Artificial Intelligence – Week 5

Instructor: Teerapong Panboonyuen

Course Repository:
https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025s1

🎯 Objective

This week, you will explore the fundamentals of **Collaborative Filtering** to build a recommender system using real-world retail transaction data.

You will:

- Understand user-item interactions and rating behavior

- Engineer a collaborative filtering pipeline using Matrix Factorization (SVD)

- Discover insights from user purchase data

- Evaluate your recommendation model using a metric like **RMSE**

- Build a system that can suggest relevant items to users based on their behavior

📚 Dataset Description

We will use the **Online Retail dataset** provided in the course GitHub repo:

📂 **Dataset URL:**
KKUAI_OnlineRetail.csv

This dataset contains:

- Invoices, product codes, descriptions
- Customer purchase quantities and unit prices
- Date of transaction and customer ID

Key Columns:

- InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country

---

📝 Assignment Instructions

1. Dataset Download & Preprocessing

- Load the dataset using pandas

- Remove missing values and invalid transactions

- Compute derived fields like **TotalPrice** = Quantity × UnitPrice

2. Insight Discovery

Extract at least **5 insights** using pandas, plots, and groupby operations. Examples:

- What are the most popular items?

- Which countries have the most customers?

- When are customers most active?

- Which products generate the most revenue?

- How frequently do customers return?

3. Build a Collaborative Filtering Model

Using Surprise, build a recommendation system using **Matrix Factorization (SVD)**:

- Construct a user-item rating matrix from quantity or frequency

- Train an SVD model

- Use cross_validate to report **RMSE** and **MAE**

4. Evaluation

Evaluate your model on:

- **RMSE** (Root Mean Squared Error)

- **MAE** (Mean Absolute Error)

- Discuss model accuracy and potential limitations

5. Bonus (Optional)

- Visualize the user-item matrix using a heatmap

- Try item-based or user-based k-NN Collaborative Filtering

- Apply dimensionality reduction (e.g., PCA) to visualize latent factors

---

🔥 **Example Approaches**

- Use SVD from `Surprise` to learn latent features

- Use collaborative filtering even without explicit ratings — treat quantities as implicit preferences

- Cross-validate using folds and compare RMSE between different algorithms

---

💡 **Deliverables**

📔 Jupyter Notebook / Google Colab:

- Code to load and clean data

- 5+ insights with visualizations

- Collaborative filtering implementation

- Model evaluation using RMSE or MAE

- Summary and interpretation of results

📄 Documentation:

- Summary of each step

- Reflections on insights from the data

- Final notes on how you could improve the model in the future

---

🎯 **Evaluation Criteria**

Criteria	Weight
Data Cleaning & Preprocessing	✅ 15%
Insight Discovery (5+ Questions)	✅ 25%
Collaborative Filtering Implementation	✅ 25%
Evaluation with Metrics	✅ 20%
Clarity, Code Comments, and Presentation	✅ 15%

🚀 **Getting Started**

💻 Setup in Colab:

```
!wget
https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025
s1/raw/main/dataset/KKUAI_OnlineRetail.csv
```

📦 Recommended Libraries:

python

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from surprise import Dataset, Reader, SVD
from surprise.model_selection import cross_validate
```

📅 **Submission Deadline:**

🕐 The final deadline will be announced and agreed upon in class. Submit your `.ipynb` notebook to the LMS or via GitHub Classroom as instructed.

Good luck and have fun building your first real recommender system!