

DS 5720

Sovann Chang and Jiaying Liang

Project Proposal

Next Product Recommendation and Prediction

- Dataset
 - Amazon KDD Cup Multilingual Shopping Session Dataset
(<https://www.aicrowd.com/challenges/amazon-kdd-cup-23-multilingual-recommendation-challenge>)
 - Two separate .csv files for training. products_train contains a list of products and some attributes of those products (name, brand, description, color, etc). sessions_train contains a list of items that a user visited in a session along with the next item visited in the session.
- Project Goals
 - Predict the next item that will be clicked on by the user in a session, given the previous items clicked on in the session
 - May return a list of the N most likely items to form a recommendation list
- Models
 - We plan to use some sort of graph neural network, exact specifics unknown
 - Will almost surely include transformers/NLP models to process text data
- Rough Timeline
 - 3/28-4/4: Explore the data, basic EDA, use EDA results to trim data to non-rare products. Research methods/models for processing training data
 - 4/4-4/11: Setting up data, preliminary modeling
 - 4/11-4/16: Refining models, realistically fixing errors from previous week, creating presentation
 - 4/16-4/25: Clean repository, finalize models, write report