Assignment 3 Due Date - 19th November 10 marks

This is a challenge problem. Before you start working, you should read about the dataset here. It is a dataset collected and curated at IIITD. So you won't find any paper using the dataset apart from the one mentioned.

So far you have primarily used the ratings information for recommendations. I briefly discussed in class how associated metadata (Movielenes 100K and 1M) can be used for improving accuracy. This dataset has a lot more associated information for users and movies.

The challenge is to get the best results on this dataset. You are free to use anything under the sun. Implementations available in Github or from author's sites or your implementation of papers …. Anything is acceptable.

You have to run the standard 5 fold cross validation and compute the NMAE and RMSE for each fold and also the average over 5 folds. The 5 folds will be given to you.

You will be marked based on your performance and effort, i.e. the best results will be given top marks. You will also have to show the results from the methods you have tried.

IF I SEE IDENTICAL RESULTS (BEST TECHNIQUES AND TECHNIQUES TRIED) FOR TWO DIFFERENT PEOPLE BOTH WILL BE CHARGED WITH PLAGIARISM.

You have to fill the following table. You can try 2-3 or more methods and show all the results.

	Method 1	Method 2	…..	Best Method
Fold-1				
Fold-2				
Fold-3				
Fold-4				
Fold-5				
Average				

References:

[1] Method 1 Ref

[2] Method 2 Ref

.

.

.

[n] Best Method Ref

During evaluation you may be asked to reproduce one or more results from the table and asked questions on the method you used. TA's will use their discretion to award final marks based on your effort and understanding.