IT492: Recommendation Systems

Lab Assignment - 03
[Deadline: 27th March 2022, Sunday 11:59 PM]

This assignment involves a comparison of evaluation metrics when the recommendation is modeled as a (i) rating prediction, (ii) classification, and (iii) ranking problem.

Marking scheme and requirements

Full marks will be given for (1) working, readable, reasonably efficient, documented code that achieves the assignment goals and (2) for providing appropriate answers to the questions in your Google colab file (`name format: LA03_rollnumber_name`) submitted via Google Classroom on the **assigned dataset only**. Please find the dataset allotted to you below.

Please refer to the Dataset Allocation

Sr No	Student Id	Student Name	Student email	Registration Type	For LA-03
1	202118004	ABHISHEK SINGH	202118004@daiict.ac.in	AUDIT	Food Reviews (Kaggle 2019)
2	202018004	P SARAN PANDIAN	202018004@daiict.ac.in	AUDIT	
3	202018026	AAKANKSHA SHAH	202018026@daiict.ac.in	AUDIT	
4	202111002	SHARMA HARSH DHARMENDRAKUMAR	202111002@daiict.ac.in	AUDIT	
5	202111029	GORASIYA RAGHAV NARESH	202111029@daiict.ac.in	AUDIT	
6	202018042	ABHIJEET KUMAR	202018042@daiict.ac.in	REGULARADD	last.FM (Hetrec 2011)
7	202111010	KEVIN JITENDRABHAI JADIYA	202111010@daiict.ac.in	REGULARADD	
8	202111035	VANSH RAHUL BHANJIBHAI	202111035@daiict.ac.in	REGULARADD	
9	202111048	MANSURI PINJARA MOHAMMED JUNED HANIFBHAI	202111048@daiict.ac.in	REGULARADD	
10	~~202112030~~	~~ARPITHA SREENIVASAN~~	~~202112030@daiict.ac.in~~	~~REGULARADD~~	
11	201801466	PARMAR SIDDHRAJ YOGESHBHAI	201801466@daiict.ac.in	REGULARADD	
12	202121004	SANDHYA KUMARI	202121004@daiict.ac.in	REGULARADD	
13	202116003	AMBUJ MISHRA	202116003@daiict.ac.in	REGULAR	Movielens 20M (Grouplens 2016)
14	202116004	ARPITA NEMA	202116004@daiict.ac.in	REGULAR	
15	202116008	RAHUL KUMAR	202116008@daiict.ac.in	REGULAR	
16	202116009	RAHUL THAKUR	202116009@daiict.ac.in	REGULAR	
17	202116011	ROHAN BAGHEL	202116011@daiict.ac.in	REGULAR	
18	202116001	ABHISHEK YADAV	202116001@daiict.ac.in	REGULAR	
19	202116002	AKSHAY KAUSHIK	202116002@daiict.ac.in	REGULAR	

Links to download the Datasets

- *last.FM (Hetrec 2011)*
- *Movielens 20M (Grouplens 2016)*
- *Food Reviews (Kaggle 2019)*

IT492: Recommendation Systems

Please adhere to the lab policy on the course website

- Cite resources and give credit where it's due. If you happen to discuss the questions with your peers, please mention your collaborators in your report/assignments.

- Acts of plagiarism will not be tolerated and will result in a straight ZERO for that assignment.

- Students who don't submit their assignment by 27th March 2022, Tuesday 11:59 PM will simply get ZERO.

Main Assignment (15 Marks in Total)

Dataset Analysis (2 Mark)

1. Explore the dataset and present a summary of the same (e.g. number of users and items, minimum/maximum/average number of ratings for an item, information apart from ratings present in the dataset). Plot a histogram of item ratings to visualize the distribution.

Recommendation (7+3 = 10 Marks)

2. Compare the performance of modeling recommendation problem as the following:

   a. Rating prediction
   b. Classification task
   c. Ranking a list problem

   Evaluate the performance with RMSE and MAE for rating prediction, precision, recall and F1-score for classification, and nDCG and MRR for ranking. Implement the evaluation metrics from scratch. You are free to use models of your choice for each of the above parts and can use Surprise and Sklearn. Also, show the surprise, diversity, novelty, and popularity of the recommended items.

Re-ranking Approaches (3 Marks)

3. Use MF to find out top-100 recommendations for twenty user. Apply re-ranked approaches on these top-100 items to diversify and finally select top-10 recommendations for each user. Now, determine a relation between diversity and relevance of the top-10 recommendations for the selected users.