

Capstone Two

House Prices - Advanced Regression Techniques

Goal

The goal of this project is to build regression models to leverage attributes of properties to predict sales prices and practice feature engineering, RFs, and gradient boosting as well as to provide useful house price prediction for setting data-driven budgets for buyers and insightful tools for setting more reasonable prices when selling properties.

Problem Identification

How much are the residential homes in Ames, Iowa given all aspects/attributes of the properties?

Approach

I will follow the steps as in the first guided capstone as the following:

1. Problem Identification: this proposal
2. Data Wrangling: data collection, data organization, data definition, data cleaning
3. Exploratory data analysis: explore the data relationships of all the features and understand how the features compare to the response variable
4. Pre-processing: create a cleaned development dataset to use to complete the modeling step
5. Modeling: build two to three different models and identify the best one, and review model outcomes
6. Documentation: finalize code, review the results, finalize documentation
7. Presentation - a slide deck with 10-20 slides

Data sources

The data is available from [kaggle](#).

Ames House Dataset with 79 attributes and sales prices of 1,460 properties. Features examples:

- Exterior features: Exterior material quality, type of foundation, masonry veneer type
- Interior features: heat condition, central air conditioning, kitchen quality
- Location feature: Zoning classification, slope of property, physical location within city limits

Deliverable

I will create a github repo containing the work I complete for each step of the project including Jupyter notebooks, a final report, and a slide deck.