First step of the project

Business understanding

Major music streaming apps offer users a high level of recommendation algorithm. But I think the book market has a weak recommendation system.

There are so many books in the world, that the book which suits my taste may not be one of the brand-new best sellers on display.

So I will make an algorithm that can recommend books to individuals based on reviews of Amazon books. I think it's a low-cost marketing, but high-efficiency way.

If it is useful to customers, they will buy more books and seller's profit will be increased. Their sales will be business success criteria.

The amount of data is large enough. Rather, It'll be important to make the program operate neatly because of the large amount.

Among the reviews, the program has to find the ones that match the user's preference. If they are satisfied with the book of these reviews, data-mining is well done.

Data understanding

I will get Amazon books reviews data from Kaggle(URL: https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews)

There are two csv files(Books_ranking.csv, books_data.csv). They contain the books' title, authors, category, description, price, image, publisher, published date, review(score, text) and users profile who wrote them.

The main focus is the text review, but other feathers also should be used properly to make the algorithm. From the text reviews the program has to exclude useless words(I, you, this, book, interesting …) from the text reviews and extract helpful words(touching, comic, romantic …) to analyse associations between them and users' preference.

Because anyone can write these reviews, the amount is large, but it may not be good quality. So, it is important to improve the quality of them. For example, the

algorithm can only use the reviews that are more than a certain length and/or with good ratings.

Planning the project.

1. Receive user input(organise the GUI) and transform to data.
2. Build an algorithm to separate helpful words and useless words from reviews.
3. Build an algorithm which recommends books based on extracted words.
4. Refine the algorithms(step 2, 3) checking results.
5. Devise how users can access the program.
   :If it works on the web, it’s easy for users to use it rather than downloadable program. But I don’t know now that I can do it alone. And providing program by web is not related to this course(Data Science), so I don’t think it is a meaningful part of this project. I’ll decide it after making algorithms.

The program will use methods we learn during the course. The most important part is to find the best way to get nice recommendation system.