|  |
| --- |
| **[SW 사용 설명서]**  이상형 월드컵을 통한 서울시 자취방 추천 |
|  |



**2022-1 데이터분석 캡스톤 디자인**

**팀F**

**목차**

[1. 파이썬 모듈 3](#_Toc433487631)

[2. 크롤링 3](#_Toc433487635)

[3. 전처리 4](#_Toc433487641)

[4. 이상형 월드컵 5](#_Toc433487642)

[5. 추천알고리즘 6](#_Toc433487643)

6. [메일 코드 7](#_Toc433487644)

[7. 실행 9](#_Toc433487631)

[8. 성능평가 11](#_Toc433487631)

1. **파이썬 모듈**



본 코드는 위와 같은 여러 파이썬 모듈을 사용하고 있다. 따라서 사용자가 코드를 실행할 때 위의 모듈을 사전에 설치하고 불러와서 코드 사용에 문제가 없도록 진행한다.

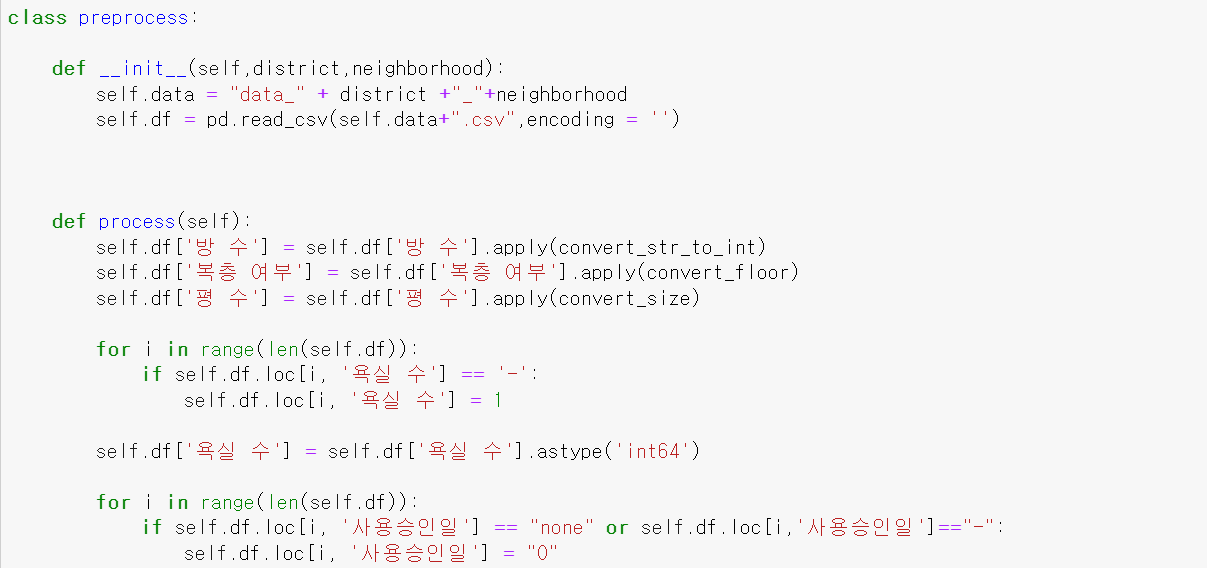
1. **크롤링**

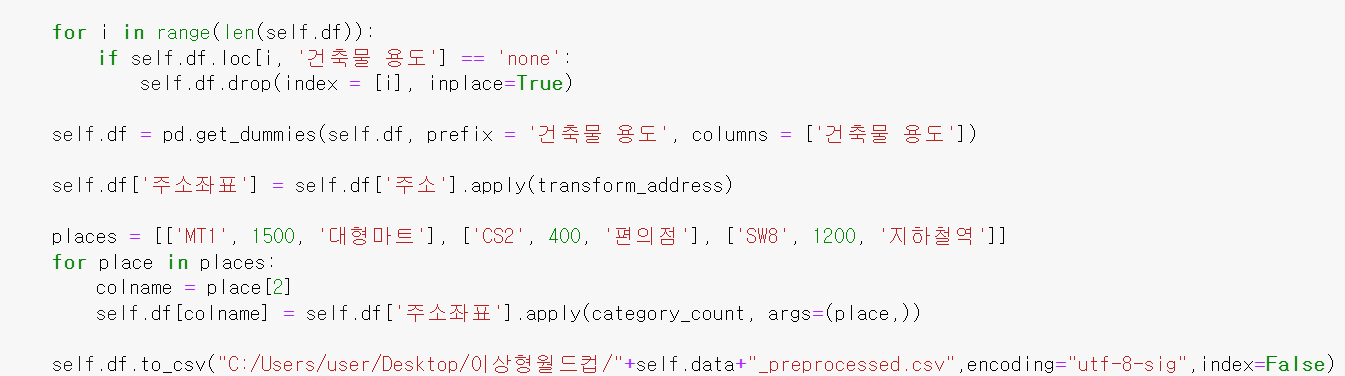


크롤링 코드는 사용자로부터 월세, 전세의 여부에 따른 rent값, 강북구, 강서구, 은평구와 같은 서울시의 구를 입력 받는 district, 우이동, 논현동과 같이 동을 입력 받는 neighborhood가 있어야 진행이 가능하다. 이는 실행 부분에서 이상형월드컵 GUI에서 받아와서 이용하게 된다.

크롤링은 크롬과 크롬의 웹드라이버를 이용하여 파이썬 모듈인 셀레니움을 통해 진행된다. 사용자의 PC에 크롬이 사전에 설치되어야 하며 소스코드 압축파일의 크롬드라이버를 이용한다. 이때 크롬 드라이버와 크롬의 버전이 일치하는 지 확인하고 일치하지 않는다면 본 소스코드 압축파일에 있는 크롬 드라이버는 삭제하고 <https://chromedriver.chromium.org/downloads> 에서 본인 크롬 버전에 맞는 드라이버를 설치한다. 크롤링은 croll class의 crolling 함수를 통해 진행되며 최종적으로 csv파일을 만들어서 저장한다. 이때 만약 한글 깨짐 현상이 나타나면 인코딩을 맞춰서 진행하도록 한다.

1. **전처리**





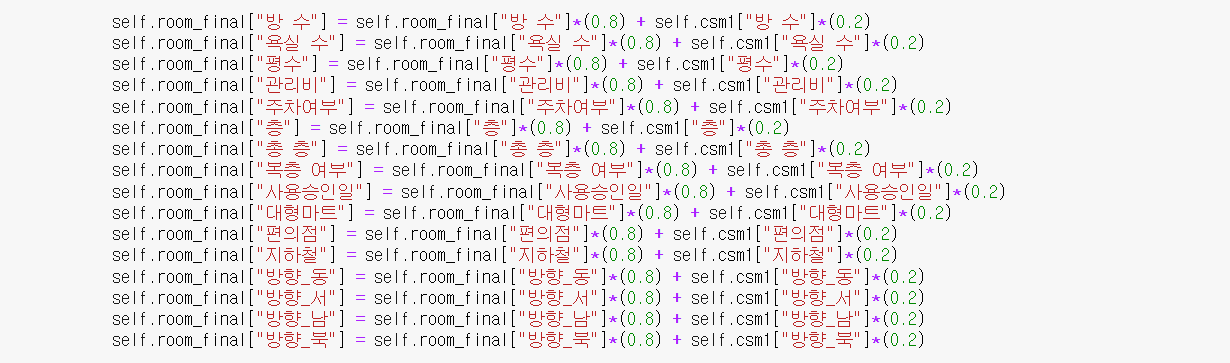
전처리 코드는 preprocess 클래스를 이용하여 진행된다. 최초 prerocess를 선언 시에 입력 받은 구, 동이 자동으로 들어가서 이미 크롤링 때 만들어진 파일을 자동으로 가져오게 된다. 그 후 process함수를 실행하면 여러 column들에 대하여 추천 알고리즘에 적용될 수 있도록 전처리가 이루어진다. 최종적으로 전처리가 완료되면 새로운 csv파일을 만들게 된다. 이때 새로운 csv파일의 이름은 사용자입력 data\_구\_ 동\_preprocessed.csv파일로 만들어진다.

1. **이상형 월드컵**

텍스트이(가) 표시된 사진

자동 생성된 설명

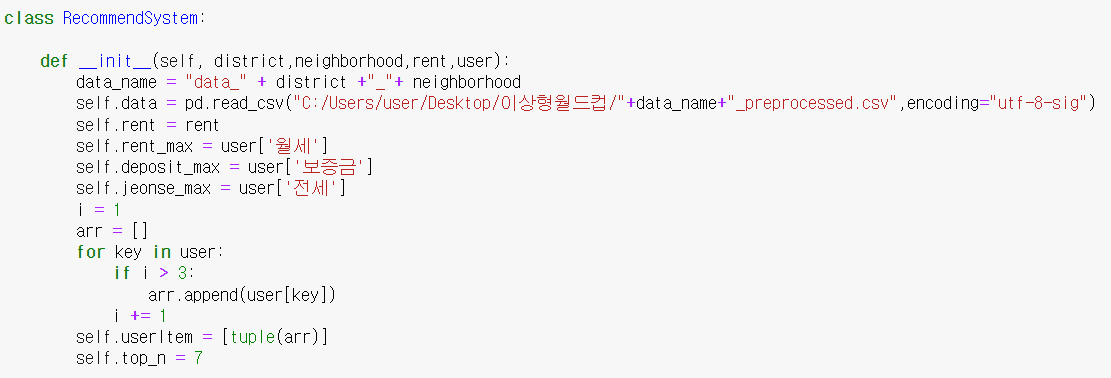
이상형월드컵은 16강 진행을 전제 조건으로 방 조건을 dictionary 형태로 담고 있는 list c1~16, 방 조건을 이상형월드컵 GUI에서 출력할 수 있도록 하는 list rc1~16, 이상형월드컵 진행을 위한 image 정보를 담고 있는 list im1~16을 이용한다. 방이 선택되면 16강에서는 각 list의 16에 방 정보들이 입력되고 진행되면서 8, 4, 2, 1에 선택된 방의 정보들이 입력된다.



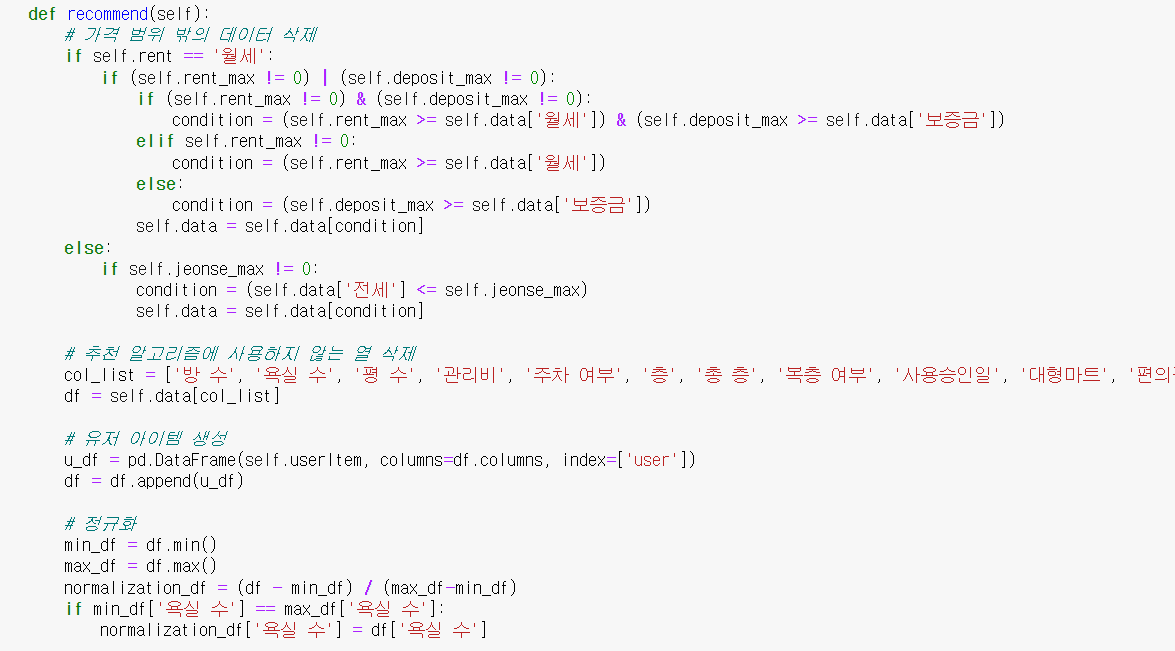
추천 알고리즘 적용을 위해 하나의 가상 매물을 생성해야 하는데 사용자의 선호도 파악을 위해 4강 이상 진출한 매물들을 이용하여 매물이 하나가 아니게 된다. 그래서 가중치를 부여하는데 4강매물에 대해서는 결승에 올라간 매물에 0.7, 탈락한 매물에 0.3을 곱하여 더하는 방식을 이용했고, 우승매물은 0.8, 패자부활전 매물은 0.2를 곱하여 더하는 방식을 이용해 각 컬럼을 합쳐 하나의 가상 매물을 생성하게 된다.

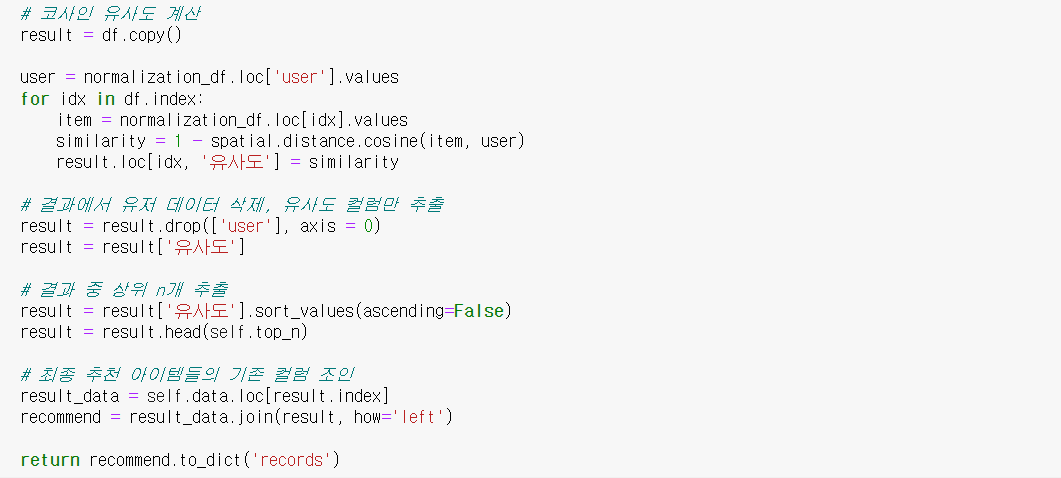
GUI는 tkinter 모듈을 이용해 생성했다.

1. **추천알고리즘**



추천 알고리즘은 RecommendSystem 클래스를 이용하여 진행된다. 최초 RecommendSystem을 선언 시에 구, 동, 월세여부, 월세, 보증금, 전세 값, 사용자 아이템을 input으로 받는다. input으로 받은 구, 동에 해당하는 데이터셋을 가져온다.



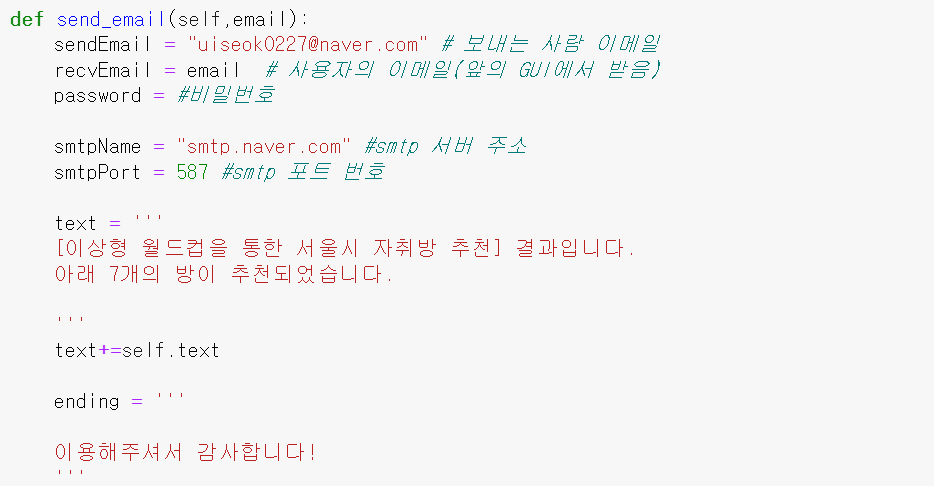


추천 알고리즘은 함수 recommend를 통해 수행된다. recomment 함수는 가격 범위 밖의 데이터 삭제, 추천 알고리즘에 사용하지 않는 열 삭제, 유저 아이템 생성, 정규화, 코사인 유사도 계산, 결과에서 유저 데이터 삭제 및 유사도 컬럼만 추출, 결과 중 상위 n개 추출, 최종 추천 아이템들의 기존 컬럼 조인 순으로 수행된다. 그리고 결과로 데이터를 dictionary 형태로 바꿔 return 한다.

1. **메일**



메일 코드의 경우 mailer 클래스로 구현했으며 추천알고리즘을 통해 나온 딕셔너리 형태의 방을 리스트에 넣은 리스트를 room\_list로, 월세, 전세 여부를 rent로 받게 된다.



이때 send\_mail을 통해 메일을 보내게 되는데 sendEmail 변수에는 서버메일 혹은 개인 메일을 입력하며 password에는 그 비밀번호를 적게된다. 여기서 메일은 네이버 서버를 이용하므로 꼭 네이버 메일과 비밀번호를 적도록 한다. 만약 다른 메일을 원한다면, smtpName과 smtpPort를 변경해 줘서 다른 사이트를 이용할 수 있다.

1. **실행**



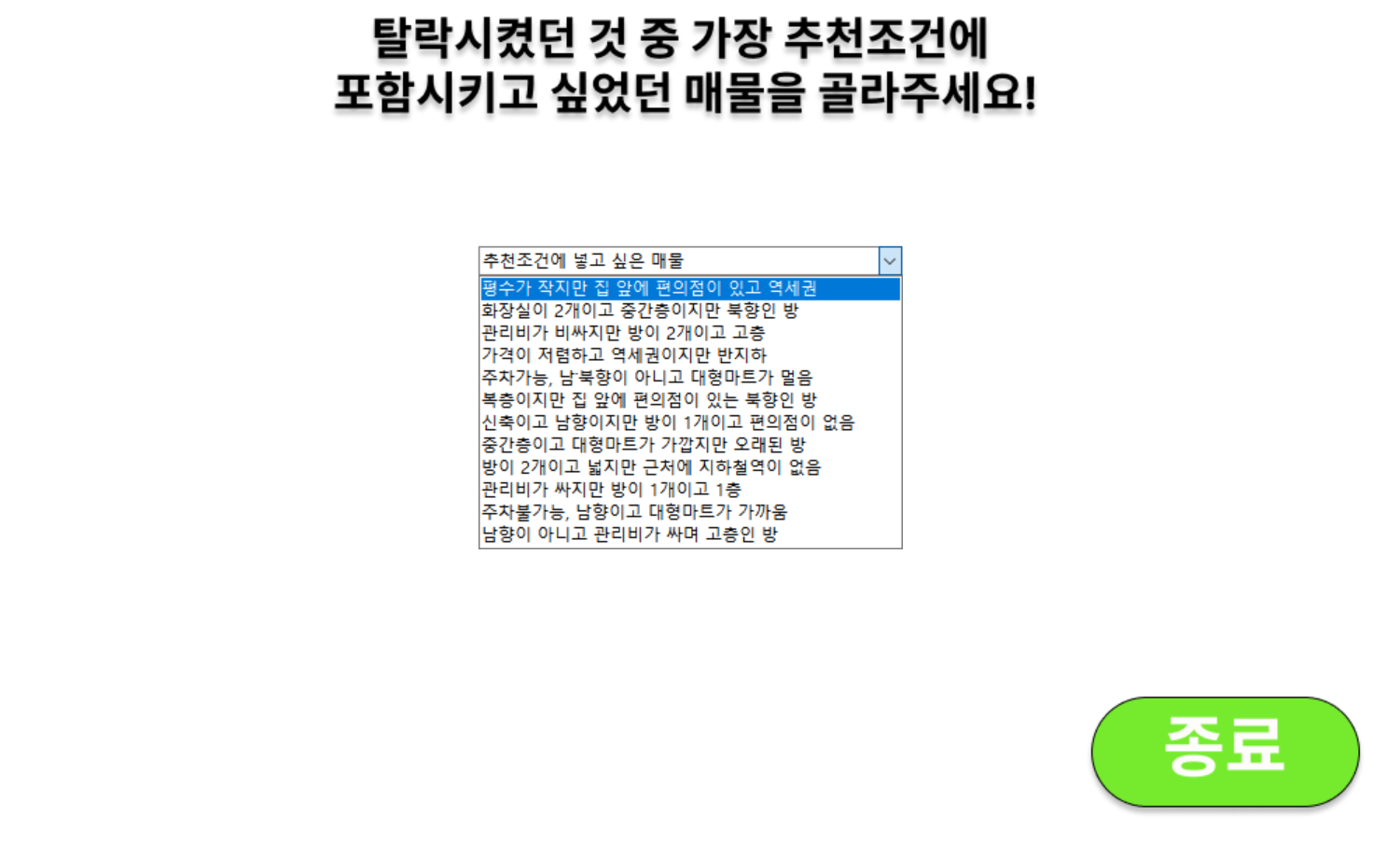
코드 실행의 경우 start\_gui를 통해 진행되는데 이때 꼭 소스코드 압축파일의 이미지 파일이 있어야 진행이 된다. 코드를 실행하면,



다음과 같은 창이 뜨게 되고 사용자는 각 입력 칸에 정보를 입력해준다. 그후 시작버튼을 누르면 이상형 월드컵이 진행되게 된다.

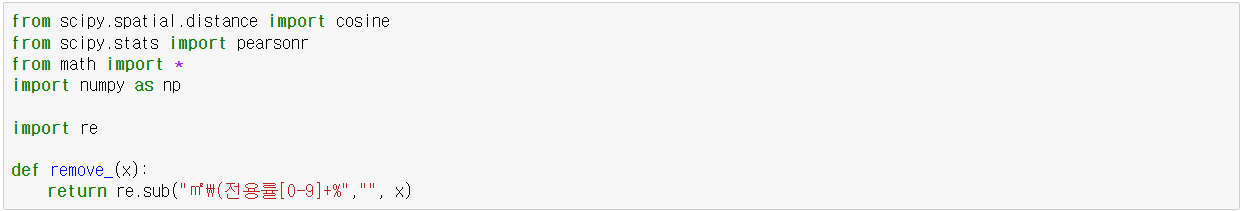


각 라운드별 원하는 혹은 선호하는 방을 선택하면 다음 라운드로 넘어가게 된다.



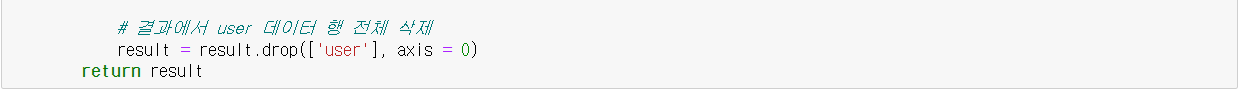
모든 라운드가 끝난 후에 패자부활전으로 넘어가는데 이때 4강전에 탈락한 매물을 선택하여 자신의 선호도 방을 구체화시킬 수 있다. 위와 같은 진행이 끝나면 매물의 추천이 네이버 메일을 통해 결과로 보내지게 된다.

1. **성능평가**



텍스트이(가) 표시된 사진

자동 생성된 설명 텍스트이(가) 표시된 사진

자동 생성된 설명 

필요한 모듈을 호출한 뒤, *RecSys\_for\_Eval* 클래스를 구현한다. 필요한 데이터와 변수를 정의하고, 코사인 유사도와 유클리드 거리 계산 함수를 정의하고, 수집한 사용자의 요구사항에 맞게 데이터를 정리하고, 사용자의 선호도를 바탕으로 생성된 매물 데이터 즉 user 아이템을 포함시켜 추천에 사용할 데이터만 남긴 뒤 전체 데이터를 정규화한다. 각 아이템별 코사인 유사도와 유클리드 유사도를 계산하고 user 아이템 행을 삭제한다.

텍스트이(가) 표시된 사진

자동 생성된 설명 텍스트이(가) 표시된 사진

자동 생성된 설명

본 프로젝트의 추천 모델과 랜덤으로 아이템을 추천하는 랜덤 모델을 대상으로 코사인 유사도와 유클리드 유사도를 기반으로 한 성능을 측정한다. 데이터프레임에서 각 유사도에 해당하는 열만 남긴 뒤 결측치를 삭제하고 내림차순으로 정렬하여 상위 N 개의 데이터만 남기는 top-N 아이템 추천 결과에서 아이템별 Intra-list Similarity 값을 구한다.

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

성능 평가 결과를 각 비교군 별로 막대그래프로 시각화한 결과를 출력한다. 그 예시는 다음과 같다.

