CMPE 256 - Advanced Data Mining
Spring 2022
Group 1

Project Proposal

Team members: Harshil Bharatkumar Darji, Vineeth Chitturi, David Archer

Team name: Havida

Project Title: Infinite Books

Project Description:

We propose to build a book recommendation system based on book ratings and user data. We plan to use data from "Goodreads" to learn and explore some subtle and advanced skills in the recommendation field. Then we'll set the evaluation model to pick the best one in our experimental solutions.

Proposed Methodology and Dataset:

Dataset Link: https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/home
From the above website we will be using two datasets, "Books" and either one of "Shelves" or "Reviews" to solve our problem statement. The datasets were collected in late 2017 from goodreads.com, where we only scraped users' *public* shelves, i.e. everyone can see it on the web without login.

Methodology:

Using the datasets, we aim to recommend books to users either by genre or by book title. Since the dataset is huge, we will try to use various techniques to provide recommendations in an optimized way. The implementation will be done in Python 3.10, and all testing will be in Jupyter Notebook. We will be using Collaborative based filtering and other Machine Learning Methods like SVD, Alternating Least Squares etc. to predict ratings. We will also try to predict ratings using third party libraries like surprise. We will try some different method combinations and the exact methodology will depend on the progress.