



## **Problem Statement:**

Recommender systems are information search and decision support tools used when there is a set of options or when the user lacks the domain-specific knowledge essential to take decision. It differentiates the data based on various algorithms and suggests the user with the most relevant product. The suggestion of the products is based on the customers' traits and needs.

In the book recommendation system,

- The aim of the proposed system is to provide one single platform that provides recommendations of most popular book among all the users for providing better user experience and helping the user to keep up with the trend.
- The proposed system provides various types of recommendations on the basis of book author, publication, ratings, location, age that in turn enhances the quality of information provided by system.

## **Dataset :**

We are using the Book-Crossing Dataset. It contains 278,858 users with demographic information providing 1,149,780 ratings about 271,379 books.

Link for the dataset:

<https://grouplens.org/datasets/book-crossing/>

The above dataset contains 3 tables:

- **BX-Users :** Contains information regarding the user such as
  - User ID
  - Location
  - Age
- **BX-Books:** Contains information regarding the user such as
  - isbn
  - book\_title
  - book\_author

- year\_of\_publication
- publisher
- image\_url\_s
- image\_url\_m
- image\_url\_l

`book-title`, `book\_author`, `year-of-publication`, `publisher` values are obtained from Amazon Web Services.

URLs linking to cover images are also given, appearing in three different flavours (*Image-URL-S*, *Image-URL-M*, *Image-URL-L*), i.e., small, medium, large. These URLs point to the Amazon web site.

- **BX-Book-Ratings:** Contain book rating information such has
  - user\_id
  - isbn
  - book\_rating( It is expressed on a scale from 1-10)

## **Related Works:**

We have referred the following articles regarding the recommender system. The research papers focuses on analysing the k nearest neighbors collaborative filtering and Item based collaborative filtering techniques. They have looked into different techniques to compute item-item similarities based on correlation vs. cosine similarities between item vectors. The authors have used different techniques for obtaining recommendations from them (weighted sum vs. regression models). They have experimentally evaluated the results and compare the results of the various recommendation algorithms and have suggested that item-based algorithms provide better performance and quality than user-based recommendation algorithms.

- [Item-Based Collaborative Filtering Recommendation Algorithms](#)
- [Understanding Basics of Recommendation Engines](#)
- [Quick guide to build Recommendation Engine in Python](#)
- <https://ieeexplore.ieee.org/document/7684166>

## **Approach Overview:**

We are aiming to develop a book recommendation system that suggests a book to the user considering different criteria such as user ratings; author or publication based on user provided various attributes of a book. The following steps would be taken in developing the model:

- **Data Pre-processing:** We will pre-process the above mentioned dataset so that we can transform the raw data into understandable format that can be used for further processing in our models. We will check for any inconsistent, incomplete and/or lacking behaviours in the data such as null values and encode the categorical values to necessary numerical values.
- **Recommendation:** It provides book domain related recommendation to the user. The model will be an all in one recommendation guide for customer searching for books.

## **Technology Stack:**

- Python,
- Pandas and
- Scikit-learn.

## **Algorithms:**

The various types of algorithms used for the recommendation systems that we are going to implement are:

- **K-Nearest Neighbor Algorithm:**

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. We will be using kNN as a classifier to implement our algorithm. We will use kNN to recommend similar books based on the similarity in the attributes such as author, publication, age and year the book was published for the given book.

- **Collaborative Filtering**

Collaborative filtering is executed in two stages:

First, similarity between all the items is found. In user-user systems, similarity functions can use normalized ratings.

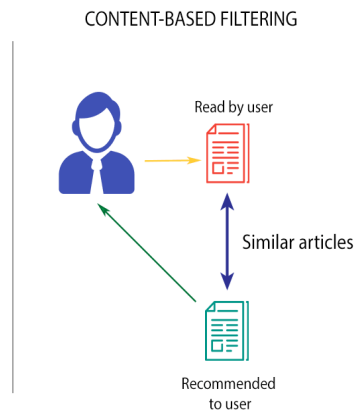
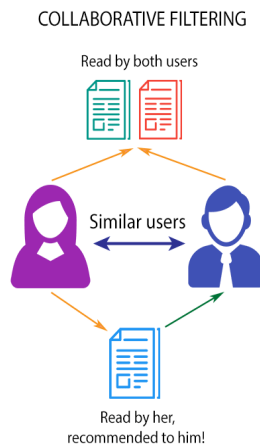
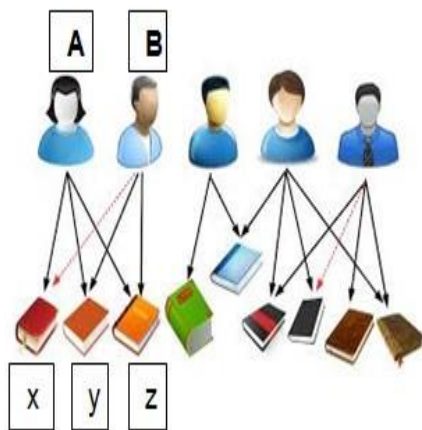
Second, the system executes a recommendation stage. It uses the most similar items to a user's already-rated items to generate a list of recommendations. Usually this calculation is a weighted sum or linear regression.

- **User Based Collaborative Filtering:**

The Collaborative filtering based algorithms unlike content based algorithms are based on the past behaviour and not the context. In User based collaborative filtering we find look alike users based on similarity metrics and offer recommendation which similar users have chosen in the past. It computes every pair of similarity between the users and tries to recommend items based on user's having similar interest. Thus we will be able to recommend books based on the similarity of users based on their age-group and location.

- **Item Based Collaborative Filtering:**

The Item based Collaborative filtering is similar to the previous mentioned algorithm but instead of finding the similar users, we try to find the similar books. We will create a matrix for similar items for the given book based on which we can easily recommend the books similar to given books based on author, publications etc.



Collaborative Filtering: "People who read book X, will also tend to read book Y highly, and you haven't read item Y yet, so you should try it"

### **Team Members:**

- KuldeepSinh Chudasama - 801045027
- Sneha Gopinath - 801053588
- Nisarg Shah - 801060905
- Freny Savalia – 801075313

### **Timelines:**

Outline of the tasks are as follows:

### **Project Timeline:**

- 20th September: Literature Survey.
- 2nd October: Project Proposal submission.
- 4th October: Data Gathering and Pre-processing.
- 10th October: Understanding the algorithms.
- 18th October: Model implementation.
- 30th October: Mid-Progress Report.
- 10th November: Training the models.
- 22nd November: Model evaluation.
- 28th November: Final Report
- 4th December: Submission of Final Report.
- 11th December: Poster Presentation for the built recommendation system.

Tasks	Names
Exploratory Data Analysis	Sneha Gopinath, Freny Savalia

Data Preprocessing	Nisarg Shah, KuldeepSinh Chudasama
KNN Algorithm	Freny Savalia, Nisarg Shah
Item Based Collaborative Filtering	KuldeepSinh Chudasama, Nisarg Shah
User Based Collaborative Filtering	Sneha Gopinath, Freny Savalia
Training the models	Sneha Gopinath, KuldeepSinh Chudasama
Model Evaluation	All the members of the team
Proposal and Report	All the members of the team

## **Questions This Project Will Answer:**

This project will answer the following questions:

- How machine learning models can be useful for real life recommendation scenario?
- Which machine learning algorithm achieves best results for our system?
- Detail Analysis of the different book attributes
- Unified Platform for book search on the attribute given by the user

## **Things That We Expect to Learn:**

In this project, we will learn how to build and evaluate a machine learning model in real life scenario. This includes the following:

- Data pre-processing and refining for optimal use in the system.
- Understanding various types of recommendation techniques: K Nearest Neighbors, User-based and Item-based Collaborative Filtering.
- Practical implementation of various machine learning algorithms to build a good recommender system.
- Enhancements that can be made to optimize the system for better performance.

## **Is your idea novel?**

There are various different approaches that have been proposed in this domain to recommend different items for different categories. We will strive to work hard in this domain to provide better results and implementation than the existing systems. This work will present a promising starting point towards designing an accurate and reliable book recommendation system.

