Fedshi Machine Learning Assignment

Building a Real-time personalised recommender system

Your task is to develop a recommendation system for an e-commerce platform. The recommendations should be based on the similarity of the audiences. You can use this dataset:

- Books
- Users
- Rankings

Your primary goal is to design a system that can effectively learn representations for both users and items, enabling highly relevant recommendations based on their similarity. We are particularly interested in seeing your approach to transforming these diverse data sources and using them to solve the problem. This Assignment is designed for 3-5 days but if you need more time to complete it based on your standards we might extend it.

Please provide the following deliverables:

1. **Jupyter Notebook(s):**
    - **Data Preprocessing and Feature Engineering:** Demonstrate how you would clean, preprocess, and engineer features from the raw user interaction data and item metadata.
    - **Model Architecture and Training:** In a clear language communicate the architecture of the system you have designed this should ideally include the data pipelines, feature pipelines, training and inference architecture. And you considerations for serving the models.
    - Describe the loss function you would use to train this model.
    - Your assignment should have a read.me with clear instructions for the evaluator to deploy the model, run it, and see the results.
    - In a separate section discuss all of the metrics you have identified and are tracking to measure the efficiency of your model. Explain why you have decided to measure these metrics.
2. **Dashboard**
    - Develop a **user-friendly interactive Streamlit, gradio (or any other preferred dashboard)** to showcase your recommendation model.
    - **Input Section:** Allow a user to select a `User_ID` (from a predefined list or by input) to get personalized recommendations. Alternatively, allow selecting an `Item_ID` to find similar items.

- **Output Section:** Display the top-N recommended items for the selected user, or the top-N similar items for the selected item. Include relevant item details (e.g., name, category, image if available).

3. **PowerPoint (PPT) Presentation:**

    - **Project Overview:** Briefly introduce the problem and the goal of the recommendation system. Your audience are from the growth, product, and commercial teams.
    - **Data Understanding:** Summarize the dataset used, its key entities (users, items, interactions), and relevant features.
    - Clearly explain the **conceptual design of your model**.
    - Explain how **similarity** is defined and calculated to generate recommendations.
    - Explain how would you approach solving the **Cold Start** problem for new items
    - Articulate the **advantages** of this architecture for scaling to large datasets and capturing complex relationships.
    - **Key Findings & Evaluation:** Present the main results from your model training and evaluation. Suggest recommendations, if any, for your target audience. List the technical metrics you have identified and their impact on the business.

4. **Code, Data, and Model Versioning:**

- Your code should be in a Git repository.
- While it is not expected from you to use a complete DVC but briefly explain what would be your approach for versioning data as well as features.
- Also briefly discuss your Model versioning approach, optionally, use tools like Apache Airflow to illustrate your approach.

If you have used AI for delivering any part of the assignment please share your workflow, prompts, and thought process.