Sudha Vijayakumar   **SJSU.ID**: 014749488

Hybrid Rating Prediction for Book Recommendation System

Learning Objective/ Goal
1) To predict book ratings for the given 'train.csv' using different approaches.
2) To use the book's metadata, learn and build a memory-based recommendation system listed in the article

Source code: https://github.com/sudha-vijayakumar/CMPE256-BOOK-RECOMMENDATION-SYSTEM.git

Input files (7)   -   **input/metadata/** goodreads_books_children.json.gz, goodreads_reviews_children.json.gz
-    **input/** train.csv, test.csv
-    **processed_data/** books.csv, reviews.csv, books_with_review.csv (will be generated at step-1,2

Output (3)
-    **output/** predictions-svd-75.csv, predictions_Fast_ai.csv, final_submission.csv, book_with_reviews.html, User_Ratings.html

HOW TO RUN | ML Pipeline (Run the notebooks in the below order)

No.	STEP	Filename
1	Read metadata as JSON	1_JSON_To_CSV.ipynb
2	Merge reviews with books	2_Merge_reviews_books_user_data.ipynb
3	EDA	3_EDA_using_Pandas.ipynb

4	**Model Training**
	1) Identifying the best model for the train data
	    a) AutoSurprise - SVD
	
	    b) Fastai - Matrix Factorization
	
	2) Train & make predictions for the test.csv using Surprise - SVD[Matrix Factorization]
		4.0_Auto_Surprise_BestModel.ipynb
		Run-time: limited to 6hours
		
		4.1_Surprise-SVD_Book_Rating_Prediction.ipynb
		Run-time: 1hour
		
		4.2_Fastai_Book_Rating_Prediction.ipynb
		Run-time: 50mins
		
		Submission: [0.7 * 4.1_(prediction)] + [0.3 * 4.2_(prediction)]
		
		6_submission.ipynb

5	**Recommendation System** Collaborative filter - Memory-based - PearsonR	5_Recommendation_using_Colab_Filtering.ipynb

Learning outcome
1) Experimented with different approaches on how to predict ratings for the given dataset.
2) Built basic recommendation systems using the practical approaches taught in this course(Collaborative filtering using Pearson similarity).

Future work To try HGN baseline for sequential recommendation (**Rank_1) Source**