

## Final Project

### Using Azure Machine Learning to predict house prices

#### Problem Statement:

In 2017, around 5.57 million existing homes were sold in the US (Statista, U.S. existing home sales 2005-2018). Determining the optimal sell price is a critical decision of the seller. If priced too low, the seller leaves money on the table. If priced too high, the house will sit on the market unsold. A negative perception can result when a house is on the market for a considerable amount of time, or when the price is reduced often. Using existing sales, how accurately can we predict the selling price of a home before it is put onto the market? We will attempt to build such a system to predict housing prices using Azure Machine Learning.

#### Overview of the Technology:

To solve the problem, I created an experiment in Azure Machine Learning studio using data I copied from Kaggle. I inspected the data, looking for any missing values. The data includes 81 features, including sale price. Not every feature contributes to the accuracy of the predictive model. Some features are redundant, unnecessary and extraneous. In fact, some features can reduce the effectiveness of the model. I employed feature selection to select the most relevant features that will optimize the effectiveness of the model. Once the features were selected, I split the data into training and validation. Both data sets were used to train and evaluate multiple regression algorithms. I selected the model with Coefficient of Determination, from Evaluate Model module, closest to 1.0.

#### High Level Steps:

- 1) Download data
- 2) Clean data
- 3) Transfer data to Azure Machine Learning Studio
- 4) Build experiment in Azure Machine Learning Studio
- 5) Identify features to be used to predict prices
- 6) Evaluate various techniques to discover optimal algorithm to predict prices

#### Data Source:

Kaggle House Prices: Advanced Regression Techniques

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

#### Hardware Used:

Windows 10 64 bit processor desktop

#### Software Used:

Microsoft Azure Machine Learning Studio (<https://studio.azureml.net/>)

Anaconda 5.0 distribution of Python 64 bit (<https://www.anaconda.com/>), which includes:

- Python 2.7.14
- Jupyter (visualization tool)

Deep Azure@McKesson

Steve Klonicki

**YouTube Links:**

2 Min: <https://youtu.be/kb44rs7QztQ>

15 Min: <https://youtu.be/GQZ3NAmyNxA>

**GitHub:**

<https://github.com/we814/AMLDeepDive>