**Literature Review**

**Nathan Rollins**

**Mapping of Real Estate Prices Using Data Mining Techniques**

This is a recent report from the Czech Republic. Their dataset is impressive, collecting a large number of samples every month since 2007. From this data, they made an interesting distinction between the listing price (the price a home is listed for) and final sale price. They showed that the two were not to be guaranteed to be the same. In fact, they differed by 10-15% on average. The scope, however, of this project is extremely large. It was focused on all of the Czech Republic. It focuses on a more macro level prediction and statistical presentation, which means it is an invaluable study to learn the landscape of the entire country, but not as fitted to the particular cities themselves.

**Data Mining and Analysis and Predictions of Real Estate Prices**

Using a dataset derived from 1000 local, randomly chosen homes, this team created two models—one utilizing decision trees and the other neural networks. This data was taken from sales in the 2012 to 2013 period. For better or worse, the dataset only included properties of each property and not any kind of environmental influences, such as inflation or interest rates. They found that their neural network system found better results than their decision trees system, however, they were not able to guaranteed identify the cause of this. This was a troubling case, as they mentioned two of their sources had found the opposite to be true without a clear cause.

**Using Data Mining and text Mining techniques in Predicting the Price of Real Estate Properties in Dubai**

An interesting case where the team demonstrated a regression-based approach to creating their model. It should be noted that their prediction percentage was improved by utilizing text mining. Overall, a worthy example of how to perform linear regression in creating a predictive model.

**A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study**

A study conducted on a small dataset, (roughly five hundred entries), consisting of Boston suburban homes. The team would take in certain variables for each house, store and treat it, and then feed them into their two prediction algorithms: the Cascade Forward Back Propagation system and the Feed Forward Back Propagation system, which are both differently structured neural networks. These systems were tested to predict, given the correct values, the price of existing homes, and did so with a 96% accuracy rate. It is an interesting study with a clear explanation of their two systems, as well as the regression and statistical analysis explained. The only problem is how small their sample size was, and the potential problems that come with that.

**Works Cited and Reviewed:**

Agarwal,Vaishai, Gan, Victor, and Kim, Ben. “*Data Mining Analysis and Predictions of Real Estate Prices*.” Issues in Information Systems, vol. 16, issue IV. Pp. 30-36, 2015.

Bahia, Itedal Sabri Hashim. “*A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study*.” International Journal of Intelligence Science, 2013.

Hromada, Eduard. “*Mapping of real estate prices using data mining techniques*.” Elsevier Ltd, 2015.

Khashan, Deena Younis Abo. “*Using Data Mining and Text Mining Techniques in Predicting the Price of Real Estate Properties in Dubai*.” The British University in Dubai, 2014.