

AI 상담 교사 **Fine-tuning**용 데이터셋

1. 데이터셋 구성 개요

1.1 핵심 기능별 데이터 분류

- 적성 평가 데이터: 개인의 잠재 능력과 성향 파악
- 능력 진단 데이터: 현재 학습 수준과 인지 능력 측정
- 진로 추천 데이터: 적성과 능력을 바탕으로 한 진로 방향 제시
- 상담 대화 데이터: 학습자와의 효과적인 소통 방법

1.2 다중 지능 이론 기반 평가 영역

- 언어적 지능: 읽기, 쓰기, 말하기, 듣기 능력
- 논리수학적 지능: 수학적 사고, 논리적 추론 능력
- 공간적 지능: 시각적 인지, 공간 지각 능력
- 음악적 지능: 리듬, 선율, 음성 인식 능력
- 신체운동적 지능: 운동 능력, 손재주, 협응력
- 인간친화적 지능: 타인 이해, 소통, 리더십 능력
- 자기성찰적 지능: 자기 인식, 감정 조절, 목표 설정 능력
- 자연친화적 지능: 자연 관찰, 분류, 패턴 인식 능력

2. 공개 데이터셋

2.1 심리학 및 교육학 연구 데이터

2.1.1 적성 검사 표준화 데이터

- **Holland** 직업 적성 검사: RIASEC 모델 기반 직업 흥미 데이터
- **Strong Interest Inventory**: 직업 관심도 측정 표준 도구
- **Myers-Briggs Type Indicator (MBTI)**: 성격 유형 분류 데이터
- **Big Five** 성격 검사: 5요인 성격 모델 측정 결과
- **Cattell 16PF**: 16가지 성격 요인 측정 데이터

2.1.2 인지 능력 평가 데이터

- **Wechsler Intelligence Scales**: 웨슬러 지능 검사 표준 데이터
- **Stanford-Binet Intelligence Scales**: 스탠포드-비네 지능 검사
- **Raven's Progressive Matrices**: 비언어적 추론 능력 측정
- **Cognitive Assessment System**: 인지 과정 평가 도구
- **Woodcock-Johnson Tests**: 인지 능력 종합 평가

2.1.3 학습 스타일 데이터

- **VARK Learning Styles:** 시각, 청각, 읽기, 운동감각 학습 유형
- **Kolb Learning Style Inventory:** 경험 학습 이론 기반 유형
- **Felder-Silverman Learning Style Model:** 공학 교육 학습 스타일
- **Dunn and Dunn Learning Style Model:** 환경적 요인 포함 학습 유형

2.2 직업 및 진로 정보 데이터

2.2.1 직업 정보 데이터베이스

- **O*NET Database:** 미국 직업 정보 네트워크 데이터
- 워크넷: 한국고용정보원 직업 정보
- 커리어넷: 한국직업능력연구원 진로 정보
- **BLS Occupational Outlook Handbook:** 미국 노동통계청 직업 전망
- **ESCO:** 유럽 기술, 역량, 자격, 직업 분류 체계

2.2.2 산업 및 직업 트렌드 데이터

- **World Economic Forum Future of Jobs:** 미래 직업 전망 보고서
- **McKinsey Global Institute:** 노동 시장 변화 연구 데이터
- **OECD Employment Outlook:** 고용 전망 통계
- 한국고용정보원 직업 전망: 국내 직업별 고용 전망
- **Burning Glass Technologies:** 실시간 구인 정보 분석

2.3 교육 성취도 및 진로 연계 데이터

2.3.1 학업 성취도 추적 데이터

- **PISA 데이터:** 국제 학업 성취도 평가 결과
- **TIMSS 데이터:** 국제 수학·과학 성취도 추이 변화 연구
- **NAEP 데이터:** 미국 국가 교육 진보 평가
- 국가수준 학업성취도 평가: 한국 학생 학력 평가 결과

2.3.2 대학 진학 및 전공 선택 데이터

- **College Board 데이터:** SAT 성적과 대학 전공 선택 상관관계
- **NCES 고등교육 통계:** 미국 대학 진학 및 졸업 통계
- 한국교육개발원 교육 통계: 국내 진학 및 전공 선택 현황
- **QS World University Rankings:** 대학 및 전공별 순위 정보

2.4 진로 상담 사례 데이터

2.4.1 상담 대화 코퍼스

- **Counseling and Psychotherapy Transcripts:** 상담 대화 전사 자료
- **AIBL 상담 대화 데이터베이스:** AI 기반 상담 연구 데이터
- 진로상담 사례집: 한국진로교육학회 발간 상담 사례
- **Career Counseling Case Studies:** 해외 진로 상담 사례 모음

2.4.2 진로 발달 이론 데이터

- **Super's Career Development Theory:** 수퍼의 진로 발달 이론 사례
- **Krumboltz's Social Learning Theory:** 사회 학습 이론 기반 사례
- **Gottfredson's Theory:** 직업 포부 발달 이론 데이터
- **Holland's Theory:** 홀랜드 이론 적용 사례

3. 비공개 데이터셋

3.1 실제 교육 기관 협력 데이터

3.1.1 초·중·고등학교 진로 상담 데이터

- 진로 상담교사 면담 기록: 실제 학생-상담교사 대화 전사
- 진로 탐색 활동 결과: 직업 체험, 진로 캠프 참여 후기
- 진로 희망 설문: 학년별 진로 희망 변화 추적 데이터
- 학부모 상담 기록: 진로 선택에 대한 가족 내 논의 과정

3.1.2 대학 진로 서비스 데이터

- 대학 진로 상담센터: 대학생 진로 고민 및 상담 기록
- 취업 지원 프로그램: 취업 준비 과정 및 결과 데이터
- 인턴십 경험 보고서: 현장 실습 후 진로 인식 변화
- 졸업생 추적 조사: 진로 선택과 실제 취업 현황 비교

3.2 심리 검사 및 평가 데이터

3.2.1 종합 심리 평가 결과

- 병원 신경심리 검사: 인지 능력 정밀 평가 데이터
- 교육청 학습 진단: 학습 부진 및 우수 학생 진단 결과
- 영재교육원 선발 평가: 영재성 판별을 위한 다면 평가
- 특수교육 진단: 학습 장애 및 발달 지연 평가 자료

3.2.2 성격 및 적성 종합 평가

- 청소년 상담복지센터: 청소년 심리 상담 및 검사 결과
- 대기업 인적성 검사: 취업 과정에서 실시하는 적성 평가
- 공무원 시험 심리 검사: 공직 적합성 평가 데이터
- 군대 적성 분류: 군 복무 중 실시하는 적성 검사 결과

3.3 종단 연구 데이터

3.3.1 개인별 성장 추적 데이터

- 유치원부터 고등학교까지 종단 연구: 장기간 개인 발달 추적
- 학습 성취도 변화: 과목별 성적 변화와 적성 발현 패턴
- 관심 분야 변화: 나이에 따른 흥미와 관심사 변화 추적
- 진로 인식 발달: 직업 이해도와 진로 성숙도 변화

3.3.2 환경적 요인 영향 분석

- 가정 환경 변수: 부모 직업, 교육 수준, 경제 상황 영향
- 학교 환경 효과: 학교 유형, 교육 프로그램이 진로에 미치는 영향
- 지역사회 요인: 거주 지역의 산업 구조와 진로 선택 관계
- 동료 집단 영향: 친구 관계와 진로 결정에 미치는 상호 작용

3.4 실제 진로 결정 및 성과 데이터

3.4.1 진로 결정 과정 분석

- 대학 전공 선택 과정: 전공 결정에 영향을 준 요인 분석
- 직업 선택 스토리: 실제 직업 선택 과정의 상세한 기록
- 진로 변경 사례: 전공이나 직업을 바꾼 경우의 배경과 과정
- 창업 및 특수 진로: 일반적이지 않은 진로 선택 사례

3.4.2 진로 만족도 및 성공 지표

- 직업 만족도 조사: 현재 직업에 대한 만족도와 적성 일치도
- 진로 성공 사례: 적성을 잘 발견하고 성공한 사례 분석
- 진로 실패 사례: 부적절한 진로 선택으로 인한 어려움 사례
- 재교육 및 전직: 성인기 진로 변화와 재교육 성과

4. 특수 집단 대상 데이터

4.1 특별한 도움이 필요한 학습자

4.1.1 학습 장애 학생 데이터

- 난독증 학생 진로 데이터: 읽기 장애가 있는 학생의 진로 적응
- **ADHD** 학생 적성 평가: 주의력 결핍 학생의 강점 영역 발견
- 자폐 스펙트럼 진로 사례: 자폐성향 학생의 특별한 재능 활용

4.1.2 영재 및 우수 학생 데이터

- 과학 영재 진로 추적: 과학 분야 영재의 진로 발달 과정
- 예술 영재 사례: 음악, 미술 등 예술 분야 영재 교육 결과
- 조기 입학생 데이터: 월반이나 조기 입학한 학생의 진로 패턴

4.2 다문화 및 특수 환경 학습자

4.2.1 다문화 가정 학생

- 이중 언어 환경: 다국어 능력을 가진 학생의 진로 장점
- 문화적 배경 영향: 부모의 출신국이 진로 선택에 미치는 영향
- 정체성 탐색: 다문화 청소년의 진로 정체성 형성 과정

4.2.2 경제적 어려움 가정 학생

- 저소득층 진로 지원: 경제적 제약 하에서의 진로 개발 전략
- 장학금 수혜자 추적: 교육 기회 확대가 진로에 미친 영향
- 멘토링 효과: 사회적 지원이 진로 발달에 미치는 효과

5. 데이터 통합 및 분석 모델

5.1 종합 평가 모델 구성

5.1.1 다차원 적성 분석

- 인지 능력 프로필: 8가지 지능 영역별 강약점 분석
- 성격 특성 매핑: 5요인 성격과 직업 적합성 연결
- 흥미 패턴 분석: 홀랜드 6각형 모델 기반 직업 흥미 매핑
- 가치관 평가: 개인의 가치 체계와 직업 가치 일치도 분석

5.1.2 학습 성취 패턴 분석

- 과목별 성취도 추이: 시간에 따른 과목별 성적 변화 패턴
- 학습 전략 효과성: 개인별 효과적인 학습 방법 도출
- 동기 수준 변화: 내재적·외재적 동기의 변화와 성취도 관계
- 메타인지 능력: 자기 학습 조절 능력과 학업 성취 관계

5.2 예측 모델 개발

5.2.1 진로 적합성 예측

- 직업 성공 가능성: 개인 특성과 직업 요구사항 매칭도
- 전공 적합도 예측: 대학 전공 선택 성공 가능성 예측
- 진로 만족도 예측: 장기적 진로 만족도 예상 수준
- 진로 변경 리스크: 진로 변경 가능성과 시기 예측

5.2.2 개인화 추천 시스템

- 단계별 진로 로드맵: 현재부터 목표 직업까지의 단계별 계획
- 보완 영역 제시: 부족한 능력이나 경험 영역 식별
- 대안 진로 제안: 주 진로와 유사한 대안적 진로 옵션
- 지원 자원 연결: 진로 개발에 필요한 교육·경험 기회 안내

6. 데이터 품질 관리 및 윤리

6.1 데이터 품질 보장

6.1.1 신뢰성 검증

- 검사 도구 타당도: 사용된 심리 검사의 과학적 타당성 확인
- 평가자 간 일치도: 여러 전문가의 평가 결과 일치도 검토
- 종단적 안정성: 시간 경과에 따른 측정 결과의 안정성 확인

- 교차 검증: 다양한 방법으로 같은 특성 측정 시 일치도 확인

6.1.2 대표성 확보

- 표본의 다양성: 성별, 지역, 사회경제적 배경의 균형
- 문화적 편향 제거: 특정 문화에 치우치지 않은 보편적 기준
- 시대적 적합성: 변화하는 사회와 직업 환경 반영
- 개인차 고려: 극단적 사례와 평균적 사례 모두 포함

6.2 윤리적 고려사항

6.2.1 개인정보 보호

- 민감 정보 익명화: 개인 식별 가능한 모든 정보 제거
- 동의서 확보: 데이터 수집과 사용에 대한 명시적 동의
- 접근 권한 관리: 연구 목적으로만 제한된 접근 허용
- 보관 기간 제한: 연구 완료 후 개인 데이터 안전한 폐기

6.2.2 편견 방지

- 성별 편견 제거: 성별에 따른 직업 고정관념 배제
- 사회적 편견 방지: 계층, 지역, 출신에 따른 차별 요소 제거
- 능력주의 함정: 다양한 형태의 재능과 가치 인정
- 진로 다양성 존중: 전통적이지 않은 진로 선택도 긍정적 평가

6.3 지속적 개선 체계

6.3.1 피드백 수집

- 사용자 만족도: AI 상담을 받은 학습자의 만족도 조사
- 예측 정확도: 실제 진로 결과와 AI 예측의 일치도 추적
- 전문가 검토: 진로 상담 전문가의 AI 성능 평가
- 사회 변화 반영: 새로운 직업과 사회 트렌드 적시 반영

6.3.2 모델 업데이트

- 정기적 재훈련: 새로운 데이터로 모델 성능 개선
- 알고리즘 개선: 더 정확한 분석 방법론 도입
- 평가 기준 조정: 변화하는 교육과 직업 환경에 맞는 기준 개발
- 다학제적 접근: 교육학, 심리학, 사회학 등 다양한 관점 통합