Ti Zhang 1004424517

My price prediction process includes data cleaning, visualization, and modeling.

For data cleaning, I checked to see whether there are missing values in each feature given. And lucky enough, there isn’t any missing data. Next I identify numerical and categorical features in order to convert categorical features types into numerical classification for calculation purposes.

After getting a clean categorical feature set, I did a visualization by plotting each features’ distribution into histogram to check and make sure there aren’t any columns with low variances in order to make the model more accurate. Again all the columns are having reasonable variance and there is no need to make modification.

For the modeling part, I seperate the training dataset into 2 parts of test and training. I implemented two models: the first model is using XGB Regressor to implement a Gbtree. It’s an auto model of gradient descent of tree type with a penalty on complexity. The second model is using Light GBM Regression with L2 regression analysis. I fit the two training sets with the two models, and obtained two results with similar variance. Finally I fit the two models with all the date sets and used both of them to predict the test set. I combined the two sets of different results by taking their mean values, and imported the prediction value into the submission file.