

# 2023 [Machine Learning Project] Personalized Book Search

## Problem

You're driving the team to develop an innovative book search engine. Assume that you have an inclusive catalogue containing information about 10k English books. Over time you've also collected users' interactions with books including `marked` and `ratings`.

Your goal is to provide users a personalized search experience for any type of book-hunting intent query. More formally, your goal is to provide code that takes as input a user id  $u$  and a query string  $q$ , and gives as output a ranked list of books  $b_1, b_2, \dots, b_n$ . We expect your solution to have an information retrieval component.

The data is downloadable [here](#). This link also contains a description of the dataset.

## Deliverables

- All the code you wrote to solve the problem, including the data exploration, model, and feature generation.
- You should explain and provide all data and code necessary to evaluate your system.
- What we are looking for here is the quality of the personalization output with your choices of the evaluation metric. Your code only needs to return the list of books for a query in an easily checkable form **but you don't need to implement any UI to render the search results**.
- Instructions on how to test your code with a query.
- A short document answering the following questions:
  - What is your problem formulation?
  - What are the goals and non-goals of your model and system?
  - What features did you consider?
  - What model did you use and why?
  - What was your evaluation method for this? Why did you select this method? Please include details on how you evaluated your model.
  - What features would you like to add to the model in the future if you had more time?

- What other things would you want to try before deploying this model in production.

## Hints

- If you are reading this problem carefully, you'll realize that this problem is not solely a recommendation system.
- It's your choice to decide which part of the search experience that your algorithm can apply and what this new experience will look like. And yes, you don't need to use all the data/resources that are provided to you.
- Feel free to use the [existing notebook](#) from the repository if that's useful to you.

Please **do not** post this question or your solution publicly on the internet. Feel free to set your own time constraints and scope down as needed and (please do not ruin your weekend for this!). We sincerely appreciate the fact that you're willing to spend any time whatsoever on this. Thank you and have fun!

**neeva**