

**SK네트웍스 Family AI과정 6기  
 데이터 수집 및 저장 데이터 수집 보고서**



**□ 개요**

* 산출물 단계 : 데이터 수집 및 저장
* 평가 산출물 : 데이터 수집 보고서
* 제출 일자 : 2025-02-07
* 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-1TEAM>
* 작성 팀원 : 노원재

| **데이터 수집 목적** | **1. 사업보고서**   * 기업 개요, 사업 개요, 재무제표 정보를 분석하여 기업 건전성 및 주요 사업 내용 파악 * 재무제표 정보 시각화 * 재무제표 정보 기반 추이 추론   **2. 뉴스**   * 실시간 트렌드와 기업 관련 이슈를 분석하여 긍/부정 추론 및 키워드 시각화 제공   **3. 재무회계개념체계**   * 기업 재무 데이터 분석 시 참고할 표준 회계 용어 및 기준 확보 * 챗봇 구현시 적절한 해석 제공 |
| --- | --- |
| **데이터 수집 방법** | 1. 사업보고서  * 임직원 정보: api 호출 후 데이터를 csv로 저장 * 사업보고서 원본: api 호출 후 xml 형식으로 저장 * 회사 개요 및 사업 개요: 사업보고서 원본에서 추출 후 csv로 저장 * 재무제표: txt 형식으로 다운로드 후 저장  2. 뉴스  * 빅카인즈(BigKinds)에서 ‘자동차’를 키워드로 한 뉴스 1달 분량 다운로드 * 네이버에서 ‘자동차’ 키워드 검색 후 해당 페이지 뉴스 실시간 크롤링  3. 재무회계개념체계  * 한국회계기준원에서 PDF 파일 다운로드 |
| **수집 데이터(요약)** | 1. 사업보고서  * 기업 개요 및 사업 개요(csv), 재무제표(txt) * https://opendart.fss.or.kr/guide/main.do?apiGrpCd=DS001  2. 뉴스  * ‘자동차’ 키워드 관련 국내외 주요 뉴스 (약 20000건) * 빅카인즈와 네이버 크롤링으로 데이터 수집 (실시간 DB 저장) * 뉴스 키워드 데이터  3. 재무회계개념체계  * 재무회계개념체계(PDF) |