DATA643: Recommender System

Final Project Proposal

Harpreet Shoker, Rose Koh, Summer 2018

- **Overview**

As supply chain management and retail tech are growing at a faster pace, maintaining a log and predicting customer's next purchase is a highly valuable use case for a Data science student. Leveraging the skills and coding techniques we learnt in DATA643, our group wants to develop a recommender system that predicts the user's next purchase, using Instacart market basket data. This project is officially declared as open source and open challenge by Instacart.com.

- **Data**

The dataset for project is a relational set of files describing customers' orders over time. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.

Source: https://www.kaggle.com/c/instacart-market-basket-analysis/data

- **Deliverables**

A github repository with data, code, interpretation, visualization to understand model.

- **Workflow**

1) Data preprocessing
2) Matrix Factorization
3) Evaluation
4) Optimization
5) Finalization of the model

- **Resources**

Advanced mathematical techniques, Spark distributed computing.

- **References**

https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2