



도서 추천 알고리즘

AI 17 정미정

목차

-
- 서비스 소개
 - 파이프라인
 - 데이터 소개
 - 서비스 시연
 - 보완할 점



서비스 소개

하루 평균 독서 시간



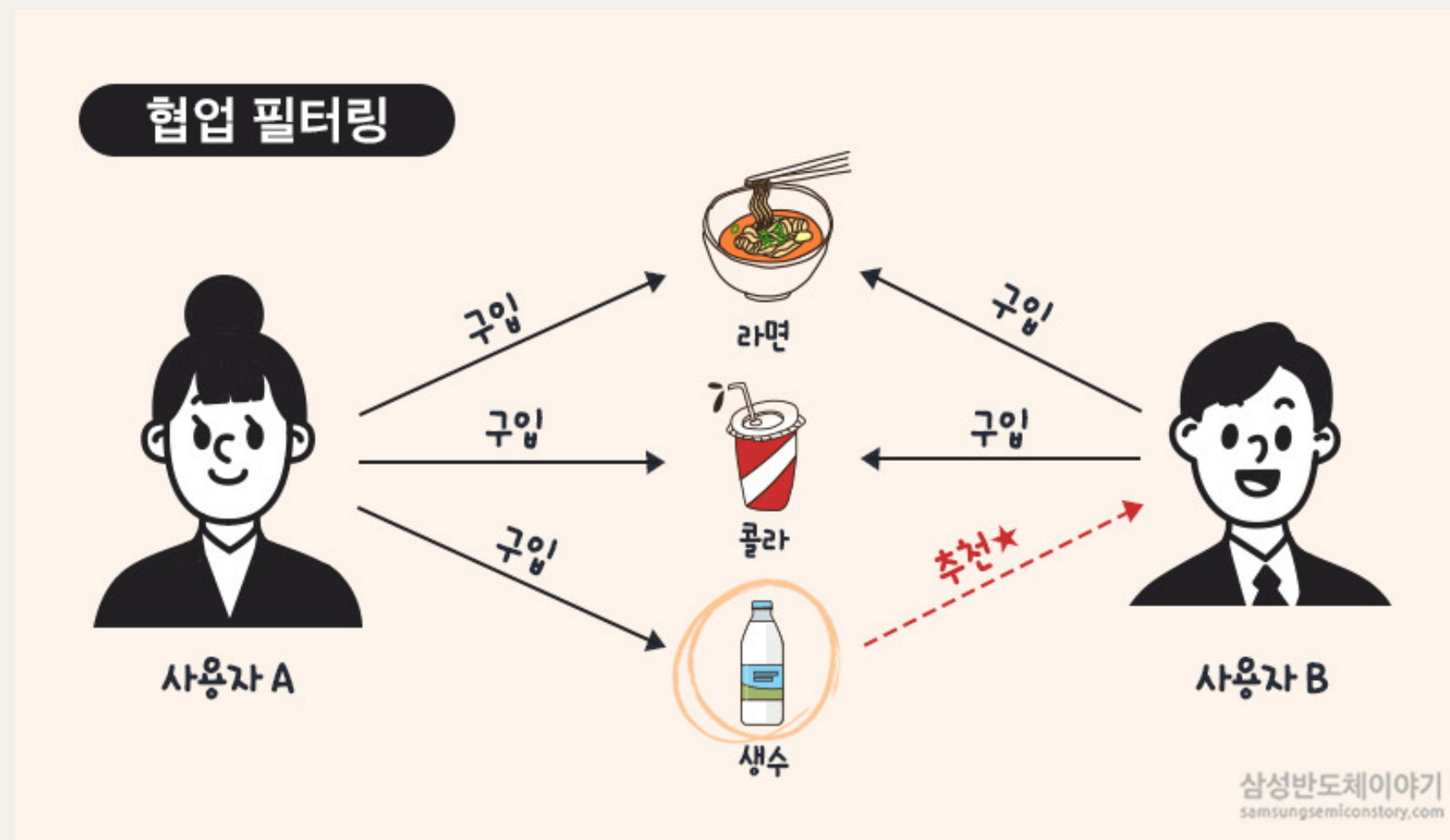
출처: 통계청, 2019 생활시간조사

독서 인구 및 1인당 독서 권수



출처: 통계청, 2019 사회조사

서비스 소개



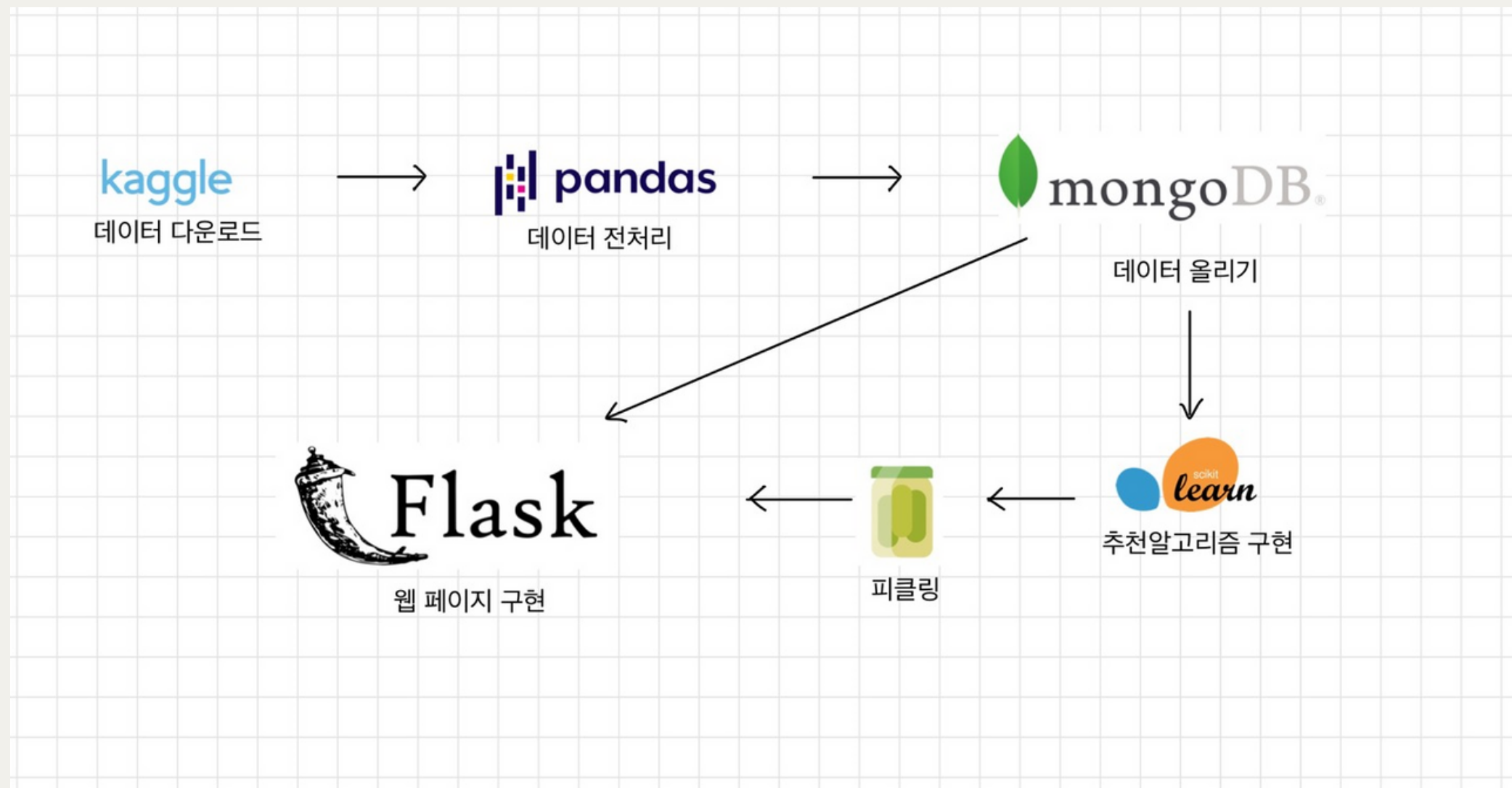
사용자들의 평점 정보



코사인 유사도

유사도가 가장 높은 10개의 책 추천

파이프라인



1. 캐글에서 csv 파일 다운로드
2. 데이터 전처리 & merge
3. pymongo를 활용해 DB에 데이터 올리기
4. sklearn을 활용해 코사인 유사도 계산
5. 결과를 피클링을 통해 .pickle 파일로 저장
6. 데이터베이스 내의 파일과 피클 파일을 불러와서 웹 페이지 내 추천 알고리즘 구현

데이터 소개

캐글: Book Recommendation Dataset

Books.csv



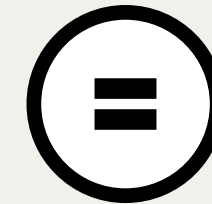
Ratings.csv

책 정보를 담고 있는 데이터 셋

- ISBN: 도서번호
- Book-Title: 책 제목
- Book-Author: 작가
- Year-of-Publication: 출판연도
- Publisher: 출판사
- Image-URL-M: 표지 이미지 URL

사용자들이 책에 매긴 점수를 담고
있는 데이터 셋

- User-ID: 유저 아이디
- ISBN: 도서번호
- Book-Rating: 평점정보



Rating_Book.csv

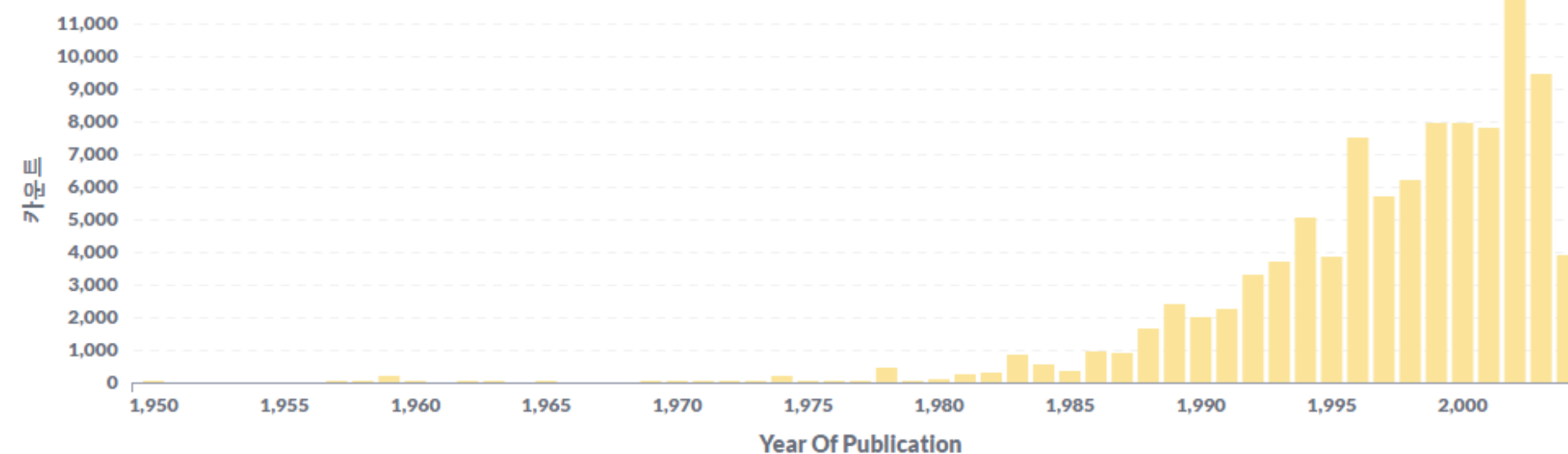
데이터 소개

책 추천 알고리즘 데이터

26,846
user의 수

2,434
책의 개수

출판년도



ratings의 분포



3.13
평균 Ratings

7.76
평균 Ratings(0점 제외)

서비스 시연

한계점

1. 책 제목만을 사용할 수 있음
2. 데이터베이스 내에 입력되어 있는 책 제목 그대로 입력해야 알고리즘 작동
3. 개인의 평점 정보 미사용

THANK YOU

감사합니다

