

# 마스터클래스

- 8조

- kpt 회고(pm, 서로의 task에 대해서 잘 알기)
- 초기 브레인스토밍, 방향 설정 → eda → data processing → baseline → model → ensemble
- 결측치 처리, interaction의 파생 변수 생성, tokenizer 활용, 이미 본 영화와 유저별 마지막 영화 이후에 대해 필터링
- 다양한 모델을 활용하기 위해 RecBole 라이브러리 사용하여 베이스라인 구성 - 오픈소스여서 관리가 제대로 되지 않아서 오류가 많이 존재함
- recall@k에 맞춘 submit 파일의 유사도를 측정하는 파일을 작성하여 사용
- 모델 가중치, 순위 가중치 사용, top-k를 변경해가며 앙상블 진행
- 비대면 협업을 보완하기 위해 메타버스(게더타운) 활용, 노선에 최대한 많은 내용 기록하고 공유, 노선 데이터베이스 활용, 깃 플로우를 브랜치를 개인마다 사용하는 것이 아니고 목적에 따라 사용(data, model 등), 깃 커밋 컨벤션 설정, 깃 pull request 활용
- 개별 스터디 진행

- 4조

- eda: 콜드스타트 문제 확인, 인기있는 영화에 많이 몰림(long-tail)
- 룰베이스 성능을 모델 성능의 기준으로 활용
- sasrec 분석: negative sampling 인덱싱, max\_len, data augmentation(최근 시점 슬라이싱 윈도우)
- test set 구성: 중간에서 마스킹된 아이템 - EASE, next 아이템 - sasrec
- 시드 앙상블 진행