Project Proposal

Introduction

The purpose of this proposal is to give an overview of the recommendation system for books. In general, the recommendation system is used to predict the items that may be of interest to the user according to the user's preferences and behaviors. In this proposal, a recommendation system is used to recommend books based on book reviews and book titles. This will build trust between the user and the service provider. This proposal will describe the data source, size and columns, the algorithm, and tools used to build this recommendation system.

Design

Book recommendation based on book reviews and book title

Data

About the dataset:

- Books: information of each book.
- Ratings: User rating records for each book.
- Users: User information (Age, Location)

Explore datasets:

- The books dataset is containing 271360 entries, 5 columns
- The user's dataset is containing 278858 entries, 3 columns
- Ratings dataset is containing 526356 entries, 3 columns

Algorithms

The book recommendation system will build using a recommendation system: user-based collaborative filtering using the nearest neighbors algorithm. I will use two models the first model is the nearest neighbors (Classification) model, using a brute algorithm, which will calculate the find the distance of every point to every other point. to make a prediction I will use the result of the nearest neighbors model into the K nearest (Clustering) model.

Tools

Data processing: pandas, Numpy.

Modelling: The nearest neighbors model, using a brute algorithm and the K nearest.

Visualization: Matplotlib.

Writing the code: Jupyter