

# Topic and group proposal

*“Recommendation System”*: Exploration and Comparison of current methods based on ML  
Advanced ML Course Project,

Fall 2024

Zakarya El Mimouni , Ahmed Khairaldin, Amine RAZIG

*This summary note aims to describe of the general topic, what we plan to do for your the project, our references (books, papers, blog posts, github repos, etc.) for the topic, the dataset we will use, and a brief description of how you plan to implement your code.*

*Note: the project references are likely to be completed/enriched later, as are the ML methods, depending on the difficulties encountered in the project.*

## Description of the General topic:

A recommendation system is a class of machine learning tools designed to suggest relevant items to users based on their preferences, behaviors, and other users' activities. They are widely used across e-commerce, streaming platforms, social media, and online advertising, aiming to enhance user experience by delivering personalized content or product suggestions.

## Database:

- <https://nijianmo.github.io/amazon/index.html#complete-data>

## Methods :

There are three types of collaboration systems: the « collaborative filtering », « Content-Based Filtering » and hybrid recommendation systems witch combine the two previous ones.

### 1. Collaborative Filtering

Collaborative filtering recommends products to users based on the behavior of other users with similar preferences. CF methods work on the principle that users who agreed on items in the past are likely to agree again. There are two main types:

- User-Based Collaborative Filtering: This approach identifies users who have similar preferences (based on ratings or clicks) and recommends items that similar users liked.
- Item-Based Collaborative Filtering: Instead of focusing on user similarity, this method finds items that are similar based on user ratings or interactions.

We can try to implement this method using: **matrix factorization techniques** like **Singular Value Decomposition**, which reduce the dimensionality of the data matrix, capturing latent factors that explain user-item interactions.

### 2. Content-Based Filtering

Content-based filtering recommends products based on item characteristics and user preferences. This approach analyzes the features of items (e.g., category, brand, price, description, etc.) and suggests items similar to those the user has interacted with in the past.

In content-based Filtering we can use **NLP** similarity techniques such as tf-idf and Cosine similarity.

Otherwise, we also can use **Neural Networks**: embeddings (Word2Vec, BERT embeddings) for textual data, we can capture richer semantic information about items.

### 3. Hybrid Recommender Systems

Hybrid methods combine collaborative filtering and content-based filtering to leverage their strengths and mitigate their individual weaknesses. In the project we can try some popular hybrid approaches include:

- **Weighted Hybrid**: This approach assigns different weights to recommendations generated by collaborative filtering and content-based filtering.
- **Feature-Augmented Collaborative Filtering**: In this hybrid, content-based features are added to collaborative filtering algorithms. For instance, user preferences and item characteristics are used as input features in matrix factorization, allowing the system to consider both user preferences and item content.
- **Deep Hybrid Models**: Neural networks allow the combination of user and item embeddings from both collaborative and content-based filtering. A multi-layer perceptron or a convolutional neural network can be trained to learn relationships between users and items based on both behavioral and content-based data.

### 4. Evaluation

To assess the performance of our recommendation system, we will use some metrics like metrics like Mean Absolute Error, Root Mean Squared Error, Precision, Recall, etc..

## Flow of the Code Project (All in Python) :

Data Preprocessing:

- Preprocess textual data using tokenization, stemming, and vectorization techniques like TF-IDF or word embeddings.

Building Collaborative Filtering Models:

- Implement SVD for matrix factorization.
- Build user-based and item-based collaborative filtering models using libraries like Scikit-Learn.

Building Content-Based Filtering Models:

- Create item profiles using metadata.
- Use cosine similarity or neural embeddings to identify similar items.

Combine with Hybrid Techniques:

- Experiment with hybrid models (weighted hybrid, feature-augmented collaborative filtering, etc.) to combine collaborative and content-based methods.
- Train deep hybrid models if using neural networks, concatenating collaborative and content-based embeddings as input.

Visualization:

- Récapitulatifs des des résultats et visualisation par des graphiques dans la mesure du possible

## References :

Papers :

- "Item-Based Collaborative Filtering Recommendation Algorithms", Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl (2001)
- Empirical Analysis of Predictive Algorithms for Collaborative Filtering, John S. Breese ,David Heckerman , Carl Kadie
- "Matrix Factorization Techniques for Recommender Systems", Authors: Yehuda Koren, Robert Bell, Chris Volinsky (2009)
- "Deep Neural Networks for YouTube Recommendations", Paul Covington, Jay Adams, Emre Sargin (2016)
- Recent Developments in Recommender Systems: A Survey  
Yang Li, Member, IEEE, Kangbo Liu, Ranjan Satapathy, Suhan Wang, Member, IEEE, and Erik Cambria, Fellow, IEEE
- "Neural Collaborative Filtering", Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, Tat-Seng Chua (2017)

Web :

- [https://en.wikipedia.org/wiki/Singular\\_value\\_decomposition](https://en.wikipedia.org/wiki/Singular_value_decomposition)
- <https://youtu.be/G4MBc40rQ2k?si=6cbPoP8LaufhoBSd>
- <https://www.youtube.com/playlist?list=PLQY2H8rRoyvy2MiyUBz5RWZr5MPFkV3qz>

Not available :

- "Content-Based Recommendation Systems", Michael J. Pazzani, Daniel Billsus (2007)

Sujet annexe :

## **Image-based Recommendations on Styles and Substitutes**

<https://cseweb.ucsd.edu/~jmcauley/pdfs/sigir15.pdf>