

# SHAP-XAI 기반 STR 보고서 자동화 시스템

SW기초교육  
AJOU  
SOFTCON

과 목 명 금융프로그래밍1

학 과 소프트웨어학과

지도교수 김현균

이 름 이에림

## 연구배경 및 필요성

### ① 금융 사기의 심각성과 규제 환경

금융 사기는 매년 전 세계적으로 막대한 피해를 발생시키는 심각한 문제이며, 국내 금융기관은 의심거래 발생 시 30일 이내 STR(의심거래보고)을 제출해야 하는 규제를 따른다. 기존 규칙 기반 탐지 방식은 새로운 사기 패턴에 취약하고 오탐률이 높으며, 머신러닝 기반 탐지는 높은 정확도를 보이지만 판단 근거를 설명하기 어려운 ‘블랙박스’ 한계를 가진다. 이는 감독기관 보고와 내부 검증에서 신뢰성 저하로 이어질 수 있다.

### ② 선택한 모델과 이유

이 문제를 해결하기 위해 본 프로젝트는 대용량·고차원 금융 거래 데이터 처리에 강점이 있는 LightGBM과 안정적인 예측을 제공하는 Random Forest를 활용했다.

### ③ XAI와 SHAP 기법의 필요성과 특징

XAI(Explainable AI)는 인공지능 모델이 내린 예측 결과의 이유를 사람이 이해할 수 있도록 설명하는 기술로, 금융 사기 탐지처럼 규제와 신뢰가 중요한 분야에서 필수적이다. 기존 머신러닝 모델은 ‘블랙박스’ 특성으로 인해 예측 근거를 파악하기 어렵지만, XAI는 이를 시각화와 수치로 해석 가능하게 만든다.

이 프로젝트에서는 XAI 기법 중 SHAP(Shapley Additive Explanations)을 적용했다. SHAP은 게임이론 기반 방법으로, 각 피처가 예측 결과에 기여한 정도를 정량화하며 전역(모델 전체)·로컬(개별 거래) 수준에서 모두 해석이 가능하다. 예를 들어, 사기 거래로 분류된 경우 거래 유형, 금액 대비 잔액 비율, 송금 전 잔액 등이 각각 얼마나 영향을 미쳤는지를 한눈에 파악할 수 있어, 실무자가 규제 준수 보고서 작성 시 명확하고 신뢰성 있는 근거를 확보할 수 있다.

## 프로젝트 진행과정

본 프로젝트는 총 636만 건의 PaySim 시뮬레이션 금융 거래 데이터를 대상으로, 데이터 전처리 → 모델 학습 및 최적화 → XAI 분석 → 웹 대시보드 개발 → LLM 기반 STR 보고서 자동화의 5단계로 구성되었다.

```
(xai_env) yelim@yelim-ui-MacBookPro XAI % python scripts/generate_report.py
[✓] 모델 로드 완료 : /Users/yelim/Desktop/XAI/outputs/models/model.joblib
[✓] SHAP 분석 수행 중...
[✓] 개별 거래 SHAP 설명 생성 중...
[✓] 개별 거래별 LLM 설명 생성 중...
[✓] 개별 거래 SHAP 시각화 저장 중...
[✓] LLM 설명 생성 중...
[✓] 보고서 생성 중...
[✓] 보고서 생성 완료 : /Users/yelim/Desktop/XAI/outputs/reports/STR_Report.md
```

### 1단계 : 데이터 분석 및 전처리

PaySim 데이터셋은 실제 금융 거래 패턴을 모방한 6,362,620건의 시뮬레이션 데이터로, 거래 시간(step), 거래 유형(type), 거래 금액(amount), 잔액 정보, 사기 여부(isFraud), 은행 사기 플래그(isFlaggedFraud) 등 원본 11개 컬럼으로 구성된다.

데이터 분석 결과, 사기 거래는 8,213건(0.13%)에 불과해 정상 거래 대비 약 1:773 비율로 극도로 불균형했다. 이를 해결하기 위해 SMOTE 기법을 적용해 정상/사기 거래를 1:1 비율로 맞췄다.

피처 엔지니어링 과정에서는 거래 유형 원핫 인코딩(5개), 금액 관련 피처(2개), 송금·수취인 잔액 변화량(6개), 거래 발생 시간 관련 피처(2개) 등 총 15개 핵심 파생 피처를 생성했다.

이후 시간 순서를 고려해 데이터를 훈련(60%), 검증(20%), 테스트(20%)로 분할하고, StandardScaler를 적용해 모든 피처의 스케일을 통일했다.

### 2단계 : 모델 학습 및 최적화

LightGBM과 Random Forest 모델을 학습에 사용했으며, Grid Search를 통해 n\_estimators=200, max\_depth=7, learning\_rate=0.1을 최적 파라미터로 결정했다. 불균형 데이터 특성을 고려해 Accuracy 대신 F1-Score와 AUPRC를 중심으로 평가한 결과, 테스트 데이터 약 127만 건(전체 데이터의 20%)에 대해 정밀도 86.7%, 재현율 95.1%, F1-Score 90.7%, AUPRC 74.2%를 달성했다.

### 3단계 : XAI 분석(SHAP)

모델이 특정 거래를 사기라고 판단한 이유를 수치와 시각화로 제공하기 위해 SHAP 분석을 적용했다. 전체 데이터셋에 대해 주요 영향 피처는 거래 유형(type\_CASH\_IN),잔액 대비 거래 금액 비율, 송금 전 잔액이었으며, 개별 거래별로 SHAP Waterfall Plot을 통해 각 피처의 긍정·부정 기여도를 시각적으로 제시했다. 예를 들어, 잔액이 0원인 계좌에 고액 입금 후 단시간 내 인출되는 패턴이 대표적인 사기 탐지 근거로 나타났다.

### 4단계 : 웹 대시보드 개발

Flask 백엔드와 HTML/CSS/JS 프론트엔드로 대시보드·상세분석, 보고서 생성 3개 화면을 구현했다. 대시보드 페이지에서는 정밀도·재현율·F1·AUPRC 등 핵심 지표와 분류 현황, 상위 SHAP 피처를 한눈에 제공하며, 상세 분석 페이지에서는 SHAP Bar/Beeswarm로 전역 중요도와 분포를, Waterfall Plot로 개별 거래의 판단 근거를 시각화한다.

### 5단계 : LLM 기반 STR 보고서 자동화

보고서 화면에서 거래 ID와 보고서 유형을 선택하면 GPT-4o-mini가 SHAP 결과를 자연어로 해석해 규제 보고 형식의 초안을 자동 생성한다.

## 프로젝트 성과 및 기대효과

깃허브 링크 : [https://github.com/yelim8902/XAI\\_AML](https://github.com/yelim8902/XAI_AML)

### [메인 페이지]

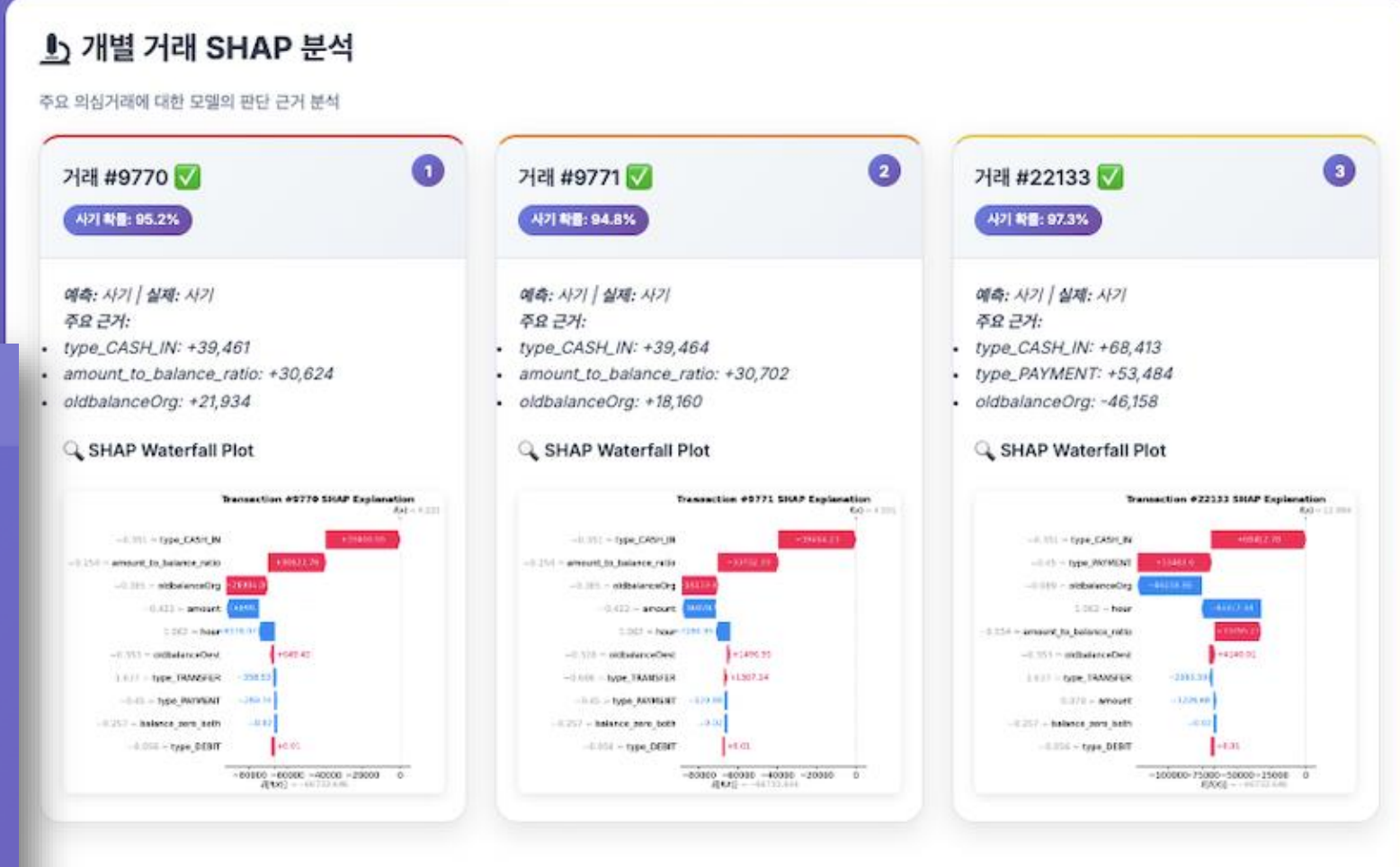
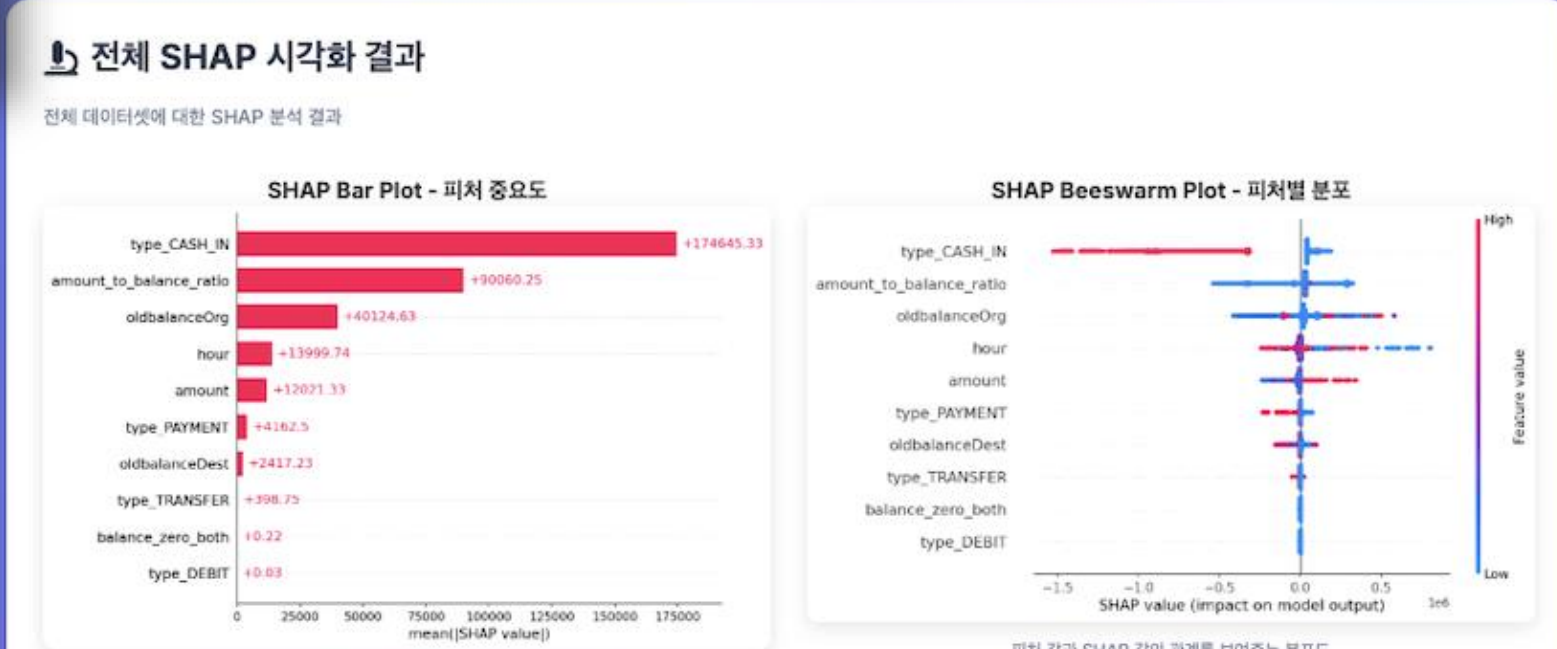
➤ 이 페이지는 XAI 금융 사기 탐지 시스템의 대시보드 역할을 한다

- **모델 성능 요약:** 정밀도 86.7%, 재현율 95.1%, F1-Score 90.7% 등 핵심 성능 지표를 한눈에 볼 수 있다.
- **전체적인 모델의 성과:** PaySim 데이터 127만 건으로 실험해서 F1-Score 90.7%로 사기를 탐지했다.
- **실시간 위험도 모니터링:** 위험 거래 현황: 탐지된 위험 거래들의 상태와 위험도를 색깔로 구분해서 보여준다.

### [상세 분석 페이지]

➤ 이 페이지는 AI 모델이 어떻게 사기를 탐지하는지 자세히 분석한다.

- **SHAP 시각화:** 전체 데이터셋에 대한 SHAP Bar Plot과 Beeswarm Plot으로 피처별 영향력을 시각적으로 보여준다.
- **주요 피처 설명:** 사기 탐지에 가장 중요한 영향을 미치는 상위 피처들을 자세히 설명한다. 현재 데이터에서는 거래 유형(type\_CASH\_IN), 잔액 대비 거래 금액 비율, 송금 전 잔액 등이 판단에 주요한 요소로 쓰임을 알 수 있다.
- **개별 거래 분석:** 각 거래별로 SHAP Waterfall 차트를 통해 어떤 특성이 사기 판단에 기여했는지 상세히 분석한다.



### [보고서 생성 페이지]

➤ 이 페이지는 AI가 SHAP 분석 결과를 바탕으로 보고서를 자동으로 작성해준다.

거래 ID와 보고서 유형(상세 분석/경영진 요약/규제 기관)을 선택하면 AI가 자동으로 보고서를 작성한다.

본 프로젝트는 금융 사기 탐지의 정확도·투명성·속도를 동시에 혁신적으로 향상시킨다. SHAP 기반의 해석 가능성을 통해 탐지 근거를 명확히 제시함으로써 감독기관 신뢰와 규제 준수를 지원하고, LLM 자동 보고서 생성으로 수 시간 소요되던 STR 작성 업무를 1~2분 내로 단축하여 업무 효율을 극대화할 수 있다.

나아가 다양한 금융기관과 거래 유형으로의 확장이 가능해, 차세대 금융 보안 솔루션으로서의 가치를 가질 수 있다.

