

CS 5683: Big Data Analytics

Project-1 - Data Files

README

All datasets given in this project are sampled versions of data openly published as ground-truth datasets¹ for recommendations systems.

views.txt

This file contains movie watch history of **5982** customers. Each line in this contains watch history of one customer. *We do not provide user-ids in this data as it is irrelevant for Project-1.*

An example line present in **views.txt**:

3114::1721::2687::1028

, where each integer given above represent a movie ID. You can find ID's up to 3900. Each ID is separated by two colons (::) to make it easy for you to process the data.

mappings.txt

This file act as a dictionary to map IDs to movie names. Each line represent a ID to name mapping. Example:

9::Sudden Death (1995)
10::GoldenEye (1995)

, **Sudden Death (1995)** and **GoldenEye (1995)** are movie names of IDs **9** and **10** respectively. The ID and its corresponding movie name is separated by two colons (::) to make it easy for you to process the data.

get_recommendations.txt

This file is used only for the **bonus task**. The input lines of this file look very similar to **views.txt**. Each line in this file represents watch history of one customer. For each watch history in this file, the bonus task is to get 3 movie recommendation using Association Rules you extracted in the previous task.

¹ DOI=<http://dx.doi.org/10.1145/2827872>