	catboost_RS_9	catboost	reandom_search	17833209665	0.8311	98599.62
13	ranger_RS_8	ranger	reandom_search	18664013766	0.8233	104687.41
11	ranger_RS_6	ranger	reandom_search	19247620835	0.8177	105450.93
5	catboost_model	catboost	basic	19360719763	0.8167	106146.88
1	ranger_model	ranger	basic	19576847580	0.8146	107434.42
14	ranger_RS_9	ranger	reandom_search	19697622426	0.8135	107765.48
9	ranger_RS_4	ranger	reandom_search	20715464044	0.8038	118773.93
15	ranger_RS_10	ranger	reandom_search	21544995521	0.7960	118006.63
7	ranger_RS_2	ranger	reandom_search	22108929810	0.7906	120877.50
8	ranger_RS_3	ranger	reandom_search	22256116902	0.7892	119869.43
57	xgboost_bayes	xgboost	bayes_opt	22555626424	0.7864	102368.05
12	ranger_RS_7	ranger	reandom_search	22593918905	0.7860	124135.21
55	catboost_RS_10	catboost	reandom_search	24152398593	0.7713	117958.45
46	catboost_RS_1	catboost	reandom_search	26398915574	0.7500	117316.09
47	catboost_RS_2	catboost	reandom_search	26804384558	0.7462	118101.94
49	catboost_RS_4	catboost	reandom_search	28318444262	0.7318	118531.87
59	lightgbm_bayes	lightgbm	bayes_opt	30491942706	0.7113	144150.93
41	lightgbm_RS_6	lightgbm	reandom_search	30620875705	0.7100	141682.15
44	lightgbm_RS_9	lightgbm	reandom_search	30620875705	0.7100	141682.15
2	xgboost_model	xgboost	basic	32071601238	0.6963	105173.44
45	lightgbm_RS_10	lightgbm	reandom_search	32251131474	0.6946	123289.78
37	lightgbm_RS_2	lightgbm	reandom_search	32467502648	0.6925	125564.04
6	ranger_RS_1	ranger	reandom_search	35863980017	0.6604	163046.35
40	lightgbm_RS_5	lightgbm	reandom_search	37786957183	0.6422	165050.36
23	xgboost_RS_8	xgboost	reandom_search	40068645159	0.6206	152800.86
4	lightgbm_model	lightgbm	basic	41241249076	0.6095	134809.68
36	lightgbm_RS_1	lightgbm	reandom_search	41241249076	0.6095	134809.68

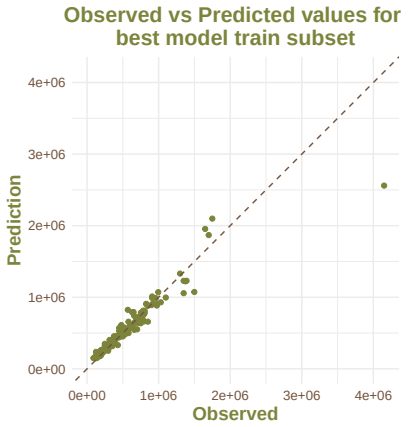
no.	name	engine	tuning	mse	r2	mae
42	lightgbm_RS_7	lightgbm	reandom_search	41241249076	0.6095	134809.68
3	decision_tree_model	decision_tree	basic	41440869163	0.6076	149750.18
26	decision_tree_RS_1	decision_tree	reandom_search	41440869163	0.6076	149750.18
27	decision_tree_RS_2	decision_tree	reandom_search	41440869163	0.6076	149750.18
28	decision_tree_RS_3	decision_tree	reandom_search	41440869163	0.6076	149750.18
29	decision_tree_RS_4	decision_tree	reandom_search	41440869163	0.6076	149750.18
30	decision_tree_RS_5	decision_tree	reandom_search	41440869163	0.6076	149750.18
31	decision_tree_RS_6	decision_tree	reandom_search	41440869163	0.6076	149750.18
32	decision_tree_RS_7	decision_tree	reandom_search	41440869163	0.6076	149750.18
33	decision_tree_RS_8	decision_tree	reandom_search	41440869163	0.6076	149750.18
34	decision_tree_RS_9	decision_tree	reandom_search	41440869163	0.6076	149750.18
35	decision_tree_RS_10	decision_tree	reandom_search	41440869163	0.6076	149750.18
58	decision_tree_bayes	decision_tree	bayes_opt	41440869163	0.6076	149750.18
10	ranger_RS_5	ranger	reandom_search	42728183325	0.5954	172026.77
50	catboost_RS_5	catboost	reandom_search	43330637883	0.5897	154339.48
51	catboost_RS_6	catboost	reandom_search	43550223044	0.5876	165031.32
43	lightgbm_RS_8	lightgbm	reandom_search	47730865411	0.5480	147206.19
16	xgboost_RS_1	xgboost	reandom_search	48801968891	0.5379	157457.28
18	xgboost_RS_3	xgboost	reandom_search	48801968891	0.5379	157457.28
38	lightgbm_RS_3	lightgbm	reandom_search	51296493750	0.5142	148548.51
39	lightgbm_RS_4	lightgbm	reandom_search	54933001033	0.4798	157421.98
48	catboost_RS_3	catboost	reandom_search	80927730921	0.2336	212851.62
19	xgboost_RS_4	xgboost	reandom_search	97454164693	0.0772	250462.49
22	xgboost_RS_7	xgboost	reandom_search	97509913344	0.0766	250716.62
20	xgboost_RS_5	xgboost	reandom_search	101940605622	0.0347	246008.40
25	xgboost_RS_10	xgboost	reandom_search	101940605622	0.0347	246008.40
53	catboost_RS_8	catboost	reandom_search	106744769176	-0.0108	234117.96
17	xgboost_RS_2	xgboost	reandom_search	257070718342	-1.4343	427263.76
21	xgboost_RS_6	xgboost	reandom_search	257070718342	-1.4343	427263.76
24	xgboost_RS_9	xgboost	reandom_search	257070718342	-1.4343	427263.76

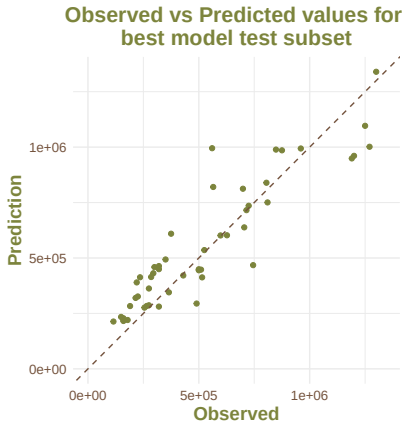
Plots for all models



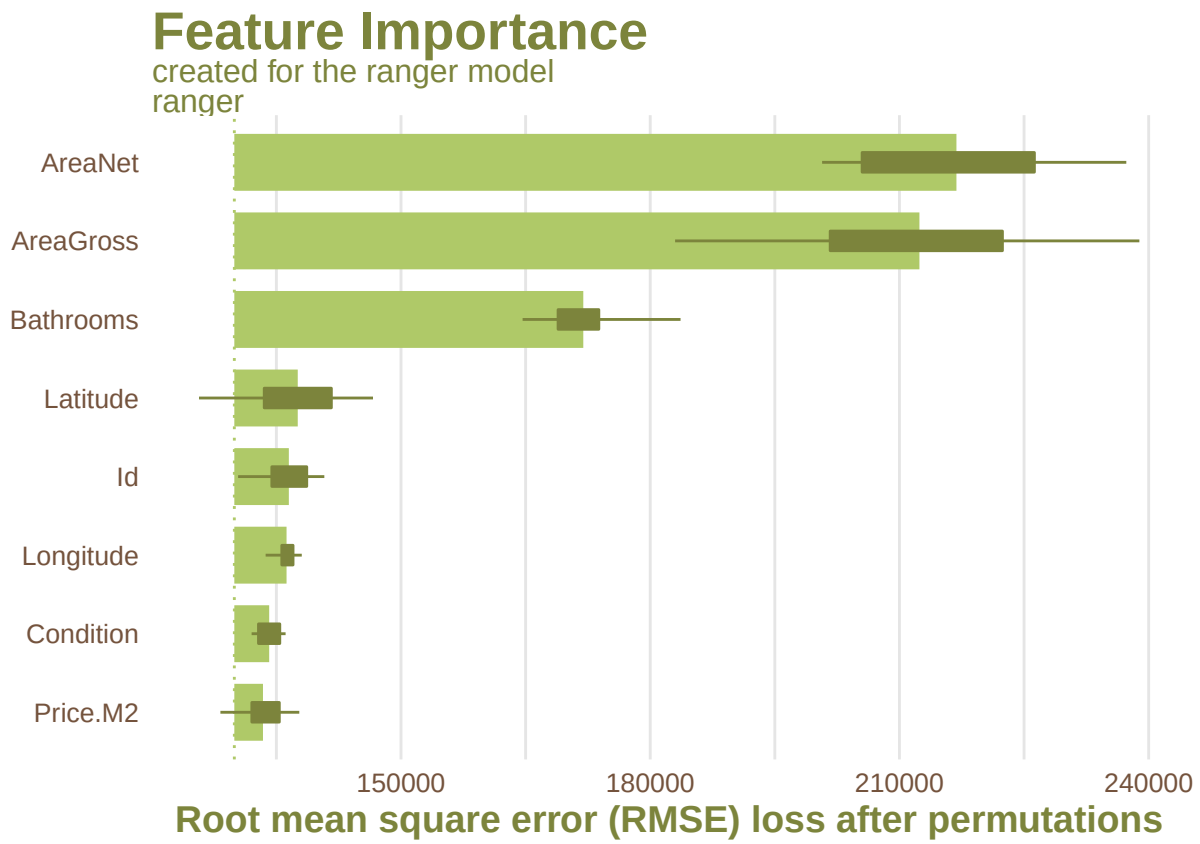


Plots for the best model - ranger_bayes





Feature Importance for the best model - ranger_bayes



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 obserwation migth 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