**SK네트웍스 Family AI과정 6기  
 데이터 수집 및 저장 프로젝트 기획서**

**□ 개요**

* 산출물 단계 : 데이터 수집 및 저장
* 평가 산출물 : 프로젝트 기획서
* 제출 일자 : 2025-01-31
* 깃허브 경로 : https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-1TEAM
* 작성 팀원 : 노원재

|  |  |
| --- | --- |
| **프로젝트 주제** | **AI 기반 기업 분석 및 투자 전략 지원 서비스**   1. 내부 고객(직원 및 팀)이 활용할 수 있는 **사업보고서 요약 및 분석 서비스** 2. 투자 및 경영 의사결정을 지원하는 **스마트 검색/분석** |
| **문제정의** | **1. 기업 공시 분석의 어려움**  사업보고서는 전문 용어가 많고 방대한 양의 데이터가 포함되어 있어 분석이 어렵다.  **2. 기존 전자공시 시스템의 한계**  검색 및 문서 열람 기능만 제공되며, 요약, 재무 분석, 리스크 평가 기능이 부족하다.  **3. 최신 트렌드 반영 부족**  기업의 성장성과 리스크를 평가할 때 뉴스 데이터를 활용하는 것이 중요하나, 기존 시스템에서는 이를 제공하지 않는다. |
| **시장조사** | **1. DART(금융감독원 전자공시 시스템)**  금융감독원이 운영하는 기업 공시 정보 플랫폼으로, 사업보고서, 감사보고서 등을 제공하지만, 핵심 정보 추출 및 분석 기능이 부족하다.  **2. 나이스 기업정보**  기업 신용평가 및 재무 정보를 제공하며, 기업의 신용등급과 주요 재무제표를 확인할 수 있지만, 보고서 형태로 제공되어 추가 가공이 필요하다.  **3. DARTPOINT.AI**  AI 기반 공시 데이터 분석 서비스로, DART의 단순 검색 기능을 보완하여 요약 및 인사이트를 제공하지만, 뉴스 데이터 연계 분석은 미흡하다.  **4. 네이버 종목정보**  주가, 기업 개요, 재무제표, 관련 뉴스 등을 제공하지만, 데이터의 깊이가 제한적이며 공시 데이터와 연계된 분석 기능이 부족하다. |
| **시스템 구성** | **1. 사업보고서 요약 시스템**  LLM을 활용하여 사업보고서의 핵심 내용을 자동 요약  **2. 뉴스 트렌드 분석 시스템**  빅카인즈 데이터를 활용하여 기업 및 산업 관련 최신 이슈 분석  **3. 질문 기반 검색(Q&A 시스템, 챗봇)**  사용자가 질문을 입력하면 AI가 기업 공시 및 뉴스 데이터를 종합 분석하여 답변 제공  **4. 기업 분석 및 데이터 시각화**  기업의 재무 상태, 리스크 요인 평가 및 시각적 데이터 제공 |
| **모델링 방안** | **1. 데이터 수집**   * 금융감독원 DART(OpenDart) → 기업 공시 데이터 크롤링 * 빅카인즈(BigKinds) → 뉴스 데이터 수집 * 한국회계기준원 → 사업보고서 내 재무정보 용어 설명 데이터 수집 * 네이버 뉴스 → 실시간 데이터 크롤링   **2. 문서 및 뉴스 데이터 전처리**   * 텍스트 정제 및 주요 키워드 추출 * LLM 기반 트렌드 분석   **3. AI 분석 및 모델 적용**   * LLM을 활용한 사업보고서 요약 * 뉴스 데이터 기반 트렌드 분석 * 사용자 질문 기반 질의응답(Q&A) 제공   **4. 결과 제공 및 시각화**   * 기업별 요약 보고서 제공 * 주요 뉴스 이슈 및 데이터 차트 시각화 |
| **사용데이터** | **빅카인즈** - 뉴스 데이터  **Dart API** - 금융감독원 사업보고서 공시 API 데이터  **한국회계기준원** - 재무회계 개념체계  **네이버 뉴스** - 실시간 크롤링 |
| **R&R** | 박창규(팀장) - PM, 데이터 수집 및 전처리  김지영 - 웹서비스, 데이터 수집 및 전처리  백하은 - 모델링, 데이터 수집 및 전처리  노원재 - 기획, 데이터 수집 및 전처리 |