

클라우드 기반 대출 상환 위험 모니터링 및 예측 플랫폼 개발

오픈소스 URL: <https://github.com/RiskFinderTeam/risk-finder-be.git>
<https://github.com/RiskFinderTeam/risk-finder-fe.git>
<https://github.com/MiLLku/dfag.git>



2025학년도 2학기
**SW 캡스톤디자인
경진대회**

팀 명 리스크파인더

지도교수 강동기

팀 원

산업체

신서현(컴퓨터인공지능학부, 4), 박준영(컴퓨터공학부, 4),
박민호(컴퓨터공학부, 4), 임다미(컴퓨터공학부, 4)

개발 동기 및 목적

[현황 및 문제 인식: 위기의 금융 환경]

- 가계부채 및 연체율 급증: 대한민국 가계부채 비율이 세계 최고 수준을 기록하고 고금리 기조가 장기화됨에 따라, 은행권 연체율이 69개월 만에 최고치(0.53%)를 경신하는 등 자산 건전성 관리가 국가적 과제로 대두됨.

- 기존 관리 체계의 한계: 급증하는 관리 대상 대비 심사 인력은 부족하여 업무 과부하가 발생하고 있으며, 수작업 위주의 사후 관리는 인적 오류(Human Error)를 유발하고 데이터에 기반한 정밀한 예측을 수행하지 못하는 구조적 한계가 존재함.

[개발 목적: 데이터 기반의 선제적 대응]

- AI 기반 선제적 위험 예측: 연체 발생 후 대응하는 사후 관리를 넘어, 빅데이터 분석을 통해 상환 위험 징후를 사전에 포착하고 부실 채권을 예방하는 'Risk Finder' 플랫폼 개발.

- 설명 가능한 AI(XAI) 도입: 단순한 예측 결과뿐만 아니라 위험 요인(Why)을 투명하게 설명하여 실무자의 신뢰도 확보 및 신속한 의사결정 지원.

- 클라우드 네이티브 환경 구축: AWS Serverless 아키텍처를 적용하여 유휴 자원을 최소화하고, 대용량 데이터를 안정적으로 처리하는 비용 효율적인 자동화 시스템 구현.

주요 기술

[1] AI 모델링 및 최적화 (Prediction Engine)

- 3-Model 앙상블: 정형 데이터에 강력한 성능을 보이는 XGBoost(35%), LightGBM(35%), CatBoost(30%) 3종 모델을 가중 평균(Weighted Blending)하여 단일 모델 대비 예측 안정성과 정확도(AUC 0.795)를 극대화함.

- Optuna 하이퍼파라미터 튜닝: 베이지안 최적화 프레임워크인 Optuna를 활용, 각 모델당 50회 이상의 시행(Trial)을 거쳐 최적의 학습 파라미터를 자동 탐색함.

[2] 금융 특화 피처 엔지니어링

- 파생 변수 고도화: 7개 테이블(신상, 대출 이력, 타사 정보 등)을 통합하고, '납부 지연 일수(DPD)', '소득 대비 상환 비율(DTI)' 등 금융 리스크를 대변하는 200개 이상의 파생 변수를 생성.

- Target Encoding (5-Fold CV): 범주형 변수(직업 등) 처리 시 5-Fold Cross Validation 기반의 타겟 인코딩을 적용하여, 정보 손실을 줄이면서도 과적합을 원천 차단함.

[3] 시스템 인프라 및 웹 개발 (Cloud & Web)

- AWS Serverless Batch: Amazon EventBridge와 ECS(Fargate)를 연동하여, 데이터 업로드 시 별도의 서버 관리 없이 예측 컨테이너가 자동 실행되는 이벤트 기반 아키텍처 구현.

- Full-Stack Web: Spring Boot(Backend)로 안정적인 데이터 처리를 보장하고, Vue.js(Frontend)로 시각화된 대시보드를 구축하여 사용자 편의성 제공.

개발 내용

[1] 하이브리드 시스템 아키텍처 (Web Zone & AI Zone)

- Web Zone (서비스 영역): 관리자 접근성을 위해 EC2 상에 웹 애플리케이션을 배포하고, RDS(MySQL)와 Redis를 활용하여 빠른 데이터 조회 및 세션 관리 환경 구축.

- AI Zone (분석 영역): 대용량 연산이 필요한 AI 모델은 Docker Container로 패키징하여 ECR에 저장하고, 배치 작업 시에만 ECS Fargate를 통해 구동되도록 설계하여 운영 비용 최적화.

[2] 자동화된 배치 스코어링 파이프라인

- End-to-End 자동화: S3 버킷에 고객 데이터(CSV)를 이용해 주기적으로 EventBridge가 트리거되어 분석 파이프라인이 시작됨.

- 처리 프로세스: [전처리 & 피처 엔지니어링] → [앙상블 모델 예측] → [등급 산출(A~E)] → [SHAP 위험요인 분석] → [DB 적재]의 전 과정이 사람의 개입 없이 자동으로 수행됨.

[3] 관리자 리스크 모니터링 대시보드

- 종합 관제 화면: 전체 고객의 위험 등급 분포, 시간 흐름에 따른 연체율 추이 등을 차트와 그래프로 시각화하여 거시적 현황 파악 용이.

- 고객 상세 분석: 특정 고객 클릭 시 모달 창을 통해 AI 산출 등급과 핵심 위험 요인(TOP 3 Feature)을 상세히 제공하여 정밀 심사 지원.

결과 및 분석

1. 비용 절감 및 효율성 극대화

수작업 모니터링을 자동화하여 업무 효율성을 높이고, 부실채권 정리 비용을 절감하여 자산 건전성을 개선함.

Serverless 아키텍처를 적용하여 유휴 자원 없이 필요한 시점에만 자원을 사용하여 운영 비용을 최적화함.

2. 직관적인 리스크 시각화

관리자 대시보드를 통해 위험 고객 목록, 등급 분포, 연체율 추세 등을 그래프로 시각화하여 복잡한 데이터를 한눈에 파악하고 신속한 의사결정을 지원함.

3. 설명 가능한 선제적 조치 (XAI)

SHAP 분석을 통해 고위험군으로 분류된 구체적인 원인(TOP 3)을 투명하게 제시함.

이를 근거로 맞춤형 상담 및 선제적 조치를 취하여 부실 채권 발생을 사전에 억제함.

