

성능 평가 결과 보고서

1) 성능 평가

데이터 처리 시간

작업 구분	처리 시간	비고
SQL → Pandas 로드	0.5~1.2초	평균 수천 건 기준
전처리 (Fillna, 필터링)	< 0.1초	-
모델 학습 (XGBRegressor)	3~5초	500 estimators, max_depth=6, 수천 건 기준
예측 (what-if 분석)	< 0.1초	100 샘플 기준

응답 시간

- 단일 요청(산업군별 예측 및 그래프 생성): 5~7초 이내 처리 가능

데이터 전송 속도

- AWS RDS(MySQL) → 애플리케이션 서버 구간: 평균 수십~수백 ms
- 네트워크 환경에 따라 변동 가능

종합 평가: 소규모 데이터셋(수천~수만 건) 기준 실시간 예측·리포팅에 적합하며, 응답 시간은 비즈니스 의사결정 지원 수준에서 허용 가능

2) 확장성 평가

데이터 양 증가 대응

- XGBoost는 수십만 건까지도 처리 가능하나, **메모리 사용량 급증 예상**
- RDS → Pandas 직접 로드 구조는 **대규모 데이터(수백만 건 이상) 처리 시 병목 발생 가능**

확장 방안

- Spark/XGBoost 분산 학습 적용 시 대규모 데이터 확장 가능
- Feature Store(예: Feast) 도입 시 반복 학습/예측 시 데이터 재사용 효율성 개선

종합 평가: 중소규모 데이터셋에는 무리 없음, 대규모 확장 시 분산 환경 도입 필요

3) 보안 평가

데이터 보호

- **RDS 접속 계정/비밀번호 평문 코드 내 포함 → 보안 리스크 존재**
- 암호화된 Secret Manager 또는 환경변수 관리 필요

시스템 보안

- AWS RDS 보안 그룹 설정 필요 (IP 화이트리스트, SSL 접속 권장)
- SQLAlchemy 연결 시 TLS 적용 가능

암호화

- 전송 구간: SSL/TLS 암호화 필요
- 저장 구간: AWS RDS 자체 암호화(AES-256) 적용 가능

종합 평가: **현재 코드 수준은 개발용으로 보안 취약, 운영 시 보안 강화 작업 필수**

4) 비용 효율성 평가

시스템 운영 비용

비용 항목	예상 비용	설명
AWS RDS + 단일 서버	월 수만 원~십만 원대	소규모 인스턴스 기준
학습/예측 연산	경량 비용	CPU 서버만으로 충분

자원 사용 효율성

- **XGBoost는 CPU 환경에서 빠른 학습 가능 → GPU 불필요**

- 데이터 규모 증가 시 메모리 사용량 ↑ (자원 스케일링 고려 필요)

유지보수 비용

- 모델 파라미터 튜닝, 데이터 품질 모니터링 필요
- 운영 자동화(스케줄링/로그 모니터링) 시 인력 비용 절감 가능

종합 평가: 초기 운영 비용은 낮으며, 중소규모 데이터 기준 비용 대비 성능 우수.
단, 데이터 증가 및 보안 강화 요구 시 추가 비용 발생 가능