**Data Science Capstone Project Proposal—Amazon product Recommendation System**

Zi Yuan

June 16, 2016

1. **What is the problem you want to solve?**

Different from conventional physical delivery systems that are characterized by a scarcity of resources, on-line stores can make anything that exists available to the customer. For example, a physical bookstore may only have several thousand books on its shelves due to limited shelf space, but Amazon can offer millions of books to their customer. Due to overwhelming number of products, it is almost impossible for customers to go through all of them and find the products they are looking for. Recommendation in the online world is becoming more and more important. I will build an online retailer recommendation system, which offering customers suggestions about what they might like to buy. This system is mainly based on Amazon product reviews and metadata (e.g. price, related products, sales rank, and brand).

1. **Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?**

This recommendation system can help users discover items they might not have found by themselves.

1. **What data are you going to use for this? How will you acquire this data?**

I am going to use the Amazon product dataset, which contains product reviews and metadata from Amazon, spanning May 1996 - July 2014.

I will acquire this data from http://jmcauley.ucsd.edu/data/amazon/links.html

1. **In brief, outline your approach to solving this problem.**

I will use two different technologies to build the recommendation system:

* **Content-based systems:** examine properties of the products recommended.
* **Collaborative filtering systems**: recommend products based on similarity measures between customers and/or products.

1. **What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

Apart from actual code, I deliver a presentation along with graphs to make sure all data is presented in a coherent and easy-to-understand way.