

*Machine Learning Bootcamp - 2020*

**Team No. 1**

**Project 4**

**September 9th 2020**

**Mentor**: Pushpesh Sharma

**Team Members**:

Jose Alvarado - Business Analyst

Jiajun Chang - Developer

Kalu Divine – Data Engineer

Nhung Nguyen – Data Scientist

**Abstract Submission**

The housing market is one of the most important markets in the US. Because of this, predicting the price of house is an important tool. With the knowledge of how a house’s price will change, investors can buy a house with the best rate at a low risk. The purpose of this model will be to use a data set to predict the housing price of several homes in a neighborhood.

The neighborhood used for the data set will be in Boston. The data set will have 13 parameters: crime rate, proportion of residential land, proportion of nonretail business, a dummy variable, nitric oxide concentration, average rooms per home, proportion of owner-occupied homes, distance from employment centers, accessibility of highways, property tax, pupil-teacher ratio, proportion of African Americans by town, lower status population and median value of owner-occupied homes. The output will be the housing price in thousands of dollars.

The proposed technique that we will use is K Means. This will cluster the data set into k different clusters. The data set will be unsupervised, and a sample will be assigned to the cluster with the nearest mean. First, we will perform iterative optimization to the clusters. We begin by initializing the cluster centers randomly and repeat until the clusters converge. As we repeat, we update the cluster labels by assigning points to the nearest cluster center and set the center to the mean of the cluster. Once converged, we then take the distance between the clusters.