**Statement of Work - SOW**

**House Sales Price Prediction**

**Prepared By:**

**Muhammad Imran**

**Final Project: AI Algorithms 1002**

**Fall 2020**

**Executive Summary**

Artificial intelligence has become the breakout technology, utilizing huge amounts of computing power to learn and identify patterns in data.  Real estate industry has been driven by personal preferences and human interactions, primarily among buyers, sellers, and real estate agents. But now, changes to the buyer-agent-seller paradigm are happening. Real estate companies adopting Artificial Intelligence will be critical to sustaining and enhancing their competitive advantage to grow in this rapidly evolving industry.

**Rationale Statement**

Property price influences many different transactions such as sales and loan. Traditionally, property estimate is determined by professional. The risk is that human tend to be biased due to interest from lender, buyer, or seller. Therefore, a predicted system based on Artificial intelligence can serve as an independent and less biased system. For first time or less experienced buyers of real estate properties, an automated price prediction system can be useful to suggest underpriced or overpriced properties in the market.

**Data Set:**

In this project, Kaggle.com data from “House Prices: Advanced Regression Techniques” competition will be used for model building and evaluation. The train dataset consists of character and integer variables and **SalePrice** is a target feature.

Following data files are available,

* train.csv - the training set and has 163 columns.
* test.csv - the test set

**Model/Architecture Approach**

**Supervised Regression models:**

House Price Prediction is a classic regression problem to predict sales price. Following models will be used and compared for best accuracy.

1. Regression model
2. XGBoost , Light GBM
3. Support Vector Machine

**Programming language and Tools** : Python notebook, Colab and Anaconda – Jupyter notebook.

**Model Evaluation matrix**

Model performance matrix will be based on:

* **R-squared**
* **Mean Square Error (MSE):**

.