**Reference**: AI Apprenticeship Framework ideas

# Recommendation Engine

## 1. Goals and Requirements

We have a dataset of sales from an online store. The dataset contains transactional data, product-related data, and user-related data. There are a lot of products that are hard for the user to navigate. When a user visits the landing page, we want the user to be shown a list of the top 10 products that align with the user's interest. Also, whenever a user adds one item to the cart, we want the user to be shown a list of the top 10 products that complement the item already in the cart.

- Develop a landing page recommendation system using transactional, user, and product data to generate a list of the top 10 list of items
- Develop an add-to-cart recommendation system using sequential purchase data to generate a list of the top 10 list of items

## 2. Milestones

1. Exploratory Data Analysis on the data to generate insights on viable features that could be used in the model and general data preprocessing
2. Development of a baseline model using Association Rule Mining for add-to-cart recommendation
3. Data preprocessing and data conversion to the format required by the landing page recommendation model
4. Development of landing page recommendation model using DeepCTR and experiment tracking using MLFlow
5. Deployment of the model in REST API through FastAPI and containerization using Docker
6. Data preprocessing and data conversion to the format required by add-to-cart recommendation model
7. Development of add-to-cart recommendation model using RecBole and monitoring using MLFlow
8. Deployment of the model in REST API through FastAPI and containerization using Docker

# 3. Success Criteria

- Evaluation metrics of models should be above acceptable threshold (varies on data)
- Recommendations given out by models should be qualitatively good
- A pipeline should be created that takes in raw data and gets up to deployment

# 4. Skills Involved

1. Data Wrangling and Preprocessing:
    - Cleaning, Transformation, Imputation, and Sampling of Data.
    - Text Preprocessing, Tokenization, Stemming, and Lemmatization for NLP tasks.
    - Stop words and Punctuation Removal for text data.
    - Encoding techniques like Bag-of-words, Tf-IDF, and various embeddings (self-learned, pre-trained, W2V, sBERT).
2. Machine Learning and Recommender Systems:
    - Understanding the basics of Machine Learning and its application to recommender systems.
    - Knowledge of different types of recommendations: Non-Personalization, Personalization, Implicit vs. Explicit Feedback, Similarity Scores, Content-Based, Collaborative, Matrix Factorization, Hybrid, etc
3. Deep Learning for Recommendations:
    - Understanding and using the deepCTR library for click-through-rate (CTR) prediction.
    - Data preparation in deeper format.
    - Model training using deepCTR models for recommendation systems.
4. Learning-to-Rank (LTR) Approach:
    - Knowledge of Learning-to-Rank algorithms such as Pointwise, Pairwise, and Listwise methods.
    - Candidate Generation and Ranking techniques for LTR-based recommendation systems.
5. Evaluation Metrics
    - Understanding and using evaluation metrics to assess the performance of recommendation systems.
6. MLflow Integration
    - Integrating the MLflow framework to track and manage machine learning experiments.
7. Web API and Deployment
    - Knowledge of FastAPI for building web APIs to deploy recommendation systems.
    - Familiarity with Docker for containerizing the deployed APIs.

8. Version Control
    - Understanding and using Git for version control during the development of the recommendation system.

# 4. Dataset

Instacart Market Basket Analysis

It consists of more than 3 million grocery order transactions with over more than 200000 users.

# 5. Resources, Courses, Upskilling Requirements

Link

# 6. Tools and Libraries

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Surprise
- PyTorch
- SBERT
- Tensorflow
- DeepCTR
- RecBole
- LightGBM

# 7. Project Process

This project is led by our principal researcher, Pradip Mainali. All tasks are directly assigned to us by him and any progress and problems that arise are discussed with him. The project process in as follows:

- Get assigned a task
- Complete a task
- Create a detailed report/presentation on the task along with the problems you faced

- Get recommendation on what to do next from Pradip dai

**Tasks**

- Data wrangling, cleaning
- EDA
- Feature Engineering
- Modeling
- Model Evaluation
- Documentation and reporting

**Git workflow**

- Work on a new branch
- Pull Request
- Code Review
- Merge

# 8. Timeline

- Present a detailed timeline with important dates and milestones.
- Include regular checkpoints for progress evaluation.

# 9. Onboarding Plan

- A presentation on the task we have done and are working on, the overview of the project, and how they will be working on it
- Upskilling
- Sharing the Documentation

# 10. History and Latest Progress

- Project Summary and Accomplishments
- Iteration on Modeling for new client data.

# 11. Trainee Evaluation Plan

Trainees will be evaluated on the basis of their upskilling scores.

**Scratch Pad:**

Type - AI engine, other projects

Foundation Course

Goals, Requirements

Teach trainees about time series

Milestones

Success Criteria

Trainees complete all the chapters and gain certifications
They do a demo project…
...

Skills involved

EDA,
Time Series Analysis,
Linear Regression,
…

Dataset

Link to the dataset,
Short description

Code Repo

Link to code repo

Upskilling Resources/ Courses

Tools and Libraries

Pandas, numpy, statsmodel, GluonTS,…

JIRA Kanban board

Link to Kanban Board
NA

Process

*What is the Project Process Methodology for the project?*
*Possible Attributes:*

- *Solo or team Project?*
- *Agile process?*
    - *Sprint Duration?*
    - *Have Daily Standups?*
    - *Sprint Planning and Retrospectives?*
    - *Kanban or Agile?*
    - *Stages in JIRA Board?*

Solo
Sprint Duration 1 month
Code Review

Timeline

History / Latest Progress

Onboarding plan

*What tasks to assign for onboarding trainees?*

Trainee Performance Evaluation