**STAT 517 -LITERATURE REVIEW**

**Housing price prediction by machine learning algorithms**

**Guang Qiu 10/03/2018**

**Introduction:**

Housing price valuation is one of most important trading decisions. In addiction to the housing price indices from Standard & Poor’s Case-Shiller home price indices and the housing price index of the Office of Federal Housing Enterprise Oversight (OFHEO), machine learning algorithm could be another way to predict the price of housing, which used to assist in prediction by building models of housing price.

**Goals:**

In order to develop a prediction model for housing prices, several machine learning algorithms were used. And people can achieve the goals as follows.

Which prediction model is better compare to other prediction models with smaller error rate?

Is there any advance compare to conventional ways with machine learning approaches such as neural network and SVM?

What can it used to do in the future after building housing price model?

**Descriptions:**

When people tried to develop housing price, the hedonic-based regression approach was used, but the models focused on the relationship between house price and housing characteristics and it had potential limitations relating to fundamental model assumptions and estimation. As the developing of artificial intelligence techniques, it could be a more popular way to approach the housing price prediction.

Housing prices are an important reflection of the economy, and housing price ranges are of great interest for both buyers and sellers. In the articles, four algorithms (C4.5, RIPPER, Naïve Bayesian, and AdaBoost) were used to compare the performance of machine learning algorithms. Despite there are four algorithms, more models are needed to predict the housing price. Housing prices could be predicted given explanatory variables that cover many aspects of residential houses. As continuous house prices, they could be predicted with various regression techniques including Lasso, Ridge, SVM regression, and Random Forest regression; as individual price ranges, they could be predicted with classification methods including Naive Bayes, logistic regression, SVM classification, and Random Forest classification. And we could also perform PCA to improve the prediction accuracy.

The goal is to create a regression model and a classification model that are able to accurately estimate the price of the house given the features. We would perform two types of supervised learning algorithms: classification and regression. While it seems more reasonable to perform regression since house prices are continuous, classifying house prices into individual ranges of prices would also provide helpful insight for the users; also, this helps us explore different techniques which might be regression- or classification-specific.

Finally, I would continue reading related articles about housing price prediction and compare them to find better model with smaller error.