Reddic Housing LLC

House Price Estimate Report

## presented by

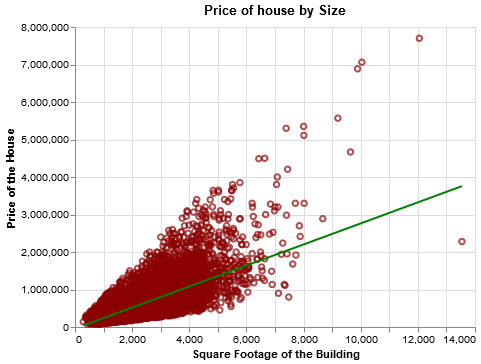
Cody Overholt

Kyle Mueller

Jordan Carlson

1. Accuracy of the Model

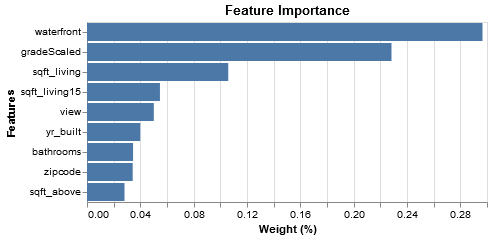
We used a gradient boosted tree model to predict the price of the houses based on the data provided. The metrics used to determine how accurate the model is are the Root Mean Squared Error and the R^2 value. R^2 is measured on a scale between 0 and 1 with scores closer to 1 meaning higher accuracy. The Root Mean Squared Error is the average difference between the actual values and the predicted values of the regression model. The R^2 value was 0.83339 and the RMSE was around $130 thousand. That may seem like a lot but the properties that were more expensive deviated from the prediction line by a much greater margin than those that were below as shown in this graph.

1. Grouping of data for increased precision

Grade comparison:

|  |  |  |
| --- | --- | --- |
| Low (4-6) | Average (7) | High (8-10) |
| RMSE: $82,206  : 0.62 | RMSE: $75,336  : 0.77 | RMSE: $121,964  : 0.87 |

III. Feature Importance

The Following Graph shows how each feature weighed into the overall model.

The biggest factor in determining the price of a property was whether or not it was a waterfront property. The second most important was the grade of the property followed by the size of the house.

1. Conclusion

Ensemble learning and XGBoosted trees enabled the production of a very accurate model. Data dropped from the learning algorithm included: price', 'date', 'yr\_renovated', 'id', 'lat', 'long', 'grade', 'condition'.

Adjusting parameters of the model increased its accuracy and prevented overfitting from occurring. By grouping the data by housing type or by quality grade, the model was able to be even more accurate by reducing RMSE and increasing R2.

This ensemble model will accurately predict the pricing of houses desired and enable an efficient regression system for your business. Thank you.

1. Python Notebooks

Below are Github Gist links to the notebooks we used during this case study:

<https://gist.github.com/JordanCarlson7/264deef0993ce63f5f5df3231882200d>

<https://gist.github.com/codeholt/58cd947ee37450ab731f346712bd4470>