

1. 조건에 따른 모델 성능 비교

번호	조건	AUC	학습 시간
1	no_components=60 learning_rate=0.01 user_alpha=1e-4 item_alpha=1e-4 epochs=100	<pre> --- 5. 최종 평가 결과 --- Train AUC: 0.9091 Test AUC: 0.7599 ----- </pre>	소요 시간: 343.82 초
2	no_components=60 learning_rate=0.01 user_alpha=1e-3 item_alpha=1e-3 epochs=30	<pre> --- 5. 최종 평가 결과 --- Train AUC: 0.8484 Test AUC: 0.7698 ----- </pre>	소요 시간: 97.85 초
3	no_components=30 learning_rate=0.01 user_alpha=1e-3 item_alpha=1e-3 epochs=20	<pre> --- 5. 최종 평가 결과 --- Train AUC: 0.8167 Test AUC: 0.7649 ----- </pre>	소요 시간: 44.52 초
4	no_components=40 learning_rate=0.01 user_alpha=1e-3 item_alpha=1e-3 epochs=30	<pre> --- 5. 최종 평가 결과 --- Train AUC: 0.8352 Test AUC: 0.7764 ----- </pre>	소요 시간: 76.74 초
5	no_components=30 learning_rate=0.01 user_alpha=1e-3 item_alpha=1e-3 epochs=30	<pre> --- 5. 최종 평가 결과 --- Train AUC: 0.8286 Test AUC: 0.7749 ----- </pre>	소요 시간: 66.70 초
6	no_components=20 learning_rate=0.01 user_alpha=1e-3 item_alpha=1e-3 epochs=30	<pre> --- 5. 최종 평가 결과 --- Train AUC: 0.8135 Test AUC: 0.7688 ----- </pre>	소요 시간: 52.10 초

num_threads=4 로 병렬 처리 중일 때, 각 CPU thread 가 데이터를 처리하는 순서는 매번 달라질 수 있다.

-> 비결정적인 순서로 인해 최종 결과에 차이가 발생할 수 있다

변경 조건:

no_components: 잠재 요인 수 - LightFM 모델이 사용자(User)와 아이템(Item)의 특성을 표현하는 잠재 벡터(Latent Vector)의 차원 수를 결정

learning_rate: 학습 진도 - 오류를 수정하는 보폭

user_alpha: 사용자 L2 정규화 강도 - 사용자 특징(Feature) 및 잠재 벡터에 대한 L2 페널티

item_alpha: 아이템 L2 정규화 강도 - 아이템 특징(Feature) 및 잠재 벡터에 대한 L2 페널티

epochs: 학습 반복 횟수 - 모델이 전체 학습 데이터를 몇 번 반복하여 학습할지 결정하는 횟수

-> overfitting 위험성이 가장 적고, Test AUC 성능이 가장 좋은 4 번 모델 사용

2. 활동량 기반 필터링에 따른 모델 성능 비교

해당 모델을 사용하여 활동량을 기준으로 노이즈 제거(유저: 책 추천 5 회 미만, 책: 언급 3 회 미만) 후 성능을 비교]

```
--- 데이터 필터링 시작 (User 최소 5회, Item 최소 3회) ---
-> 유저 제거 후 상호작용: 925554개
-> 아이템 제거 후 최종 상호작용: 693157개
--- 필터링 완료: 1031136개에서 693157개로 감소 (32.78% 감소) ---
```

기존 AUC	노이즈 제거 후 AUC
--- 5. 최종 평가 결과 --- Train AUC: 0.8352 Test AUC: 0.7764 -----	--- 5. 최종 평가 결과 --- Train AUC: 0.7716 Test AUC: 0.7250 -----

노이즈 제거 후 오히려 AUC 감소:

노이즈 제거 조건 완화(유저: 책 추천 3 회 미만, 책: 언급 1 회 미만) 후 재시도

```
--- 데이터 필터링 시작 (User 최소 3회, Item 최소 1회) ---
-> 유저 제거 후 상호작용: 958242개
-> 아이템 제거 후 최종 상호작용: 958242개
--- 필터링 완료: 1031136개에서 958242개로 감소 (7.07% 감소) ---
```

기존 AUC	노이즈(기준 완화) 제거 후 AUC
--- 5. 최종 평가 결과 --- Train AUC: 0.8352 Test AUC: 0.7764 -----	--- 5. 최종 평가 결과 --- Train AUC: 0.8306 Test AUC: 0.7635 -----

결론: 활동량을 기반으로 하는 노이즈 제거는 성능 향상에 큰 도움이 되지 않음

-> 기존의 4 번 모델로 프로젝트 진행

3. 기존 사용자 도서 추천

Book Recommendation System

Enter the user ID (Exit: q):276726

Top 5 books recommended to user ID: 276726

1. Rocks, Gems and Minerals : A Golden Guide from St. Martin's Press (Golden Guide from St. Martin's Press) (By: Paul R. Shaffer)
ISBN: 1582381321, Predicted score: -90.6251
2. Rocks and Minerals : A Golden Photo Guide from St. Martin's Press (Golden Photo Guide from St. Martin's Press) (By: Cally Oldershaw)
ISBN: 0312289219, Predicted score: -90.8066
3. Seashells of the World : A Golden Guide from St. Martin's Press (Golden Guide) (By: R. Tucker Abbott)
ISBN: 1582381488, Predicted score: -91.4840
4. North American Indian Arts : A Golden Guide from St. Martin's Press (Golden Guide) (By: Andrew Hunter Whiteford)
ISBN: 1582381453, Predicted score: -91.8824
5. Microscopic Life: A Golden Photo Guide from St. Martin's Press (By: Theresa Greenaway)
ISBN: 0312300557, Predicted score: -92.3038

Enter the user ID (Exit: q):11627

Top 5 books recommended to user ID: 11627

1. Rocks, Gems and Minerals : A Golden Guide from St. Martin's Press (Golden Guide from St. Martin's Press) (By: Paul R. Shaffer)
ISBN: 1582381321, Predicted score: -90.6251
2. Rocks and Minerals : A Golden Photo Guide from St. Martin's Press (Golden Photo Guide from St. Martin's Press) (By: Cally Oldershaw)
ISBN: 0312289219, Predicted score: -90.8066
3. Seashells of the World : A Golden Guide from St. Martin's Press (Golden Guide) (By: R. Tucker Abbott)
ISBN: 1582381488, Predicted score: -91.4840
4. North American Indian Arts : A Golden Guide from St. Martin's Press (Golden Guide) (By: Andrew Hunter Whiteford)
ISBN: 1582381453, Predicted score: -91.8824
5. Microscopic Life: A Golden Photo Guide from St. Martin's Press (By: Theresa Greenaway)
ISBN: 0312300557, Predicted score: -92.3038

Predicted score 순으로 상위 5 개의 도서 추천 진행

3. 새로운 사용자 도서 추천

data_preprocessing.ipynb 에서 새로운 유저에 대한 기본 데이터 입력

```
# 신규 고객 유저 데이터
new_users_data = {
    'User-ID': [300000, 300001, 300002],
    'Location': ['usa', 'georgia, usa', 'morrow, georgia, usa'],
    'Age': [30, 30, 30]
}
```

```
new_users_df = pd.DataFrame(new_users_data)
```

```
# 기존 users DataFrame에 신규 고객 데이터를 추가
users = pd.concat([users, new_users_df], ignore_index=True)
```

```
print(f"신규 고객 3명 삽입 완료. 총 유저 수: {len(users)}")
```

✓ 0.0s

신규 고객 3명 삽입 완료. 총 유저 수: 278861

```
Book Recommendation System
Enter the user ID (Exit: q):300000
Top 5 books recommended to user ID: 300000
1. Thorsons Principles of Druidry (Thorsons Principles) (By: Emma Restall Orr)
   ISBN: 0722536747, Predicted score: -47.3834
2. One-Pot Vegetarian Dishes (One-Pot Sunday Suppers Series , No 2) (By: Amy Cottle)
   ISBN: 006017319X, Predicted score: -47.6901
3. Alf: This Place Is a Zoo (Alf Storybooks, Series II) (By: Robert Loren Fleming)
   ISBN: 0026892219, Predicted score: -48.0377
4. Business as Unusual: The Triumph of Anita Roddick (By: Anita Roddick)
   ISBN: 0722539878, Predicted score: -48.3603
5. Sniglets (Snig'lit : Any Word That Doesn't Appear in the Dictionary, But Should) (By: Rich Hall)
   ISBN: 0020125305, Predicted score: -48.9128
```

```
Enter the user ID (Exit: q):300001
Top 5 books recommended to user ID: 300001
1. Thorsons Principles of Druidry (Thorsons Principles) (By: Emma Restall Orr)
   ISBN: 0722536747, Predicted score: -47.3834
2. One-Pot Vegetarian Dishes (One-Pot Sunday Suppers Series , No 2) (By: Amy Cottle)
   ISBN: 006017319X, Predicted score: -47.6901
3. Alf: This Place Is a Zoo (Alf Storybooks, Series II) (By: Robert Loren Fleming)
   ISBN: 0026892219, Predicted score: -48.0377
4. Business as Unusual: The Triumph of Anita Roddick (By: Anita Roddick)
   ISBN: 0722539878, Predicted score: -48.3603
5. Sniglets (Snig'lit : Any Word That Doesn't Appear in the Dictionary, But Should) (By: Rich Hall)
   ISBN: 0020125305, Predicted score: -48.9128
-----
```

Enter the user ID (Exit: q):300002

Top 5 books recommended to user ID: 300002

1. Thorsons Principles of Druidry (Thorsons Principles) (By: Emma Restall Orr)
ISBN: 0722536747, Predicted score: -47.3834
2. One-Pot Vegetarian Dishes (One-Pot Sunday Suppers Series , No 2) (By: Amy Cotler)
ISBN: 006017319X, Predicted score: -47.6901
3. Alf: This Place Is a Zoo (Alf Storybooks, Series II) (By: Robert Loren Fleming)
ISBN: 0026892219, Predicted score: -48.0377
4. Business as Unusual: The Triumph of Anita Roddick (By: Anita Roddick)
ISBN: 0722539878, Predicted score: -48.3603
5. Sniglets (Snig'lit : Any Word That Doesn't Appear in the Dictionary, But Should) (By: Rich Hall)
ISBN: 0020125305, Predicted score: -48.9128