

Movie Recommendation System – Milestone-2

This week I mostly started with the EDA of the dataset. The dataset contains 25 million records of movies. The ratings table I mostly concentrating on. I will be linking ratings table with the movies table to associate genres with the movies.

Any surprises from your domain from these data?

Till now, I have not faced any surprises from this domain as many of the terms already known. However, I need to still understand how genre tags and genre scores can be used.

The dataset is what you thought it was?

The dataset is mostly cleaned-up and pre-processed. However, the genres and title columns are of complex types. Title column is having movie year embedded in it while genres column is “|” pipe separated strings column. I will be processing these columns to create additional columns.

Have you had to adjust your approach or research questions?

For this week, I mostly worked on portfolio. I completed preliminary EDA for this dataset and I have not found anything that might lead to change of approach. Next Week I will be completing EDA as well as basic model, that will give me more idea if I need to change my approach or need more research.

Is your method working?

I have not faced any blocker challenge till now. I couldn't start with model creation; so, I think I can answer this question next week.

What challenges are you having?

As the size of dataset is big, I am seeing some challenges while loading this data. Also, I am not aware how to use some of the columns/feature, so I need to learn more about those columns before moving to model creation.

As I spent most of my time on portfolio creation on GitHub, I could not progress much on the project.

However, I will try to cover most of the things next week for this project.