A5
Kaggle: House Prices - Advanced Regression Techniques

Maj-Annika Tammisto

Introduction

Resources:

Training data with 1460 instances and test data with 1459 instances describing 79 aspects of residential homes in Ames, Iowa (USA);

Task:

To predict the final price of each home.

Goals

Goal 1:

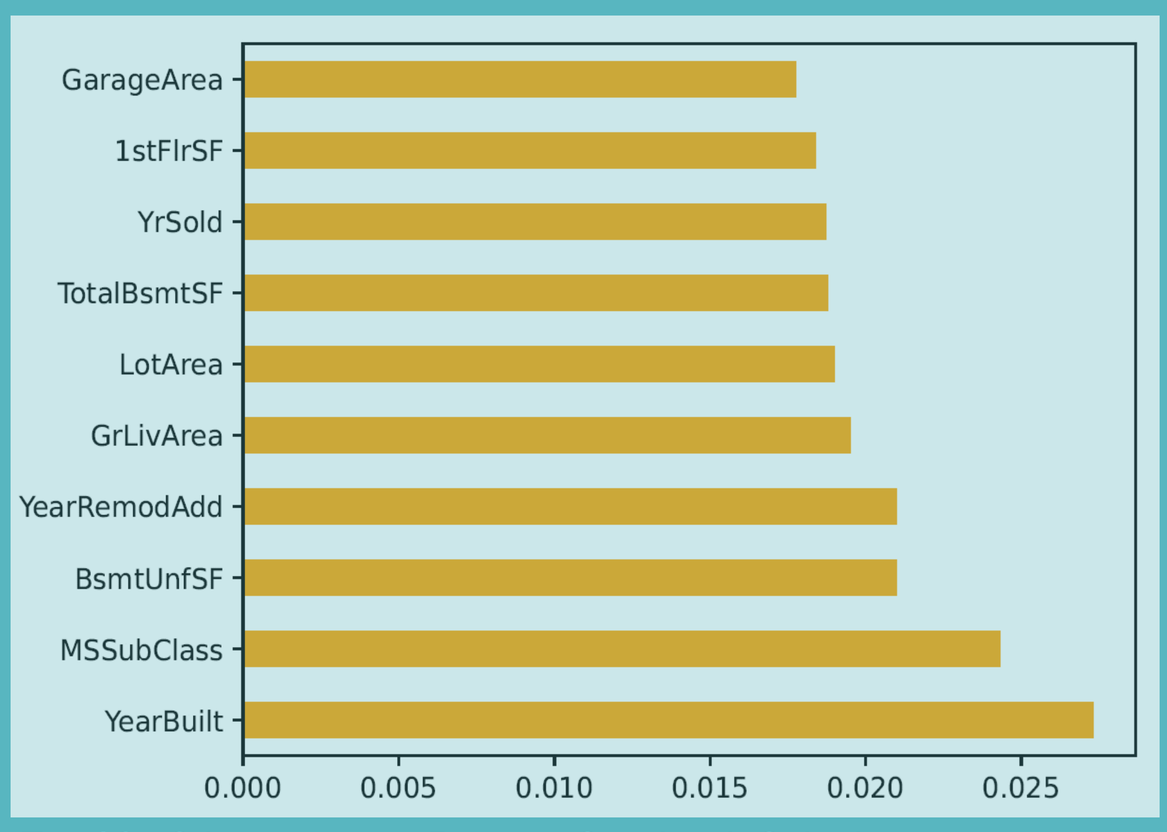
Applying (at least) three regression techniques and defining the one that achieves the best results on this specific data.

Goal 2:

Reaching a place in TOP 100 on the Leaderboard.

Data preparation

- Replacing NA values;
- One-hot encoding;
- Defining important Features;
- Deleting less important Features;
- Adjusting the number and format of Features in the training and test datasets so that they could be fitted.



Top 10 of the important Features found with Scikit-learn

Modelling

1. RandomForestRegressor
2. LinearRegression
3. XGBRegressor
4. GradientBoostingRegressor

Results

Goal 1 ✓

Goal 2 ✗ *The best RMSE value achieved with XGBRegressor: 0.12493, No 630 on the Leaderboard.*
RMSE value 0.11405 or lower required for reaching TOP 100 on the leaderboard.

Project Repository:
https://github.com/annimaj/KaggleHousePrices