

구 분	세 부 내 용		
과 제 명	머신러닝 기반 예산관리		
참여인원	3명(장정영, 원재민, 이예은)		
추진기간	'25.9월~ (진행중)		
과제목적	기관의 예산 집행 과정에서 발생할 수 있는 이상거래(비정상 지출 패턴)를 조기에 발견하고, 예산 흐름을 한눈에 파악할 수 있는 머신러닝 기반 예산관리 시스템 구축		
추진경과	단계	기간	추진내용
	기획	9월 1주~ 9월 2주	- 과제 범위 및 목표 설정 - ML 기법 및 도구 선정
	데이터 수집	9월 3주~ 10월 2주	- 과거 예산 집행 데이터 수집 - 데이터 항목 정의
	데이터 전처리 및 EDA 등	11월 3주~ 12월 3주	- 데이터 정제 및 품질 검증 - 결측치, 이상치 처리 - 탐색적 데이터 분석 - 특성(Feature) 설계 - 비지도 학습 기반 모델 개발
	테스트	-	- 모델 사용자 검증 등
			완료여부
			완료
			완료
			진행중 (비지도 기반 예산관리 모델 Prototype 개발 완료)
			미착수
개발워크 플로우	○ 전체 워크플로우		
	구분	내용	완료도
	기획	시스템 목표 설정 → 사용자 요구 조사 → 화면/기능/모델 목표 정의 ※ 기획 단계는 완료, 기능 범위 확정됨	100%
	데이터 수집	ERP RAW 데이터 확보(2020~2024) → 자동 파일 스캔·로드 → 컬럼 매핑 ※ 실제 운영 데이터 기반 학습 준비 완료	100%
	전처리	Pandas 기반 정제 파이프라인 구축 - 결측치 처리 - 날짜/금액 통일 - 이상값 제거 - 데이터 표준화 ※ RAW → CLEAN 변환 안정화됨	100%
	EDA	반복지출 패턴 / 부서 평균 대비 편차 / 지출 빈도 / 계정과목 특성 반영 ※ 모델 성능에 필요한 핵심 Feature 구성 대부분 완료	90%
	특성(Feature) 엔지니어링	반복지출 패턴 / 부서 평균 대비 편차 / 지출 빈도 / 계정과목 특성 반영 ※ 모델 성능에 필요한 핵심 Feature 구성 대부분 완료	80%

	↓		
	비지도 학습 기반 모델 개발	Isolation Forest (100%) K-means Clustering (100%) 이상치 스코어링 (80%) 임계값 설정 (70%) 가중치 보정 (40%) ※ 프로토타입 기준 단독 운영 가능한 수준	85%
	↓		
	웹 앱 개발(React + Flask)	React UI (70%) - 업로드 / 대시보드 / 상세 페이지 Flask API (60%) - 전처리 / 모델 실행 / 결과 반환 시각화 (60%) - 월별/부서별 추세 - 이상거래 히트맵 ※ 프로토타입 실행 가능한 수준	65%
	↓		
	통합 테스트	React ↔ Flask ↔ 모델 연동 점검 대용량 데이터 처리 테스트 부서별 사용자 검증(파일 단위) ※ 기본 연동 성공, 실사용 검증 단계 PDF/Excel 자동 보고서 생성(40%) 로그·감사 이력 설계(20%) 운영용 Evidence 보관 구조 구축(진행중)	40%
- 활용 도구별 워크플로우 : 모델 개발			
단계	활용도구	용도	진행상태
1. 데이터 수집	ERP 원천데이터(엑셀), Python(os, pandas)	연도별 지출데이터 자동 탐색 및 로드	완료
2. 데이터 정제 (전처리)	Python(pandas, numpy), 자체 전처리 스크립트	결측치 처리, 형식 표준화, 수치형 변수 자동 변환	완료
3. EDA	Python(pandas, matplotlib, seaborn)	지출 분포·패턴 분석, 이상값 탐색, 변수 구조 파악	90%*
3. 피처 엔지니어링	Python(pandas, numpy), Scikit-learn	반복지출 여부·부서별 평균 편차 등 지출 패턴 변수 생성	80%*
4. 이상거래 탐지 (비지도 학습)	Isolation Forest, K-means, Scikit-learn, Python(pandas, numpy), 통계기반 스코어링 모듈	거래별 이상스코어 도출 및 고 위험군 식별 두 모델 점수 결합하여 이상징후 종합 스코어 생성	85%*
6. 내부 검증	Python 리포트 출력 (Excel)	이상거래 후보건의 타당성 검증 및 개선 의견 취합	진행 예정
7. 지도학습	random forest, XGBoost 등	비지도 학습 모델 고도화를 위해 데이터 라벨링 등 진행	향후 진행
* 모델 핵심 알고리즘, 학습구조, 스코어 산출로직은 완성되었으나 실무 적용을 위해선 보정 업무 진행 필요			

1) 모델개발 : 구현화면

[illegible]

2) 웹 앱 개발 : 구현화면

[illegible]

<p>내부 검증결과 (검증의견)</p>	<ul style="list-style-type: none"> ○ 잘 작동하는 부분 <ul style="list-style-type: none"> - 반복 지출 대비 편차가 큰 건에 대해 일관된 신호(스코어) 제공 - “부서 평균 대비 특이 지출”, “소액 반복거래의 누적 패턴” 등 내부통제에서 찾기 어려운 지점을 효과적으로 탐지 - K-means 기반 클러스터링이 특정 부서의 지출 성향 차이를 잘 드러냄 ○ 보완이 필요한 부분 <ul style="list-style-type: none"> - 라벨링된 실제 부정·오류사례 부족 → 비지도학습 정확도 한계 - 일부 특수지출(행사비·사업비 등)은 문서형 정보까지 결합해야 탐지가 정교해짐 - 예산 항목별 성격 차이에 따른 가중치 설정 필요 ○ 종합 검증 의견 <ul style="list-style-type: none"> - “비지도 학습 모델만으로도 의미 있는 이상거래 시그널을 제공하고 있으나, 지출 성격별 정교화 및 지도학습 결합 시 실효성이 크게 향상될 것으로 판단됨. <p>내부통제 절차와 결합할 경우 실무자의 점검 부담을 실질적으로 줄일 수 있음.”</p>
<p>기대효과</p>	<ol style="list-style-type: none"> 1) 내부통제 강화 <ul style="list-style-type: none"> - 지출 발생 시 즉시 이상 신호 제공 → 사전 예방 중심의 통제 가능 - 실무자의 주관적 판단 의존도 감소 2) 업무 효율성 향상 <ul style="list-style-type: none"> - 연도별 지출 비교, 부서별 패턴 분석 등 반복 업무 자동화 - 기존 수작업 분석 대비 약 60~70% 시간 절감 3) 데이터 기반 의사결정 활성화 <ul style="list-style-type: none"> - 보고서 자동 생성 - 예산 누수·이상징후를 시각적으로 파악 → 빠른 조치
<p>향후계획</p>	<ol style="list-style-type: none"> 1) 지도학습 기반 정교 모델 개발 <ul style="list-style-type: none"> - 과거 오류사례·부정거래 라벨링 - 랜덤포레스트, XGBoost 등 고도화 모델 적용 - 신뢰도 기반 알림 시스템 구축 2) 부서별 예산 모니터링 자동 리포트 <ul style="list-style-type: none"> - 월간 위험지출 Top10 - 부서별 예산 사용률 및 편차 리포트 자동 생성 3) 데이터 관리 <ul style="list-style-type: none"> - 점검 이력 저장 및 감사보고용 데이터 자동 축적 - 담당자 검증 절차(R&R) 반영