

## House Sales Price Prediction

### Project Description

The objective of the project is to foresee sales price of house in Iowa, Ames. People move from one place to another considering factors like the education of their children, career driven plans and other various reasons. In such cases, Data science can lend a helping hand for foresee the price of the house they look for. So divining problem can be formulated as collecting important attribute which would make impact on sales price.

This is imputed into regression algorithms. Regression is the process of predicting value from the given metrics. This algorithm is used to fruitfully foretell the price of house.

### Dataset Description

The Ames Housing Price data set [1] recently released on Kaggle [2] is “a modernized and expanded version of the often-cited Boston Housing dataset”. It covers all the recorded house sale price in Ames, IA from January 2006 to July 2010. With 79 explanatory variables describing almost every feature of residential homes, we aimed to apply data imputation, feature engineering and machine learning modeling to achieve a better predictive accuracy on the housing price.

Each house description consist of such attributes as house area (called ‘GrLivArea’), garage capacity (‘GarageCars’), overall quality estimation of house and kitchen(‘OverallQual’, ‘KitchenQual’), distinct of the city (‘MS\_Zoning’), data about neighborhood, type of sale (‘SaleType’), year of building(‘YearBuilt’), and another similar attributes.

The dataset contains 1460 observations in the training set and 1459 observations in the test set. There are 46 categorical variables including 23 nominal and 23 ordinal ones, and 33 numeric variables in the dataset. The training set also has the sale price as response while the test set does not. Solutions are evaluated on Root-Mean-Squared-Error (RMSE) between the logarithm of the predicted value and the logarithm of the observed sales price. (Taking logs means that errors in predicting expensive houses and cheap houses will affect the result equally).

### Tools and Technology used

- a. Python.
- b. Jupyter Notebook.

### Algorithms Used: -

1. Multi-linear regression
2. Artificial Neural Network (ANN)
  - Multi-Layer Perceptron (MLP)
3. Random forest Model

### Results achieved: -

The model's accuracy in predicting house price was measured by a number of criteria. The value of score is compared to select preferred model. By using random forest, the score value was increase about 11% higher than Multi-Layer Perceptron. It can be concluded Random forest model is preferred to predict house price compared to Multi-Layer Perceptron model can be used as an alternative way to estimate house price in future.