**Capstone Project #1 – Proposal**

**The Problem**

The project will attempt to build a model that can accurately predict the relevance of search results for an online retailer in order to create an improved customer shopping experience.

**The Client**

The client in this case is Home Depot, although the methods used should be applicable to other online retailers. Home Depot currently uses human raters to evaluate changes to their search algorithms, which is a slow process that can have bias. It is hoped that the results of this project can minimize or remove the amount of human input in the process and increase the number of iterations that Home Depot can perform on their search algorithms leading to a faster and more accurate customer experience.

**The Data**

The data set contains a number of products and real customer search terms from Home Depot's website. The challenge is to predict a relevance score for the provided combinations of search terms and products. To create the ground truth labels, Home Depot have crowdsourced the search/product pairs to multiple human raters.

The relevance is a number between 1 (not relevant) to 3 (highly relevant).

Each pair was evaluated by at least three human raters. The provided relevance scores are the average value of the ratings.

**The Approach**

I will begin by performing exploratory analysis of the data to look for patterns. Then I will build predictive models and test and compare them for accuracy.

**The Deliverables**

A final report will be written to summarize the results including visualizations.