

## SK네트웍스 Family AI과정 7기

# 데이터 전처리 학습된 인공지능 모델

### □ 개요

- 산출물 단계 : 데이터 전처리
- 평가 산출물 : 성능 평가 모델
- 제출 일자 : 2025-04-14
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN07-FINAL-5Team>
- 작성 팀원 : 김나예

### 사용 데이터 및 기술

- **AP@K 평가지표**를 이용한 채용공고 추천 시스템 성능 평가 모델
  - 개요
    - 단일 유저의 지원 자료를 기반으로 **RAG** 검색 엔진과 **Hybrid** 검색 엔진을 통해 각각 공고를 추천하고, K개의 결과에 대해 **Precision Average**를 구해 두 검색 엔진의 성능을 비교하는 모델
  - 사용 데이터
    - **Ground Truth**
      - 5가지 평가 기준별 정합성 점수로 도출된 **종합 정합성 점수가 0.83 이상**인 채용공고(최고 점수: 0.85)
      - 임베딩 모델 : OpenAI Embedding Model (“text-embedding-ada-002”)
      - 갯수 : 95건
    - **RAG 추천 공고**
      - ChromaDB에 저장된 데이터를 기반으로 검색한 추천 결과
      - 갯수 : 20건
    - **Hybrid 추천 공고**
      - RAG 방식에 Elastic Search 검색 엔진을 추가해 검색한 추천 결과
      - 갯수 : 20건

구현 과정	<ol style="list-style-type: none"> <li>1. 단일 유저에 대한 지원자료(이력서/포트폴리오/자소서)에 대해 특성 추출 <ol style="list-style-type: none"> <li>a. LLM을 활용하여 평가 기준을 기반으로 특성 추출</li> </ol> </li> <li>2. 평가 기준 별 정합성 계산 <ol style="list-style-type: none"> <li>a. 직무 연관성(weight: 0.4)</li> <li>b. 산업군/도메인 일치도(weight: 0.2)</li> <li>c. 기술 스킬 매칭도(weight: 0.3)</li> <li>d. 경력/학력 수준 일치도(weight: 0.05)</li> <li>e. 지역/근무 조건 부합도(weight: 0.05)</li> </ol> </li> <li>3. 공고 별 최종 정합성 계산 후 순위 산정 <ol style="list-style-type: none"> <li>a. 평가 기준별 가중치 부여하여 총점 계산</li> </ol> </li> <li>4. Ground Truth 생성 (Top-N 공고 리스트) <ol style="list-style-type: none"> <li>a. 최종 정합성 임계값(0.83) 이상의 공고를 Ground-Truth로 설정</li> </ol> </li> <li>5. 추천 공고 결과 취합 <ol style="list-style-type: none"> <li>a. LLM 추천 결과(gpt-4)</li> <li>b. RAG 추천 결과</li> <li>c. Hybrid Search 추천 결과</li> </ol> </li> <li>6. AP@K 계산 <ol style="list-style-type: none"> <li>a. 검색 방식별 추천 결과의 Precision Average 계산 후 비교</li> </ol> </li> </ol>
수행 결과	<ul style="list-style-type: none"> <li>● Ground Truth가 각 추천 결과에 포함 되지 않아 추천 성능 비교 불가</li> <li>● 원인 분석 및 개선 작업 진행 중 <ul style="list-style-type: none"> <li>○ 가설 1 : RAG, Hybrid Search 검색 수행 시 채용공고 임베딩 시 불용어, 특수기호 등을 제대로 제거하지 않아 제대로 임베딩 되지 않았다.</li> <li>○ 가설 2 : 각 단계에서의 임베딩 방식이 달라서 정답 판단 기준이 달라졌다.</li> </ul> </li> </ul>
보완점	<ul style="list-style-type: none"> <li>● 실패 원인 분석으로 모델 개선 예정 <ul style="list-style-type: none"> <li>○ 실패 원인에 대한 가설 수립 및 해결하여 추천 성능 비교</li> </ul> </li> <li>● 성능 비교를 통해 공고 추천 시스템 개선 예정 <ul style="list-style-type: none"> <li>○ 검색 성능이 높은 엔진 사용으로 추천 기능 개선</li> <li>○ 정밀도를 높여 사용자 조건에 더 적합한 공고를 추천하는 애플리케이션을 개발</li> </ul> </li> </ul>

