Group Project

Instructions:

1.This is a group project. Each group can have up to 5 members. Each group should have a leader and submits the list of group members to the TA by **March 25th**.
2.Each group is required to deliver a 5-minute in-class presentation of your project. At the same time, each group need to submit a **technical report**, in which the process of data processing and analysis, findings and conclusions are required to be clearly stated. **Technical report and source code** must be submitted via canvas.
3.The project scope is open-ended. You are encouraged to explore innovative ideas using any combination of data analysis, machine learning models, and visualization techniques.
4.Should you have any questions, please feel free to consult with the instructor or TA directly after class or via email.

Project Overview:

In this project, you will implement, evaluate, and analyze the performance of various recommender system algorithms on real-world datasets. The focus is on algorithmic understanding, experimental rigor, and empirical analysis. You will explore the strengths and weaknesses of different recommendation approaches and investigate how they perform under various conditions and evaluation metrics.

Requirements:

1. Algorithm Implementation (35%)

Implement **at least five** recommendation algorithms from different paradigms:

Traditional collaborative filtering (user-based, item-based)
https://github.com/datawhalechina/fun-rec/blob/master/docs/ch02/ch2.1/ch2.1.1/itemcf.md
Matrix factorization methods (SVD, SVD++, NMF, or ALS)
Advanced algorithms (choose at least two):
Graph-based (LightGCN, NGCF, GraphRec)
LightGCN: https://arxiv.org/abs/2002.02126
NGCF: https://arxiv.org/abs/1905.08108
GraphRec: https://arxiv.org/pdf/1902.07243

Sequential models (GRU4Rec, SASRec, BERT4Rec)
GRU4Rec: https://arxiv.org/abs/1511.06939
SASRec: https://arxiv.org/abs/1808.09781
BERT4Rec: https://arxiv.org/abs/1904.06690

2. Experimental Design (15%)

Design and conduct comprehensive experiments that:

1. Utilize proper data splitting techniques (random, temporal, user-based)
2. Employ cross-validation where appropriate
3. Include baseline comparisons (e.g., popularity, random)
4. Analyze sensitivity to key hyperparameters
5. Address data sparsity and cold-start scenarios

3. Evaluation Methodology (5%)

Implement and apply appropriate evaluation metrics: HR@K, NDCG@K, MRR@K

4. Analysis and Insights (25%)

Provide in-depth analysis of your experimental results:

1. Compare algorithm performance across different metrics
2. Identify strengths and weaknesses of each approach
3. Visualize results to highlight patterns and insights

Technical Report Format and Structure:

Your technical report must follow the ACM Conference Proceedings double-column format and include:

Abstract; Introduction; Related Work; Methodology; Experimental Setup; Results;

Datasets:

1. MovieLens(1M): https://grouplens.org/datasets/movielens/
2. Last.FM: https://grouplens.org/datasets/hetrec-2011/
3. Yelp Academic Dataset: https://business.yelp.com/data/resources/open-dataset/

Maybe a helpful tutorial for you: https://github.com/Yelp/dataset-examples

Code Requirements (20%):

Your code submission must include:

Implementation Files:

1. Separate modules for each algorithm implemented
2. Data loading and preprocessing utilities
3. Evaluation metrics implementation
4. Experiment runner scripts

Documentation:

1. README.md with clear instructions for running experiments
2. Code comments explaining complex logic
3. Function/class documentation following standard format
4. Requirements.txt or environment.yml

Experiment Scripts:

1. Scripts to reproduce all experiments in the paper
2. Configuration files for hyperparameter settings
3. Notebooks for exploratory data analysis
4. Visualization code for generating paper figures