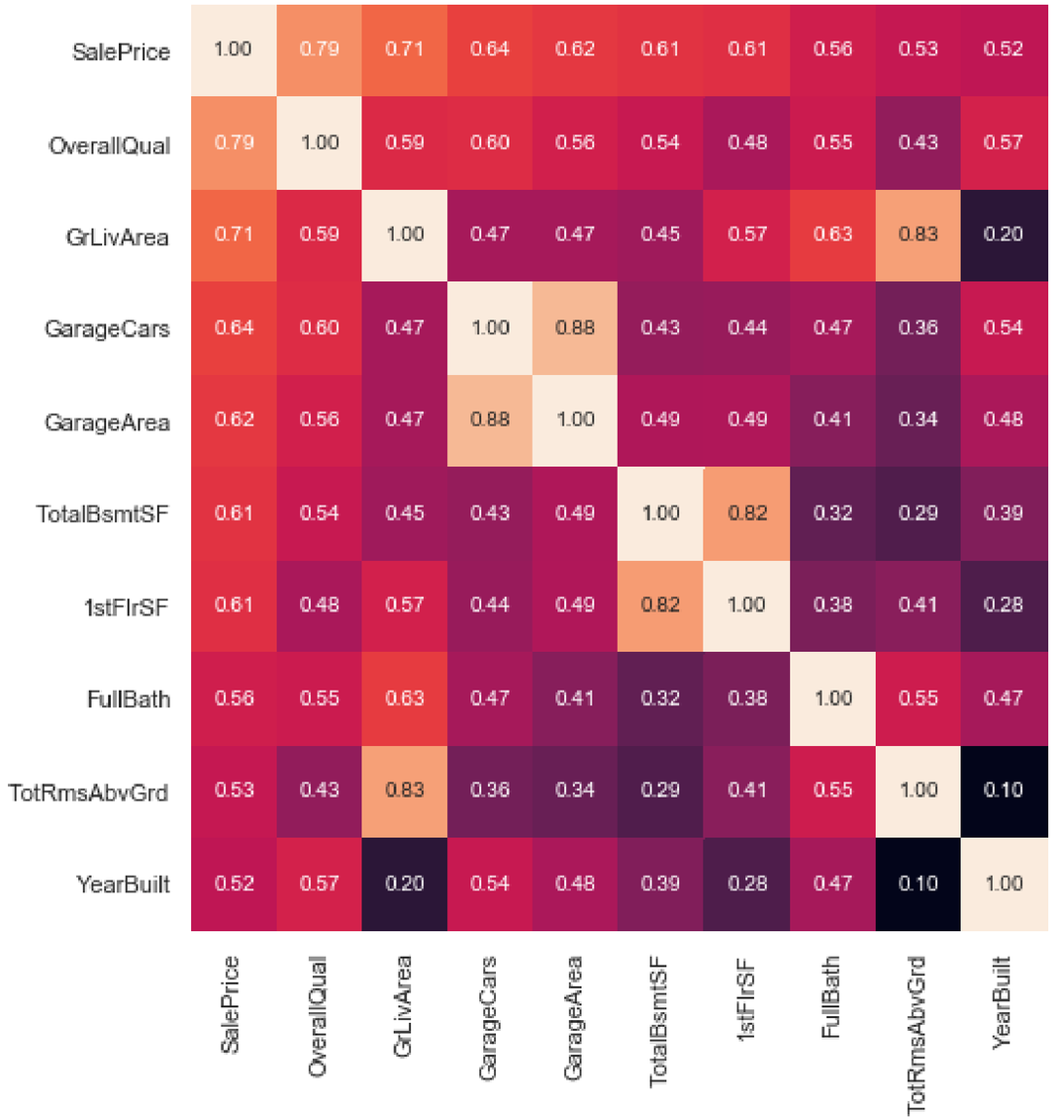
Задача Регрессии

RMSE: 0.1272

Место: 1532 из 5760

АЛЕКСЕЕВ АНДРЕЙ



HOUSE PRICES ADVANCED REGRESSION TECHNIQUES

ДАТАСЕТ

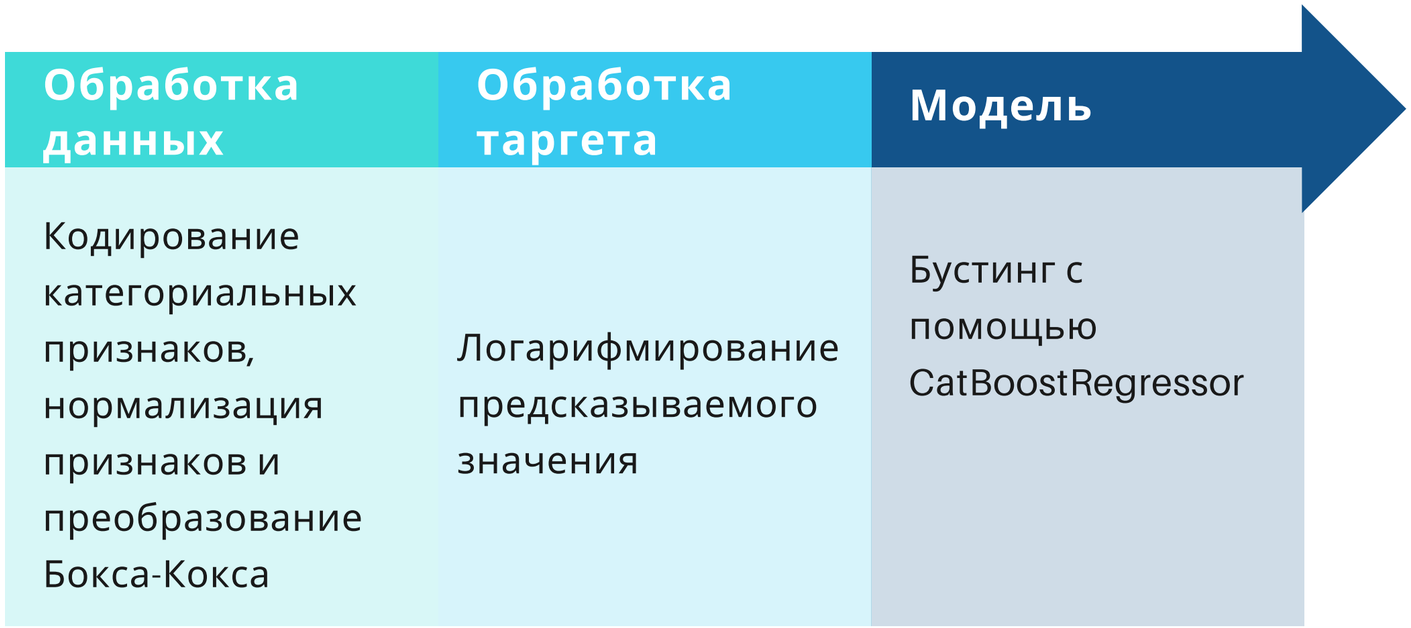
House Prices: Advanced Regression Techniques

Predict sales prices and practice feature engineering, RFs, and gradient boosting

k kaggle

ЦЕЛЬ: ПРЕДСКАЗАТЬ ЦЕНУ НА НЕДВИЖИМОСТЬ И ПОЛУЧИТЬ МИНИМАЛЬНЫЙ **RMSE**

ИТОГОВОЕ РЕШЕНИЕ



Работа с признаками

ЗАПОЛНЕНИЕ ПРОПУЩЕННЫХ ЗНАЧЕНИЙ	Многие фичи в датасете были пропущены. Вместо удаления данных столбцов было предложено заполнить их None или 0
БЫЛ ИСПОЛЬЗОВАН LABELENCODER	Данных подход показал лучшее качество, чем LabelBinarizer при обучении моделей (0.12 против 0.14)
БЫЛ ИСПОЛЬЗОВАН ROBUSTSCALER	Такая нормализация фичей позволяет привести датасет к нулевому среднему и единичной дисперсии без учета выбросов, тк для расчета статистик берутся данные между 1 и 3 квантилями
ИСПОЛЬЗОВАЛОСЬ ПРЕОБРАЗОВАНИЕ БОКСА-КОКСА	Использование данного преобразования помогло также улучшить результат регрессии
ДОБАВЛЕНА ФИЧА, КОТОРАЯ ВЫРАЖАЕТ ОБЩУЮ ПЛОЩАДЬ	Фича выражает собой общую площадь здания

Опробованные методы

Метод	**Lasso** *(**линейный**)*	**ElasticNet** *(**линейный**)*	**KNN** *(метрический)*	**CatBoost** *(**бустинг**)*	**RandomForest**
Настраиваемые параметры	-//-	-//-	-//-	-//-	-//-
Область настройки	-//-	-//-	-//-	-//-	-//-
Результат на кросс-валидации (выбран лучший)	**0.1350**	**0.1350**	**0.2063**	**0.1207**	**0.1456**
Результат на лидерборде (был отправлен только лучший алгоритм	-//-	-//-	-//-	**0.1272** ***(место 1532 из 5760)***	-//-

*Также был опробован подход с усреднением предсказаний моделей, но он не улучшил качество по сравнению с Catboost