

DATS 6202  
Machine Learning I  
Dr. Amir Jafari

*Group 10 Proposal on*  
*Maryland Property Data*  
by  
Tyler Wallett

**What problem did you select, and why did you select it?** I selected the topic of approximating the consideration of residential properties in Montgomery County, Maryland. The reason behind choosing this topic, aside from the fact that my background is in finance and it could potentially convert into a capstone project, was to challenge the conception behind the article *Getting Ahead of the Market: How Big Data Is Transforming Real Estate* by McKinsey & Company, where non-traditional features were shown to explain roughly 60% of the variation in the predictive power (McKinsey). Therefore, my goal is to attempt to answer this question: Can we successfully approximate consideration prices of residential properties from Montgomery County, Maryland only using traditional features?

**What database will you use?** The public dataset used can be found accessing this link: <https://hub.arcgis.com/datasets/maryland::maryland-property-data-parcel-points/about>. It contains roughly 2.4 million properties, with 134 features describing each, in the state of Maryland and was last updated on January 7, 2023.

**Is it large enough to train a machine learning or different algorithms?** The subset of the dataset we are interested in, 'Land usage' residential and 'County' Montgomery County, contains roughly 145,000 residential properties. So, yes it is large enough to train machine learning algorithms.

**What type of neural network will you use?** A multilayer perceptron regression model.

**Will it be a standard form of the network, or will you have to customize it?** The structure of the model will be R-S1-1, where R are the amount of inputs to be processed and S1 are the amount of neurons in the hidden layer. The transfer functions used will be *logsig()* and *purelin()*, respectively.

**What other machine algorithms will you use?** Linear regression and Random Forest regressor are other machine learning algorithms that will be used.

**What software will you use to implement the neural network or different algorithms?**  
**Why?** Python because it is the programming language that I am most familiar with, and also contains machine learning modules such as Scikit-Learn and Statsmodels that will help my analysis.

**What reference materials will you use to obtain sufficient background on applying the chosen algorithm to the specific problem that you selected?** Scikit Learn's documentation, books and Neural Network Design by Martin Hagan.

**How will you judge the performance of the network? What metrics will you use?** I will use  $R^2$  and the mean squared error with a 95% confidence interval.

**Provide a rough schedule for completing the project.** I will start working approximately three weeks prior to the deadline in order to successfully complete all of the tasks.