Model Metrics File

Hypertuning and parameters

The parameters chosen for the SVM model were C, kernel and max_iter. The hyper tuning model used was RandomizedSearchCV. There were no parameters for the Gaussian Naive Bayes model, rather the two most strongly correlated variables were selected from a correlation matrix and the model run considering only those two variables, TotalSF and GrLivArea. The RandomForestClassifier model took max_depth and n_estimators parameters. RandomizedSearchCV was the selected hyper-parameter tuning model.

Metrics

The R-squared metric was the metric chosen to score models, as the data was scaled, and the preferred model would predict values that are close to the actual values as possible, minimizing predictions that were large outliers. A metric using mean absolute error would have been preferred if we were simply preferring the model that would yield the best result on average, however, the purpose of the model is to predict a fair sale price for a home for both buyer and seller. For this use case, a large outlier would be more detrimental to the models use than models that produced predictions that were on average more a few hundred dollars more than another prediction.

Model

The chosen model was the RandomForestClassifier. This model produced a surprisingly strong r2 score of .717. The predictions were compared with actual values in the test set by finding the residuals. To visualize those residuals, I plotted the predictions over the actual values. The charts are strikingly visually similar. Lastly, the confidence interval of the residuals was examined.