**Capstone Project Proposal**

Gaurav Sureka

Foundations of Data Science

1. What is the problem you want to solve?

Recommender systems are backbones of some of the biggest companies like Amazon and Netflix . My aim is to compare different algorithms that existing recommender systems use and benchmark their in recommending the top products for a particular user based on their historical buying patterns. I will also try to combine both content-based and collaborative filtering methods into a hybrid recommemnder system for predictions on my dataset.

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?

The dataset that I have chosen is from Amazon and is specifically for the fine foods section. While Amazon already has an elaborate and closely guarded recommender system, I hope to come up with a recommender system which is better if not at par with what they already have. The performance of my recommender system cannot be judged in isolation, but I will benchmark it’s performance by running it on the test dataset that I will derive from the existing dataset.

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

I will be getting the data from Kaggle competitions website as a CSV file.

1. In brief, outline your approach to solving this problem (knowing that this might change later).

Firstly I will be going through the data and seeing what fields are important for my analysis and start working on them. As part of the EDA process, I will do some visualizations to see if there are any obvious patterns that stand out in the data. I will also try to do some text analytics based on the customer reviews to see what kind of sentiment is there among the consumers. For the recommender system, I intend to use arules, arulesViz and the recommenderLab packages in R.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck
2. Source code
3. Slides
4. Trying to come up with a short story based on it.