**Dataset Recommender of RMDSLAB**

Objective:

RMDS wants to recommend RMDS datasets to RMDS users customly based on datasets similarity and popularity. We would calculate out a recommendation score for each dataset and recommend those datasets with high recommendation scores to user. For public demonstration purpose, this project used generated data instead of read data. The process of fake data generation data is included as well.

1. **text\_gen.py(optional)**
   * generate fake dataset using aitextgen(original data not provided)
   * input: NA(trained models)
   * output: JSON format of list of projects and JSON format of dataset’s title and description
2. top\_click.py
   * generate most popular projects click history by user
   * input: NA
   * output: csv format of project click history for each user
3. file\_gen.py
   * generate project records and user upload history
   * input: project title and description
   * output: csv format of project with description, list of project and upload history
4. recom\_table.py
   * generate data for dataset including dataset information, view and download
   * input: NA
   * output: csv format of basic dataset information
5. recom.py
   * recommend 5 datasets to each user based on user similarity and datasets similarity
   * input: project record, dataset record
   * output: csv format of recommendation for each user
6. dataset\_fake.ipynb

This is a Jupyter notebook of the complete process of the recommendation system(except text gen). It could be used as an overview of the project.