H&M RECOMMENDATION SYSTEM

Team TrainDy: Aditya P., Prashanthi R., Shashidhar P., Shipra P.

H&M Group is a family of brands and businesses with 53 online markets and approximately 4,850 stores. With such an extensive range of options to shop, it is very difficult for customers to find what they want or are interested in quickly. In such a scenario, good recommendations become key in helping them find the right products.

INTRODUCTION

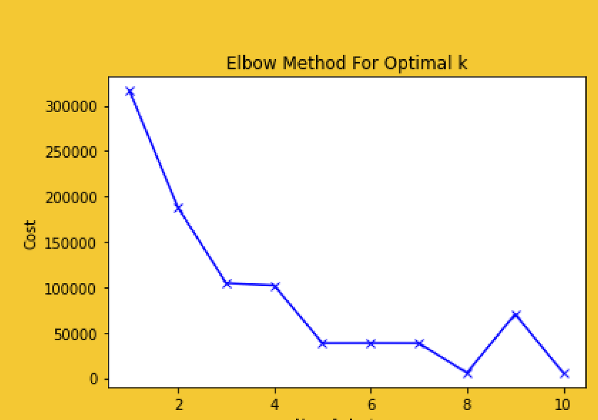
In this project, we have built a recommendation engine for H&M. We generate recommendations for a customer based on the consumption patterns of other similar users (a.k.a collaborative filtering). We use k-modes clustering to cluster similar users together based on customer metadata. To generate recommendations, we input user purchase patterns and user cluster information to a conditional VAE. This VAE trains to obtain a latent space that can then be used to sample new user purchase matrices for a specified distribution of users.

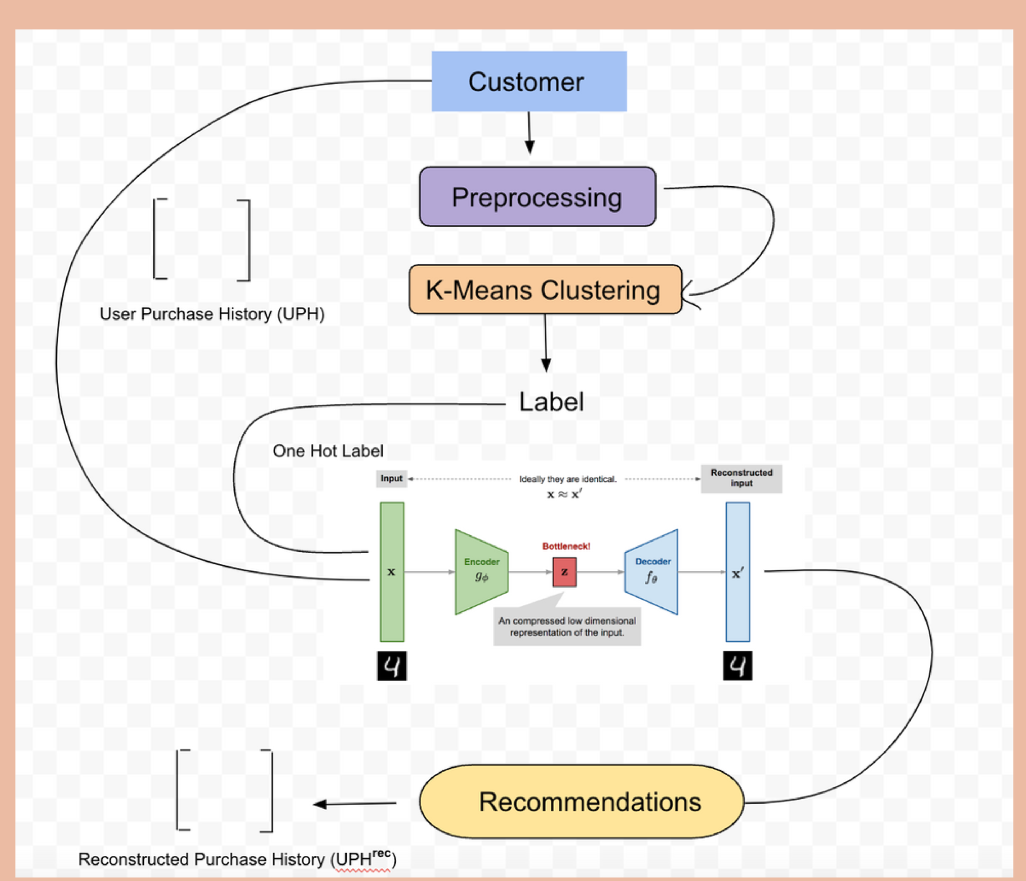
DATA

- The (anonymized) data and metadata of H&M customer purchases across time. (2017-2019). Source: Kaggle
  - articles.csv: Information about H&M products (article ID, description, etc.)
  - transactions.csv: Transaction information (customer ID, timestamp, article ID, etc.)
  - customers.csv: Customer metadata (customer ID, activity, etc.)

METHODOLOGY

- **Data:** We split our data into two parts: transactions between 2017-2018 (for training) and transactions in 2019 (for evaluation).
- **Clustering customers**: Since customer metadata attributes in the dataset were categorical, we used **k-modes clustering** to classify customers into 5 clusters (based on the elbow curve below).
- **User purchase history (UPH):** We created a matrix with customers as rows and articles as columns. This matrix would contain "1" if a user has purchased that product and "0" otherwise.
- **Recommendation engine:** We then trained a Conditional VAE on customer UPH to obtain a latent space that the model would use to sample a new UPH for a customer
- **Success Metric:** If a customer purchases at least 30% of the products recommended to them. (recall_30)





Our recommendation system has been illustrated in this diagram. We preprocess customer data, perform customer clustering and then feed customer class label and purchase history to the CVAE model to obtain recommendations

ETHICS

- **Positive implications for sustainability:** good recommendation reduce returns and carbon emissions from transportation
- **Senstive user information:** It is important to be careful as the recommendations can reveal sensitive user information
- **Systemic biases:** Customer preferences and consumption patterns can cause typecasting of people (gender, age, and neighborhood)

CHALLENGES

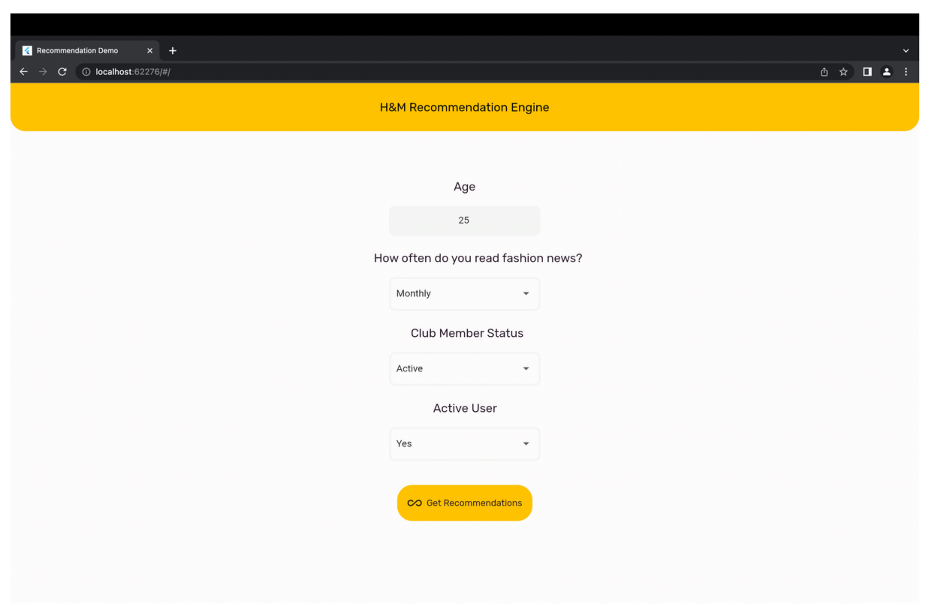
- Highly **sparse nature** of the UPH matrix: Model cheats by learning to output all zeroes while maintaining a low loss. We fixed this by using cosine-similarity/recall-based loss functions.
- **Categorical customer data:** Could not use k-means for customer clustering since customer data is not continuous/numerical. We fixed this by using k-modes clustering instead.
- **Cold starts:** For new customers, there is no past information and that deters good, personalized recommendations. We handle this by classifying the customer into a cluster and then sampling a UPH vector from the distribution of that cluster.

RESULTS AND DEMO

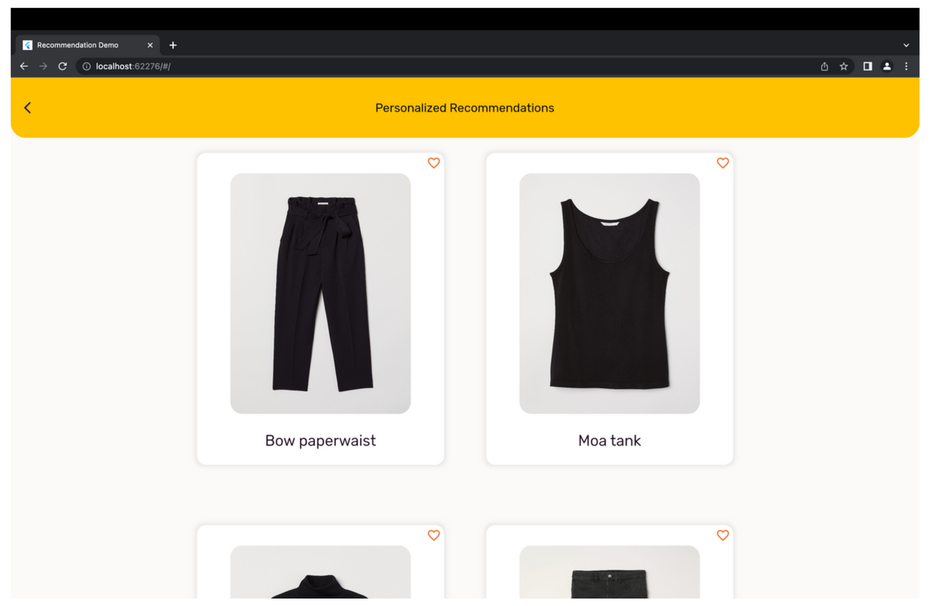
We evaluated our model using two metrics: recall_1 and recall_30:
- recall_1 is a rudimentary metric that is defined as: if a customer purchases even one of the top 20 recommended articles, it is a success
- recall_30: if a customer purchases 30% of the top-recommended articles, it is a success

We trained the model on 70% of the dataset and tested it on 30% of the dataset and obtained a recall_1 of 0.73 and a recall_30 of 0.30.

metric	success score
recall_1	0.73
recall_30	0.30



Demo of the TrainDy Web App: input user demographic



Demo of the TrainDy Web App: Recommendations

CONCLUSION

We would like to thank Prof. De Stefani and the TAs, especially our mentor TA, Thien Nguyen for their timely guidance and help.