CS-4053 Recommender System

Fall 2023

Lecture 4: Designing a Recommendation System *(Part 1)*

Course Instructor: Syed Zain Ul Hassan

National University of Computer and Emerging Sciences, Karachi

Email: zain.hassan@nu.edu.pk

Available Languages

- We can develop a recommendation engine from scratch using any of the high-level programming languages such as:
  - Python
  - Java
  - C#
  - C++
  - Julia
  - *Others…*

Python Implementation Packages

- Surprise (disallowed for project)
- Scikit-learn
- numpy
- pandas
- Word2vec
- python-recsys (disallowed for project)

- Note that we don’t always need all of these packages

C# Implementation Packages

- ML.NET
- Microsoft ML.Recommender
- NReco (a port of Mahout)
  - Disallowed for project

- Note that we don’t always need all of these packages

Datasets

- MovieLens
  - MovieLens 1m
  - MovieLens 100k
  - MovieLens 10m
  - MovieLens 20m
  - MovieLens 25m
- IMDB
  - IMDB top 250 list

Datasets

- Netflix Prize dataset
- Book-crossing dataset
- Jester
- Anime recommendation dataset
- Amazon 233m

Course Project

- The project will be of 10 marks
- You can use any of the following programming languages:
  - Python
  - C#
  - Java
  - R
- Tentative submission deadline is a week prior to the finals

Course Project

- You are allowed to use any of the public datasets that are available
- The dataset must be available at the time of demonstration

Course Project: Phases

- **Proposal Submission**
  - In 5th week (tentative)
  - Proposal must contain information on the dataset, technique(s) and language to be used
- **Review of Proposal**
  - Decision at the end of 5th week (tentative)

Course Project: Deliverables

- A working (testable) recommendation system
  - The model (technique) used must be explainable and should not be a blackbox

- A proper report with:
  - Introduction
  - Dataset details
  - Model explanation
  - Experiments
  - Discussion on results

Course Project: Evaluation

- Novelty of the technique used **20%**
- Performance of the system (to be measured by metric) **20%**
- Performance of the system (based on human judgement) **20%**
- Quality of documentation **20%**
- Honesty and effort **20%**