CSC 575 Project – Proposal
Name: Serena Yang, Zhong hua Xie
Topic: H&M Personalized Fashion Recommendations

Introduction:
In the era of big data, data analysis and machine learning can help customers make personalized recommendations. This function can save customers time and prevent users from missing out on the best option based on customers' past purchases.

H&M Group is a household name clothing brand, and this project will focus on its large amount of customer dataset. With H&M Group's enormous amounts of products, users might not quickly find what interests them or what they are looking for. In the end, they might end up not purchasing anything. Furthermore, based on the same interests that customers have bought and past orders from users, personalized recommendations can help customers positively impact sustainability, reduce returns, and minimize transportation emissions. Therefore, customized customer recommendations are essential for improving customers' shopping experience and company revenue.

Dataset:
This project aims to predict the clothes/article ID's each customer will purchase during the seven days immediately after the training data period.

There will be a folder of images corresponding to each article_id and four CSV files:

- articles.csv: details for each article_id/clothes available for purchase
  - contains 25 explanatory variables including: number, text, binary
  - 106k columns
- customer.csv: metadata for each customer_id in the dataset
  - 1.37m customers
  - 7 columns: text, binary, number
- Transactions_train.csv: include each customer's purchase on each date, as well as additional information. Duplicate rows correspond to multiple purchases of the same item.
  - 5 columns contain date, text, number
  - 1.36m customers
- sample_submission.csv: a file of lists that contains customer_id and their training prediction results, it is for the model accuracy
  - 2 columns include customers_id and list of prediction
  - 1.37m customers

The metadata ranges from simple data, such as clothing type and customer age, to text data in product descriptions and clothing images. The datasets can be found in Kaggle.

Challenge:
The project is to implement a personalized information filtering system for H&M customers. A system that can filter information and produce recommendations based on the user's transaction history and on the user's demographic.

Here is the list of functions the system is capable of by the end of the project.

- The system should be capable of handling millions of transaction data and produce relevant results based on the user's transaction history and other users with similar profiles.
    - Likeness by transactions history
    - Group by demographic
    - Collaborative filtering for recommendation (KNN)
- Ability to update user's profile using relevance feedback
    - System enables user's profile update and feedback
    - Update user's profile based on new transactions
- System index system
    - Capable of adding/removing items from the index
    - Update item's information

We will use Teams to communicate, track, and share all documents for this project. The plan is to have a meeting every week to discuss the current progress and the following tasks assigned to each person.