

Projet 1

SUMMARY

=====

This dataset was constructed to support participants in the Netflix Prize. See <http://www.netflixprize.com> for details about the prize.

The movie rating files contain over 100 million ratings from 480 thousand randomly-chosen, anonymous Netflix customers over 17 thousand movie titles. The data were collected between October, 1998 and December, 2005 and reflect the distribution of all ratings received during this period. The ratings are on a scale from 1 to 5 (integral) stars. To protect customer privacy, each customer id has been replaced with a randomly-assigned id. The date of each rating and the title and year of release for each movie id are also provided.

USAGE LICENSE

=====

Netflix can not guarantee the correctness of the data, its suitability for any particular purpose, or the validity of results based on the use of the data set. The data set may be used for any research purposes under the following conditions:

- * The user may not state or imply any endorsement from Netflix.
- * The user must acknowledge the use of the data set in publications resulting from the use of the data set, and must send us an electronic or paper copy of those publications.
- * The user may not redistribute the data without separate permission.
- * The user may not use this information for any commercial or revenue-bearing purposes without first obtaining permission from Netflix.

If you have any further questions or comments, please contact the Prize administrator <prizemaster@netflix.com>

TRAINING DATASET FILE DESCRIPTION

=====

The file "training_set.tar" is a tar of a directory containing 17770 files, one

per movie. The first line of each file contains the movie id followed by a colon. Each subsequent line in the file corresponds to a rating from a customer and its date in the following format:

CustomerID,Rating,Date

- MovieIDs range from 1 to 17770 sequentially.
- CustomerIDs range from 1 to 2649429, with gaps. There are 480189 users.
- Ratings are on a five star (integral) scale from 1 to 5.
- Dates have the format YYYY-MM-DD.

MOVIES FILE DESCRIPTION

=====

Movie information in "movie_titles.txt" is in the following format:

MovieID,YearOfRelease,Title

- MovieID do not correspond to actual Netflix movie ids or IMDB movie ids.
- YearOfRelease can range from 1890 to 2005 and may correspond to the release of corresponding DVD, not necessarily its theatrical release.
- Title is the Netflix movie title and may not correspond to titles used on other sites. Titles are in English.