Nathan Vanos and Scott Kopczynski
CPSC 310, Professor Sprint, Final Project Proposal
4/8/2019

The Power of Scikit SVM Versus the Power of our Classifiers

For our project we will predict classifications for the same dataset using both our classifiers and SKLearn support vector machine. We will be using the outcome to determine the answer to the age old question: which of the classifiers is superior? The dataset we'll be using is a table of thousands of houses with their associated pricing and upgrades. This dataset can be located at

https://www.kaggle.com/greenwing1985/housepricing/version/1#. It is formatted as a CSV file and contains a total of sixteen attributes, some of which are number of garages, number of fireplaces, bathrooms, whether or not it has solar panels, a swimming pool or a garden, etc. From this dataset we will try to predict the price of a home.

There are a few challenges that this project will entail. First off, the number of attributes that the dataset contains is greater than the number of attributes in any other dataset we've used so far. Though this by itself shouldn't be much of an issue, we will also have to discretize the prices, and we'll have to learn how to implement an SVM using Scikit learn.

The results of this project could be useful to a wide range of people, specifically homeowners, because they could use the program to predict the value of their own home. There are two main stakeholders who would benefit from this program: real estate agents and homeowners.