

SK 네트워크 Family AI 과정 7기

데이터 수집 보고서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 데이터 수집 보고서
- 제출 일자 : 2025-03-28
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN07-FINAL-5Team>
- 작성 팀원 : 박민혁, 이재철

<div>데이터 수집 목적</div>	<p>채용 지원을 돕기 위해 다양한 데이터를 수집하여 AI 기반의 분석 및 추천 서비스를 제공하는 것이 목표이다. 이를 위해 다음과 같은 주요 목적을 설정하였다.</p> <ul style="list-style-type: none">● 채용 공고 추천<ul style="list-style-type: none">○ 사용자의 이력서 및 직무 관심사를 분석하여 적합한 채용 공고를 추천한다.○ 다양한 채용 플랫폼(사람인, 캐치 등)에서 제공하는 데이터를 수집하여 최신 채용 정보를 반영한다.○ 기업별 공고 트렌드를 분석하여 사용자에게 보다 정확한 추천을 제공한다.● 예상 면접 질문/답변 생성<ul style="list-style-type: none">○ 기업별, 직무별 면접 질문 패턴을 학습하여 지원자가 예상할 수 있는 면접 질문을 생성한다.○ 면접 데이터를 기반으로 면접자의 답변 유형을 분석하고, 적절한 예시 답변을 제공한다.● 사용자 면접 답변 평가<ul style="list-style-type: none">○ 사용자의 면접 답변을 분석하여 피드백을 제공하고, 개선 방향을 제안한다.○ AI 모델을 활용하여 답변의 논리성, 직무 적합성, 표현력을 평가하고 점수화한다.○ 면접 대비 학습 기능을 지원하여, 사용자가 반복적인 연습을 통해 답변 역량을 향상시킬 수 있도록 한다. <p>데이터 수집의 필요성</p> <p>프로젝트 목표</p> <p>자연어 생성(NLG) 모델을 활용하여 채용 공고 데이터를 분석하고, 사용자의 면접 답변을 평가할 수 있는 AI 기반 채용 지원 서비스를</p>
----------------------	---

	<p>구축하는 것이 목표이다.</p> <p>학습 데이터의 중요성</p> <ul style="list-style-type: none"> ● 채용 공고 데이터: 기업별 채용 공고 데이터를 수집하여 사용자에게 적합한 공고를 추천할 수 있도록 한다. ● 면접 질문/답변 데이터: 기업별 면접 질문 및 면접자의 답변 데이터를 확보하여 면접 준비를 돕고, AI 기반 답변 평가 기능을 구현한다. ● 기업 정보 데이터: 기업 규모 및 산업군 등 기업 정보를 학습하여 맞춤형 채용 정보를 제공한다. <p>면접 준비의 어려움과 AI의 필요성</p> <p>최근 젊은 세대에서는 면접 경험이 부족하거나 대면 커뮤니케이션을 어려워하는 경향이 있다. 비대면 소통에 익숙한 환경에서 성장한 만큼, 실제 면접에서의 긴장감과 즉흥적 대화가 부담으로 작용하는 경우가 많다. 또한, 면접 준비 과정에서 어떤 질문이 나올지 예측하기 어렵고, 답변을 논리적으로 구성하는 데 어려움을 겪는 지원자도 많다. 이에 따라, AI를 활용한 면접 질문 예측 및 답변 생성 시스템은 사용자에게 실질적인 도움을 줄 수 있다. AI 기반의 면접 시뮬레이션을 통해 면접 환경을 체험하고, 자신의 답변을 피드백받아 개선할 기회를 제공함으로써, 보다 자신감 있는 면접을 준비할 수 있도록 지원한다.</p>
<p>데이터 수집 방법</p>	<p>사람인 채용공고 데이터</p> <ul style="list-style-type: none"> ● 웹크롤링 <ul style="list-style-type: none"> - 채용공고의 제목, 본문, 기업명, 기업정보, 회사위치, 이미지 URL 등의 메타데이터 수집 ● API 활용 <ul style="list-style-type: none"> - 직무 대, 중, 소분류 데이터 수집 ● 모집요강 이미지 멀티모달 <ul style="list-style-type: none"> - 이미지 S3 적재 - 이미지 형태의 모집 요강의 경우 멀티모달 비전 모델 사용하여 텍스트 추출 <p>캐치 면접후기 데이터</p> <ul style="list-style-type: none"> ● 웹크롤링 <ul style="list-style-type: none"> - 면접후기의 기업명, 면접질문, 면접팁, 어필역량, 합격여부 등의 메타데이터 수집

수집 데이터(요약)	<ul style="list-style-type: none"> - 사람인 <ul style="list-style-type: none"> - 기업정보(기업명, 기업코드) - 직무별 채용공고 - 약 21,000 건의 채용공고 데이터 - 캐치 <ul style="list-style-type: none"> - 기업정보(기업명, 기업코드, 기업규모) - 기업별 면접 후기(규모 : 대-중견기업) - 약 26,000 건의 기업별 면접 후기 데이터 <p>데이터 구성 및 특성</p> <ul style="list-style-type: none"> - 텍스트 데이터 <ul style="list-style-type: none"> - 채용 공고: 직무별, 산업군 별로 분류하여 수집 - 면접 질문/답변: 기업별 면접 질문 및 면접자의 실제 후기 포함 - 채용 정보: 기업정보 및 직무 분류 등 상세 정보 포함 - 이미지 데이터 <ul style="list-style-type: none"> - 채용 공고 및 모집요강 관련 이미지 포함 - 임베딩 데이터 <ul style="list-style-type: none"> - 채용공고 본문 - 멀티모달로 추출한 채용공고 및 모집요강 텍스트 <p>데이터 저장</p> <ul style="list-style-type: none"> - 텍스트 데이터 <ul style="list-style-type: none"> ● Postgresql ● Elasticsearch - 이미지 데이터 <ul style="list-style-type: none"> ● AWS S3 - 임베딩 데이터 <ul style="list-style-type: none"> ● Chroma
------------	---