

**SK네트웍스 Family AI과정 6기  
 모델링 및 평가 테스트 계획 및 결과 보고서**



**□ 개요**

* 산출물 단계 : 모델링 및 평가
* 평가 산출물 : 테스트 계획 및 결과 보고서
* 제출 일자 : 2025.03.07
* 깃허브 경로 : https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-1TEAM
* 작성 팀원 : 박창규

| **개요** | * 본 보고서는 데이터 모델링 및 평가 테스트 과정을 체계적으로 수행하기 위한 계획과 결과를 정리한 문서입니다. 모델링 과정, 평가 테스트 과정, 그리고 결과 보고 및 개선 방향을 세 가지 주요 단계로 나누어 설명합니다. * 본 프로젝트에서는 **Stratify AI RAG Solution**을 기반으로 기업 분석을 수행하는 AI 모델을 개발하였으며, 사업 보고서, 뉴스 데이터, 회계 용어집 등의 다양한 데이터를 활용하여 기업 정보를 정제, 검색, 요약 및 기업 전망 예측을 수행하였습니다. |
| --- | --- |
| **과정1**  **예측** | 1. 모델링 과정1.1 데이터 수집 및 전처리  * **목적**: 신뢰성 높은 데이터를 확보하고 모델링을 위한 최적의 입력 데이터를 구성 * **데이터 출처**: 내부 시스템 데이터, 오픈 DART, BigKinds 뉴스 데이터, 사용자 입력 데이터 * **전처리 과정**:   + 결측치 처리 (삭제 또는 대체)   + 이상치 탐지 및 제거   + 변수 변환 (스케일링, 인코딩 등)   + 데이터 정규화 및 피처 엔지니어링  1.2 모델 설계 및 구현  * **사용 알고리즘**:   + NLP 기반 모델: KPF-BERT (뉴스 및 기업 데이터 분석)   + 머신러닝 모델: Linear Regression (예측 모델)   + AI 검색 모델: Stratify Q (기업 분석 및 검색 기능 제공) * **하이퍼파라미터 튜닝**:   + Grid Search, Random Search, Bayesian Optimization 적용   + Cross-validation을 활용한 최적 파라미터 탐색  1.3 모델 학습 및 성능 평가  * **학습 데이터 분할**:   + 훈련 데이터(80%), 검증 데이터(10%), 테스트 데이터(10%) * **평가 지표**:   + NLP 모델: BLEU, ROUGE   + 예측 모델: MAE, MSE, RMSE, R² Score   + 검색 모델: Precision@K, Recall@K * **학습 과정**:   + 초기 모델 학습 및 기본 성능 분석   + 하이퍼파라미터 최적화 후 성능 비교 |
| **과정2**  **평가 / 테스트** | 2. 평가 / 테스트 과정2.1 테스트 환경 설정  * **테스트 목적**: 실사용 환경에서 모델의 성능 검증 * **테스트 환경**:   + 하드웨어: GPU 서버 (NVIDIA A100), CPU 클러스터   + 소프트웨어: Python (TensorFlow, Scikit-learn, PyTorch, FastAPI) * **테스트 데이터 준비**:   + 미사용 데이터셋 활용 (기업 뉴스, 재무 데이터)   + 실시간 데이터 스트리밍을 활용한 평가  2.2 평가 시나리오 및 방법론  * **기본 성능 테스트**:   + 데이터셋을 활용한 정량적 성능 평가 * **실제 환경 테스트**:   + 운영 환경과 유사한 환경에서 실제 데이터를 기반으로 평가   + 모델 응답 시간 및 실시간 처리 성능 분석 * **A/B 테스트 진행**:   + 기존 모델과 신규 모델 성능 비교   + 사용자 피드백을 통한 추가 조정  2.3 평가 결과 분석  * **주요 성능 지표 결과**:   + 검색 모델 Precision@5: 82%   + 예측 모델:  | MAE | MSE | RMSE | R² | | --- | --- | --- | --- | | 93441775852.33 | 2.56 | 160114256175.26 | 0.95 |  * + NLP 모델 BLEU Score: 89.5 * **모델 개선 필요 사항**:   + 실시간 분석 성능 최적화 필요   + 데이터 증강 기법 추가 적용 검토   + 모델 업데이트 자동화 및 성능 모니터링 강화 |
| **결과** | 3. 결과 보고 및 개선 방향3.1 평가 결과 요약  * Stratify AI RAG Solution을 기반으로 **기업 데이터 검색 및 분석**을 수행할 수 있는 AI 모델을 개발 * NLP 모델을 활용한 뉴스 요약 및 기업 리스크 분석 기능 구현 * 실사용 환경에서 예측 성능 및 검색 정확도 **80% 이상**을 기록하며, 개선 가능성이 높은 모델 구조 확보  3.2 개선 및 최적화 계획  * 데이터 전처리 강화 및 추가 피처 엔지니어링 * AI 기반 검색 모델의 문서 임베딩 최적화 * 기업 전망 예측 모델의 추가적인 데이터 학습 및 고도화 * API 기반 서비스 확장을 통한 상용화 검토  3.3 결론  * **모델 배포 가능성 검토 결과**, 현업 적용을 위한 충분한 성능을 확보하였으며, 일부 최적화를 거친 후 **단계적 배포** 진행 예정 * 실시간 질의응답 기능을 더욱 정밀하게 개선하기 위해 추가적인 하이퍼파라미터 튜닝 및 데이터 증강 기법을 적용할 계획 * 유지보수를 위해 **자동 모니터링 시스템**을 도입하여 지속적인 성능 개선을 추진 * **향후 발전 방향**: 기업 데이터와 외부 금융 데이터를 결합한 **AI 기반 투자 전략 지원 시스템**으로 확장 검토   본 보고서는 **Stratify AI RAG Solution**을 기반으로 한 AI 기업 분석 모델의 설계, 평가, 개선 방향을 포함하고 있으며, 실사용 환경 적용 가능성을 높이는 방향으로 지속적으로 발전해 나갈 것입니다. |

| **요약** | 2. 평가 / 테스트 과정2.1 테스트 환경 설정  * **테스트 목적**: 실사용 환경에서 모델의 성능 검증 * **테스트 환경**:   + localhost 환경에서 실행 * **테스트 데이터 준비**:   + Chroma DB 에 저장된 pdf 사업보고서 활용   + 요약 -> pdf를 gpt에게 pdf기반으로 질문에 대한 답변을 요약 후 결과 값 받음  2.2 평가 시나리오 및 방법론  * **기본 성능 테스트**:   + 데이터셋을 활용한 정량적 성능 평가 * **실제 환경 테스트**:   + 운영 환경과 유사한 환경에서 실제 데이터를 기반으로 평가   + 모델 응답 시간 및 실시간 처리 성능 분석  2.3 평가 결과 분석  * **주요 성능 지표 결과**:   + 검색 모델 Precision@5: 82%   + 예측 모델:  | Precision | Recall | BERTScore-F | 코사인 유사도 | | --- | --- | --- | --- | | 0.7727 | 0.8194 | 0.7950 | 0.6742 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

| **검색** | 2. 평가 / 테스트 과정2.1 테스트 환경 설정  * **테스트 목적**: 실사용 환경에서 모델의 성능 검증 * **테스트 환경**:   + localhost 환경에서 실행 * **테스트 데이터 준비**:   + Chroma DB 에 저장된 pdf 사업보고서 활용   + 검색 -> 사용자 질문의 값을 gpt를 통해 pdf파일에서 추출, 또는 pdf의 명시되어있는 데이터를 직접 확인 후 테스트셋 생성성  2.2 평가 시나리오 및 방법론  * **기본 성능 테스트**:   + 데이터셋을 활용한 정량적 성능 평가 * **실제 환경 테스트**:   + 운영 환경과 유사한 환경에서 실제 데이터를 기반으로 평가   + 모델 응답 시간 및 실시간 처리 성능 분석  2.3 평가 결과 분석  * **주요 성능 지표 결과**:   + 검색 모델 Precision@5: 82%   + 예측 모델:  | Precision | Recall | BERTScore-F | 코사인 유사도 | | --- | --- | --- | --- | | 0.6932 | 0.8067 | 0.7427 | 12.43 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |