Individual Project 1 Report

By Isaac Aguilar

What processing steps did you take?

* Create columns for the most important amenities.
* Convert amenities into a new column that has the number of amenities in the place.
* Create year columns from date related variables and remove old columns.
* Extract city and State and remove host\_location variable.
* Convert text percentages into decimals.
* Convert list of verifications to number of verifications.
* Extract bathroom and bathroom type from bathroom column.
* Imputing missing values with the mean for numerical and the mode for the categorical variables.
* Standardizing the scale for the numerical values
* Creating Dummy variables for the Categorical variables
* Doing feature selection with correlated feature variables and with the Select Percentile function.

How did you tune hyperparameters?

Using 4 to 5 fold cross validation for 4 different models. Each one using their proper Hyperparameters:

* KNN: Tunning number of neighbors
* Linear Regression (No penalty): No tunning parameters
* Ridge Regression: Tunning alpha
* Lasso Regression: Tunning alpha

How did you select your final model and what was it?

I compared the test MSEs from each of my models with their best parameters, and selected the model with the smallest one, which was the KNN model with ‘5’ nearest neighbors and ‘distance’ method for the weights parameter.

What is your out-of-sample mean absolute error?

KNN Test MSE: 1317058.186022189

What features seem to be the most important in the prediction?

* City
* Property Type
* Neighborhood cleansed
* Room Type