1851049 Kato Kentaro
1873995 Ale Vikranth

Team Name: 1873995 & 1851049
Score: 0.11358

Language: Python

Data Tidying:

1. Removed outliers with data GrLivArea > 4000 & SalePrice < 300000
2. Dropped columns which have Multicollinearity > 0.8.
3. Normalize the Target feature i.e. Sale Price
4. Apply Log-Transformation if the Skewness > 0.6.
5. Make dummy features for categorical data.
6. Replacing missing numerical values with mean of the feature.

Feature Engineering:

1. Create TotalHousePorchSF which contains all features about porch.
2. Create TotalSF which is the total square feet of floors.
3. Created OverallQual_2 which is squared OverallQual because it is mostly correlated.

Modelization:

1. Linear Regression

Training:

We have Stacked the below models to improve our prediction on Sale Price.

1. Lasso
2. Elastic Net Regression (ENet)
3. Kernel Ridge Regression (KRR)
4. Support Vector Regression (SVM)
5. Gradient Boosting Regression (GBoost)
6. Light Gradient Boosting (LGB)
7. Extreme Gradient Boosting (XGB)

After we evaluate the weight for each model, we apply the model below.

$$ENet * 0.6 + GBoost * 0.3 + LGB * 0.1$$