Main Exercise - Data Engineer

Problem: Productionalization of book recommendation script (result from the INTRO BONUS task)

Our junior data scientist wrote a proof of concept algorithm for book recommendation (it is not a clean code but you should be able to understand what is going on in the script) and your goal is to **take the script and productionalize it**:

That means some of those:

- create some reasonably detailed scheme of an app architecture - which parts should the application have, how they should be connected etc.

- do a code review to solution from junior data scientist, discuss its pros & cons

- write a code to download the data and improve its quality before it is used by the recommendation engine

- build lightweight prototype of your application, where anyone can enter his/her favourite book and the application returns a sorted list of recommended books - using the model output

- voluntarily build a simple frontend for your application (web page)

- present us with your solution and discuss it o how you think about the solution o your ability to explain your approach   o limitations of the approach, possible future improvements

Philosophy:

- the actual result and code are not that important – the journey there and potential future paths are crucial

- if you are able to follow-through with some ideas, great!

- if you just start something and have a clear idea on how to proceed, that is also useful

- you are expected to spend up to 2 evenings on the task (but there are no bounds to proactivity if you enjoy playing with it)

Data & python scripts: https://datasentics.sharepoint.com/:u:/s/EXTDataEngineerTask/
EbeD2T97tahCitR2aWOh__8BpCy1DJX8lrBOPBgF6VZulQ?e=N2c06o

available open dataset: https://www.kaggle.com/datasets/arashnic/book-recommendation-
dataset?select=Ratings.csv

Tools:

- use whatever you are comfortable in (Python, Java, Javascript, .Net, …) or feel free to use it as an opportunity to try out a new language (but be careful with that) - how to present: up to you – slides, walking through code, drawing, ...

INTRO BONUS Exercise - Data Scientist (mostly for context)

Problem: Book recommendations - „I like Lord of the Rings, what else should I read?“

Goal:

- take some data, try some approaches, produce some code, get some results
- then come and show us your solution and have a chat around it - show how you think about a specific problem, how you are able to explain what approach you used and why, think about the limitations of the approach and how things could be improved if there was more time, what you think of the results and if they make sense, etc.

Philosophy:

- the actual result and code are not that important – the journey there and potential future paths are more important
- if you are able to follow-through with some ideas then great, if you just start something and have a clear idea on how to proceed that is also useful
- the expectation is that you will spend an evening or two with the task (but there are no bounds to proactivity if you enjoy playing with the problem)

Data:

- available open dataset: https://www.kaggle.com/datasets/arashnic/book-
recommendation- dataset?select=Ratings.csv
- alternatively, feel free to grab any other relevant data set

Tools:

- use whatever you are comfortable in (R, Python, Matlab, Java, SQL,…) or feel free to use it as an opportunity to try out a new language

- it is not a contest in finding the best black-box library and blindly using it – own solutions are preferred even if they are simple

- how to present: up to you – slides, walking through code, drawing, …