Title: Build a recommendation system to recommend products to customers based on the Amazon Reviews data (Electronics Dataset)

Team: Jingxuan Bao (CIS 5190), Yaoqi Deng (CIS 5190), Guanlin Huang (CIS 5190)

Kaggle dataset: Recommending electronics product purchases on Amazon
https://www.kaggle.com/code/pritech/product-recommendation-systems/notebook

Part 1: Implementation contributions:

Because the unique item number is much lower than the unique user account number. We plan to use the item-item collaborative filtering algorithm to generate recommendations for Amazon users. Specifically, we plan to consider using the KNN model, the SVD model, or even a deep learning model like the neural network. KNN and SVD models could be trained using the prepared user-item matrix. We plan to employ cross-validation and Grid Search to tune the hyperparameters pattern of the models. If the neural network will be used, technology like PyTorch could be applied.

Part 2: Evaluation contributions:

We will measure the accuracy of our approach using cross validation. We will be using precision, recall, and f1 score to evaluate. To analyze the algorithms, we plan to tune this following parameters: "k" in the KNN model, and the latent factors in SVD. We plan to use 5-fold cross-validation and Grid Search to tune these hyperparameters and select the best performing model. We also plan to evaluate the algorithm on different dataset shifts such as temporal shifts, splitting the data into train and test sets based on time, with the most recent data as the test set.

Prior work:

1. Recommending electronics product purchases on Amazon
   https://www.kaggle.com/code/pritech/product-recommendation-systems
2. A similar collaborative filtering recommendation system for Amazon purchases
   https://www.kaggle.com/code/saurav9786/recommender-system-using-amazon-reviews
3. A Yelp related recommendation system built by deep learning model
   https://www.kaggle.com/code/zolboo/recommender-systems-knn-svd-nn-keras

Which parts of the curriculum from this class do you expect to apply?:

Our tentative plan for the recommendation is to use SVD or KNN models which will be covered later in this lecture. Some of the other factors when developing the algorithm such as the parameter to look for will be covered in the Recommender Systems lecture.

Expected challenges and risk mitigation:

One challenge we will have to deal with is the handle the potential bias from the data. User ratings are based on multiple factors such as the item price and shipping speed. When evaluating the item to recommend for users, those factors should be considered. Another challenge would be fine-tuning the model. Tuning the model to find the most ideal parameter would require a large amount of testing, and there is no guarantee that we have tested out the best parameter combination when comparing different models. Hence, we will need to perform in-depth research on available training models before training the models.

Ethical considerations and broader social impact:

The recommendation algorithms utilize users' personal data, including user id, user purchase history, and product ratings. Analyzing this data without the explicit consent of the users might be a violation of their privacy.

A successful recommendation model can have a great impact on e-commerce business owners and users. It helps business owners to manage their product listing, and by successfully recommending their products to the customer who needs it, it could potentially increase sale and has a positive impact on the economy. In terms of users, a successful recommendation helps the user to purchase what they need, which improves the user experience and boosts the sales of the e-commerce website.