

Description

Context

Today, information is growing exponentially with volume, velocity and variety throughout the globe. This has lead to information overload, and **too many choices for the consumer of any business**. It represents a real dilemma for these consumers and they often turn to denial. Recommender Systems are one of the best tools that help recommending products to consumers while they are browsing online. Providing **personalized recommendations** which is most relevant for the user is what's most likely to keep them engaged and help business.

E-commerce websites like Amazon, Walmart, Target and Etsy use different recommendation models to provide personalized suggestions to different users. These companies spend millions of dollars to come up with algorithmic techniques that can provide personalised recommendations to their users.

Amazon, for example, is well-known for its accurate selection of recommendations in its online site. Amazon's recommendation system is capable of intelligently analysing and predicting customers' shopping preferences in order to offer them a list of recommended products. **Amazon's recommendation algorithm is therefore a key element in using AI to improve the personalization of its website**. For example, one of the baseline recommendation models that Amazon uses is item-to-item collaborative filtering, which scales to massive data sets and produces high-quality recommendations in real-time.

Objective

You are a Data Science Manager at Amazon, and have been given the task of building a recommendation system to recommend products to customers based on their previous ratings for other products. You have a collection of labelled data of Amazon reviews of products. The goal is to extract meaningful insights from the data and build a recommendation system that helps in recommending products to online consumers.

Data Description

The Amazon dataset contains the following attributes:

- **userId:** Every user identified with a unique id
- **productId:** Every product identified with a unique id
- **Rating:** The rating of the corresponding product by the corresponding user
- **timestamp:** Time of the rating. We will not use this column to solve the current problem

Submission Guidelines

1. There are two ways to work on this project:

i. Full-code way: The full code way is to write the solution code from scratch and only submit a final Jupyter notebook with all the insights and observations.

ii. Low-code way. The low-code way is to use an existing solution notebook template to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their individual learning aspirations and outcomes. The below table elaborates on these two options.

Submission type	Who should choose	What is the same across the two	What is different across the two	Final submission file [IMP]	Submission Format
Full-code	Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch	Perform exploratory data analysis to identify insights and recommendations for the problem	Focus on code writing: 10 - 20% grading on the quality of the final code submitted	Solution notebook from the full-code template submitted in .html format	.html
Low-code	Learners who aspire to be in managerial roles in the future-focussed on solution review, interpretation, recommendations, and communicating with business		Focus on business presentation: 10 - 20% grading on the quality of the final business presentation submitted	Business presentation in .pdf format with problem definition, insights, and recommendations	.pdf

Please follow the below steps to complete the assessment. Kindly note that if you submit a presentation, ONLY the presentation will be evaluated. Please make sure that all the sections mentioned in the rubric have been covered in your submission.

i. Full-code version

- Download the full-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Clearly write down insights and recommendations for the business problems in the comments.
- Submit only the solution notebook prepared from the learner notebook [format: .html]

ii. Low-code version

- Download the low-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Prepare a business presentation with insights and recommendations to the business problem.
- Submit only the presentation [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated if

- it is submitted post-deadline, or,
- more than 1 file is submitted.

Best Practices for Full-code submissions

- The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
- The notebook should be run from start to finish in a sequential manner before submission.
- It is important to remove all warnings and errors before submission.
- The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
- Please refer to the FAQ page for common project-related queries.

Best Practices for Low-code submissions

- The presentation should be made keeping in mind that the audience will be the Data Science lead of a company.
- The key points in the presentation should be the following:
 - Business Overview of the problem and solution approach
 - Key findings and insights which can drive business decisions
 - Business recommendations
 - Focus on explaining the key takeaways in an easy-to-understand manner.
 - The inclusion of the potential benefits of implementing the solution will give you the edge.
- Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
- The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
- Please refer to the FAQ page for common project-related queries.

Happy Learning!

Scoring guide (Rubric) - Recommendation Systems (1)

Criteria	Points
Exploratory Data Analysis - Check the number of rows and columns and provide observations - Check data types and provide observations - Check for missing value in the data and provide observations - Summary statistics of 'rating' variable and provide observations - Create the bar plot to check the 'rating' distribution and provide observations - Check the number of unique USERS and PRODUCTS in the data and provide observations	3
Model Building - Rank based Recommendation System - Recommend top 5 products with 50 minimum interactions based on popularity - Recommend top 5 products with 100 minimum interactions based on popularity	3
Model Building - User-User Similarity-based Recommendation System - Initialize a baseline user-user similarity based recommendation system - Fit the model on the training data - Use the precision_recall_at_k function to calculate the metrics on the test data	3
Improving Model Performance - User-User Similarity-based (Optimized) Recommendation System - Perform hyperparameter tuning for the user-user similarity-based model - Build the optimized model by using tuned values of the hyperparameters - Predict the rating for a user using the optimized model	5

Criteria	Points
Model Building - Item-Item Similarity-based Recommendation System - Initialize a baseline item-item similarity based recommendation system - Fit the model on the training data - Use the precision_recall_at_k function to calculate the metrics on the test data	3
Improving Model Performance - Item-Item Similarity-based (Optimized) Recommendation System - Perform hyperparameter tuning for the item-item similarity-based model - Build the optimized model by using tuned values of the hyperparameters - Predict the rating for a user using the optimized model	5
Model Building - Matrix Factorization based Recommendation System - Initialize a baseline Matrix Factorization based recommendation system - Fit the model on the training data - Use the precision_recall_at_k function to calculate the metrics on the test data	3
Improving Model Performance - Matrix Factorisation based (Optimized) Recommendation System - Perform hyperparameter tuning for the Matrix Factorization model - Build the optimized model by using tuned values of the hyperparameters - Predict the rating for a user using the optimized model	5
Conclusion & Recommendations - Comparing different models' performance and choosing the final model with reasoning - Provide recommendations that can be acted upon to improve the business outcome	5
Report - Overall quality/ Notebook - Overall Low Code (Business Report): - Structure and flow - Crispness - Visual appeal Full Code (Notebook in HTML format): - Structure and flow - Well commented code	5
	Points 40