

**SK네트웍스 Family AI과정 6기  
 모델링 및 평가 LLM 활용 소프트웨어**



**□ 개요**

* 산출물 단계 : 모델링 및 평가
* 평가 산출물 : LLM 활용 소프트웨어
* 제출 일자 : 2025.03.07
* 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-1TEAM.git>
* 작성 팀원: 백하은

| **개요** | * **목적:** 기업 분석 및 투자 전략 수립을 지원하기 위해 **재무데이터, 사업보고서 데이터, 뉴스데이터**를 통합하는 **다중 에이전트 기반 RAG(Retrieval-Augmented Generation) 시스템**을 구축하는 것을 목표로 합니다. 사용자가 질의하면 최적의 데이터를 검색하고, 이를 분석 및 요약하여 유용한 인사이트를 제공하는 AI 기반 서비스 개발을 진행합니다.   **[세부사항]**   * **사용자의 질의를 이해하고** 적절한 데이터를 검색 및 요약하는 다중 에이전트 기반 RAG 시스템을 구축 * **기업 공시 데이터(사업보고서), 재무제표 및 뉴스 데이터를 통합하여** AI 기반 분석 및 응답 생성 * **LLM을 활용한 자연어 처리**를 통해 비전문가도 이해하기 쉬운 형태로 기업 분석 정보를 제공 |
| --- | --- |
| **데이터 전처리** | 1. **검색 및 요약 agent**  * **입력 데이터:** 기업명, 연도, 사업보고서 내용 * **출처:** DART 전자공시 사업보고서 * **전처리 과정:** * PDF → Text 변환 * Markdown(md) 형식으로 변환 * Vector Embedding 후 **ChromaDB** 저장 및 벡터 검색 최적화 * **기능:** 사용자의 질의에 따라 기업별 사업보고서 내용을 요약하여 핵심 정보 제공  1. **예측 agent**  * 입력 데이터: 기업명, 연도, 재무 항목 * 출처: DART 전자공시 대차대조표, 손익계산서, 현금흐름표 * 전처리 과정:   + 단위 변환 및 보정 * 기능 : 기업의 주요 재무 지표 분석 및 미래 재무 전망 제공.  1. **뉴스 agent**  * 입력 데이터: 뉴스 기사, 날짜, 언론사 * 출처: 네이버 뉴스 “국내 자동차” 키워드 검색 * 전처리 과정:   + 감성분석(긍정/부정)   + 키워드 추출 및 관련성 평가   + 최신 데이터 필터링 * 기능: * 뉴스의 요약 * 뉴스를 통한 산업동향 제공 |
| **기대효과** | * **통합적인 기업 분석 지원:** AI를 활용하여 사업보고서 및 재무 데이터를 빠르고 정확하게 요약. * **자연어 기반 요약 및 감성 분석:** 뉴스 데이터를 분석하여 기업 관련 주요 트렌드 및 리스크 요인을 파악. * **의사결정 지원:** 투자 및 경영 의사결정을 위한 종합 분석 정보를 제공하여 데이터 기반 의사결정을 돕는 서비스 제공. |
| **도구 및 환경** | * 프로그래밍 언어: python * 데이터베이스: MySQL, ChromaDB * LLM 모델: OpenAI GPT-4, Langchain 기반 LLM 연동 * 배포: AWS * 데이터 처리: Pandas, SQLAlchemy, PyMySQL * 모델 평가: BLEU, Rouge |
| **결론** | * **주요 성과:**   (1) **사업보고서 벡터 임베딩 후 ChromaDB 저장** → 검색 성능 개선  (2) **LLM을 활용한 자연어 응답 생성** → 데이터 해석 기능 구축  (3) **다중 데이터 소스 통합 분석 가능** → 재무데이터, 사업보고서, 거시경제지표, 뉴스 데이터 연결 |