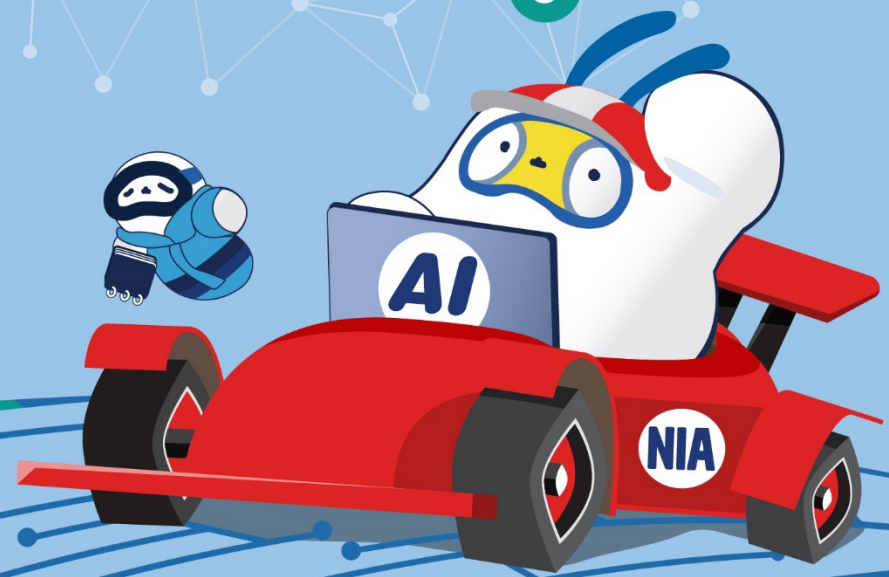


DATA CREATOR CAMP

2024 데이터 크리에이터 캠프

대학부 실습영상

8강. Item-based collaborative Filtering 실습



과학기술정보통신부


NIA 지능정보원
한국지능정보사회진흥원

목차

- 사전 학습 실습 데이터 소개
- Item-based collaborative filtering

1. 사전 강의 실습 데이터


- 해당 실습 데이터는 과제 데이터와 무관합니다.


 FOXTROT · UPDATED 7 YEARS AGO

▲ 420

New Notebook


Download (12 MB)





goodbooks-10k

Ten thousand books, one million ratings. Also books marked to read, and tags.



[Data Card](#) [Code \(96\)](#) [Discussion \(10\)](#) [Suggestions \(0\)](#)

About Dataset

This version of the dataset is obsolete. It contains duplicate ratings (same user_id,book_id), as reported by Philipp Spachtholz in his illustrious notebook.

The current version has duplicates removed, and more ratings (six million), sorted by time. Book and user IDs are the same.

**It is available at <https://github.com/zygmuntz/goodbooks-10k>. **

Usability ⓘ
8.24

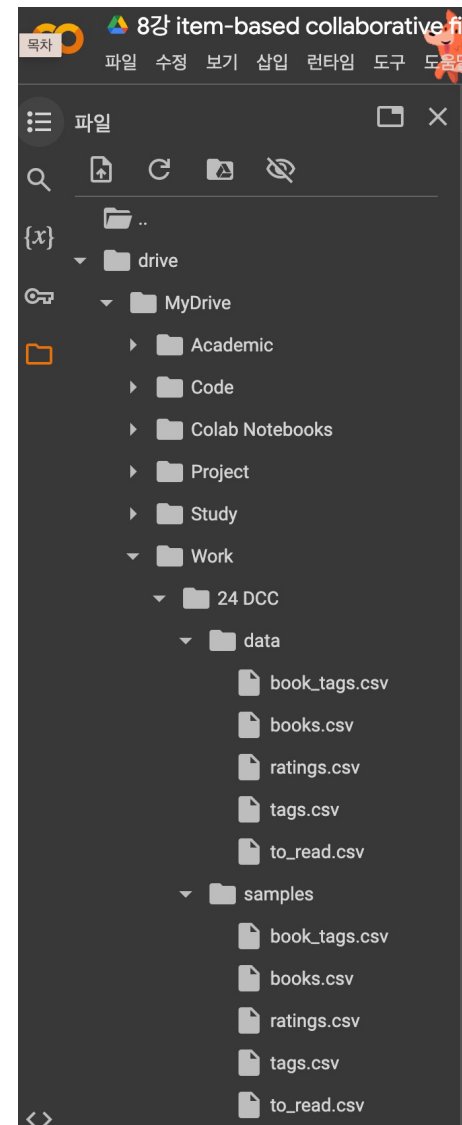
License
[CC BY-SA 4.0](#)

Expected update frequency
Not specified

Tags

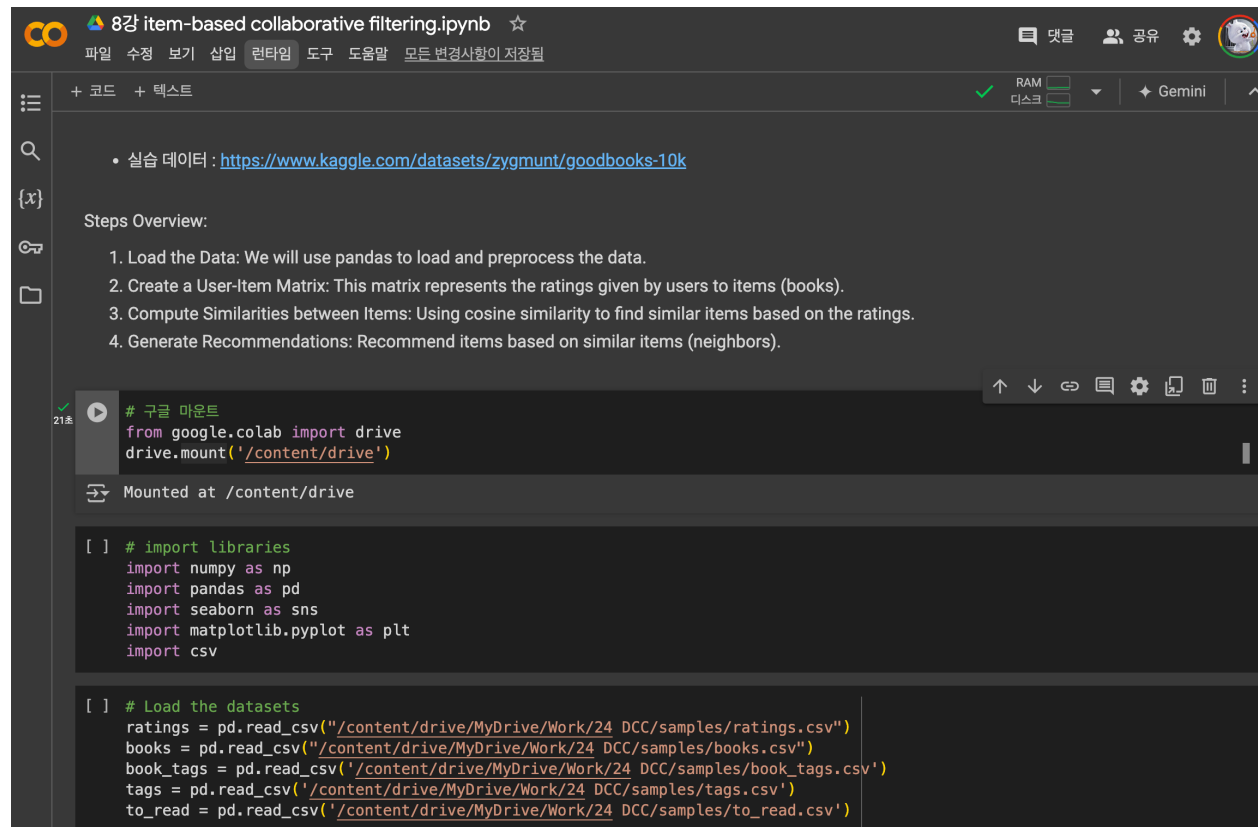
1. 사전 강의 실습 데이터

- 데이터 폴더를 마운트 한 후, 실습 준비를 마칩니다.



2. item-based collaborative filtering

- 실습 링크 : <https://colab.research.google.com/drive/1zY7xhabQUwcxKajfkA-fP5ta-uWz5Vne?usp=sharing>



```
8강 item-based collaborative filtering.ipynb ☆
파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트
✓ RAM
디스크
Gemini

• 실습 데이터 : https://www.kaggle.com/datasets/zygmunt/goodbooks-10k

Steps Overview:

1. Load the Data: We will use pandas to load and preprocess the data.
2. Create a User-Item Matrix: This matrix represents the ratings given by users to items (books).
3. Compute Similarities between Items: Using cosine similarity to find similar items based on the ratings.
4. Generate Recommendations: Recommend items based on similar items (neighbors).

# 구글 마운트
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[ ] # import libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import csv

[ ] # Load the datasets
ratings = pd.read_csv("/content/drive/MyDrive/Work/24 DCC/samples/ratings.csv")
books = pd.read_csv("/content/drive/MyDrive/Work/24 DCC/samples/books.csv")
book_tags = pd.read_csv("/content/drive/MyDrive/Work/24 DCC/samples/book_tags.csv")
tags = pd.read_csv("/content/drive/MyDrive/Work/24 DCC/samples/tags.csv")
to_read = pd.read_csv("/content/drive/MyDrive/Work/24 DCC/samples/to_read.csv")
```



이 문서의
외부 유출 및
공유를 금합니다.

본 콘텐츠는 한국지능정보사회진흥원(NIA)의 동의 없이 무단 사용할 수 없으며,
상업적 목적으로 이용을 금합니다.

DATA CREATOR CAMP

2024 데이터 크리에이터 캠프

감사합니다.



과학기술정보통신부

NIA

지능정보원
한국지능정보사회진흥원