

**SK네트웍스 Family AI과정 3기  
 데이터 전처리 학습된 인공지능 모델**



**□ 개요**

* 산출물 단계 : 데이터 전처리
* 평가 산출물 : 학습된 인공지능 모델
* 제출 일자 : 2024.12.26
* 깃허브 경로 : https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN03-FINAL-2Team
* 작성 팀원 : 이준석, 이주원

| **사용 데이터 및 기술** | * 사용 데이터   + 입력 데이터:     - Title Input : 콘텐츠 제목 정보     - Cast Input : 배우 및 출연진 관련 정보     - Genre Input : 장르 정보   + 처리 과정:     - 각 입력 데이터는 임베딩 레이어(Embedding Layer)를 통해 벡터화     - FM(Factorization Machine) Interaction Layer를 사용해 입력 간 상호작용 관계 학습 * 사용 기술   + Python, TensorFlow, PyTorch   + 딥러닝: Dense Layers, Dropout Regularization   + 머신러닝: FM(Factorization Machine)   + 데이터 전처리: Pandas, NumPy |
| --- | --- |
| **상황** | * 목적:   + 사용자 선호에 기반한 콘텐츠 추천 시스템 개발   + 입력된 메타데이터(제목, 배우, 장르)를 활용하여 최적의 추천 제공 * 배경:   + 기존 추천 모델(FM)의 한계점(비선형 관계 학습 부족)을 해결하기 위해 DeepFM 도입   + FM과 딥러닝 결합으로 콘텐츠 속성 간 상호작용 및 고차원 데이터 학습 |
| **DeepFM 모델** |  |
| **수행 결과** | * 성능 개선   + FM Interaction Layer로 특성 간 상호작용 효과적으로 학습   + ReLU 기반 비선형 학습으로 복잡한 데이터 패턴 추론 가능   + Dropout 적용으로 과적합 방지 및 일반화 성능 강화 * 추천 정확도   + 기존 FM 모델 대비 높은 추천 품질 제공   + 콘텐츠 속성을 종합적으로 반영한 개인화 추천 서비스 구현 |
| **보완점** | * 데이터 부족:   + 더 많은 사용자 데이터와 다양한 콘텐츠 메타데이터 필요 * 학습 속도:   + 모델 구조가 복잡해 학습 시간이 오래 걸릴 수 있음 * 필터링 적용으로 데이터 활용 극대화   + 배우, 장르별로 데이터가 부족해 특정 콘텐츠가 추천되지 않는 문제를 해결하기 위해 필터링 적용   + 데이터 부족 영역에 대해 협업 필터링(Collaborative Filtering)을 추가 도입하여 추천 품질을 개선 |