

내부통제 Digitalization

| 보이스피싱, 자금세탁방지, 상시감사

2025. 4. 17
금융AI1센터 팀장 김희선

Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

IV. 자금세탁방지(STR)

V. 상시감사

VI. 내부통제 Digitalization “추진방향”

Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

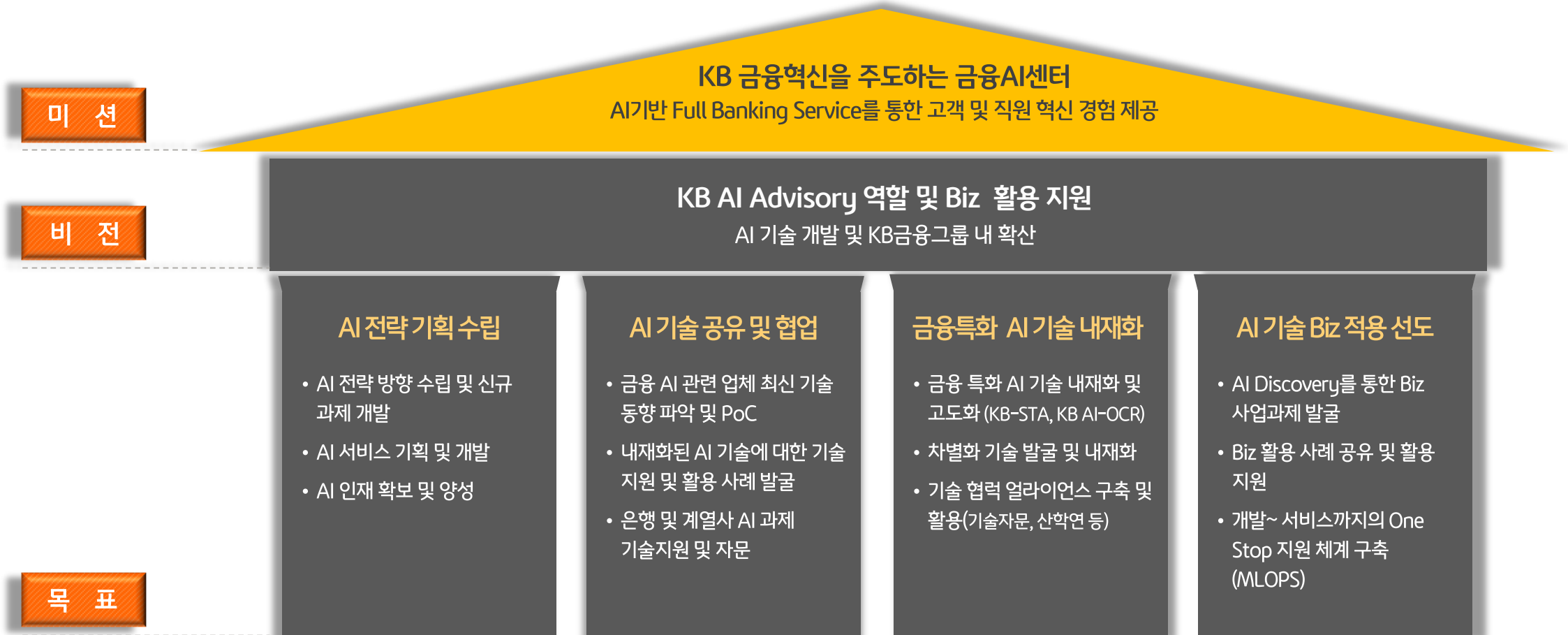
IV. 자금세탁방지(STR)

V. 상시감사

VI. 내부통제 Digitalization “추진방향”

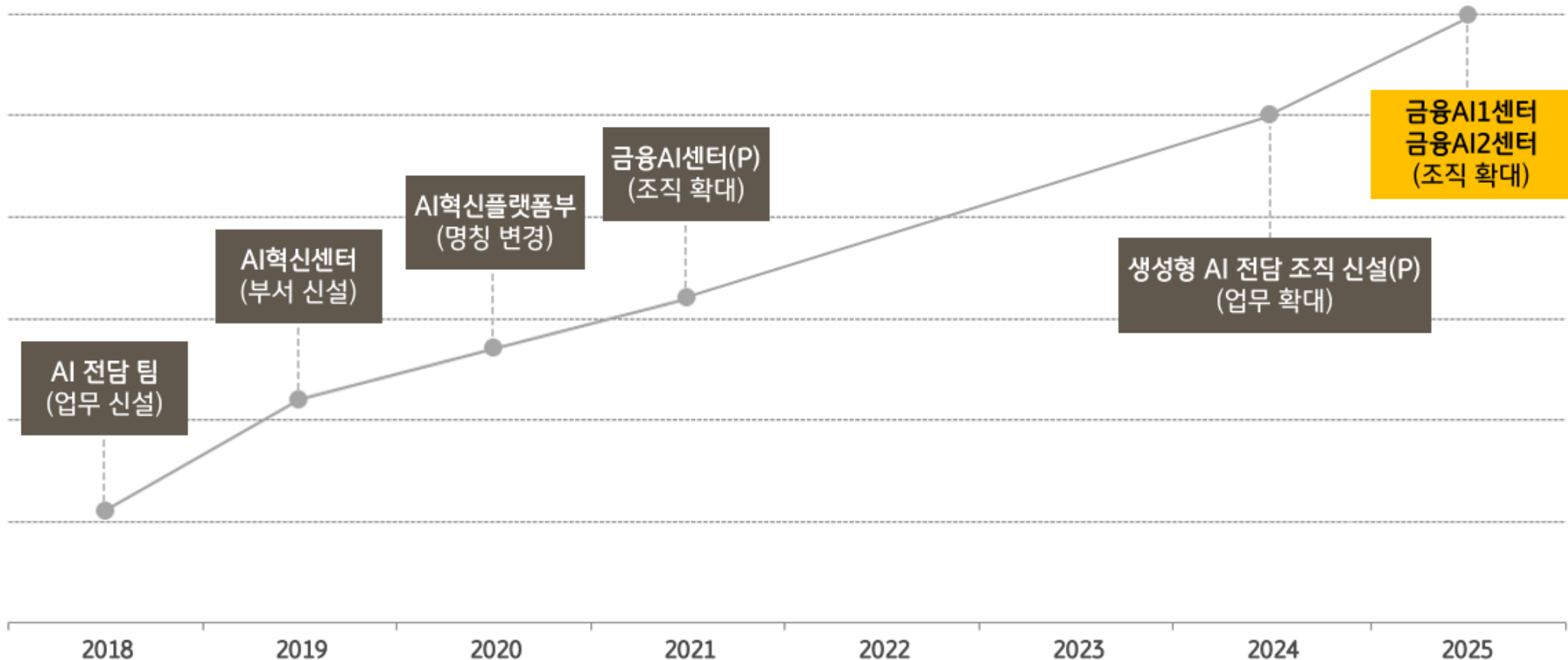
KB 금융혁신을 주도하는 금융AI센터

- 미션 | AI기반 Full Banking Service를 통한 고객 및 직원 혁신 경험 제공
- 비전 | KB AI Advisory 역할 및 Biz 활용 지원
- 목표 | ① AI 전략 기획 수립, ② AI 기술 공유 및 협업, ③ 금융특화 AI 기술 내재화, ④ AI 기술 Biz 적용 선도



금융AI센터 연혁

- 2018년 AI 전담팀으로 시작하여, 현재 AI 모델 개발, 프로세스 혁신, 신규 서비스 창출 등 KB의 디지털 혁신을 이끄는 핵심 부서로 성장
- 급변하는 AI 기술 환경과 다양한 현업 부서의 과제에 대응하기 위해 지속적으로 조직을 확대해 나갈 예정



Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

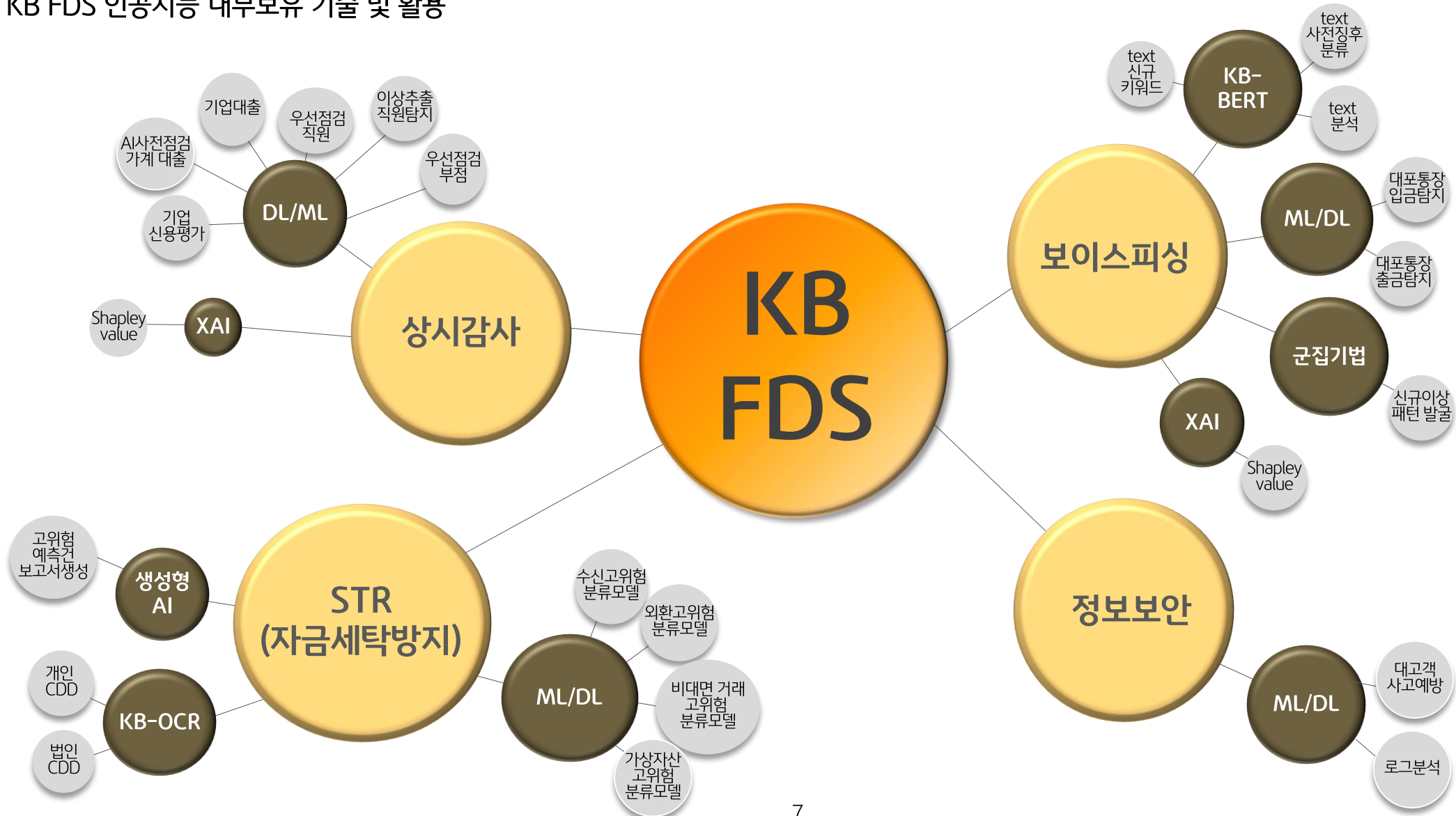
IV. 자금세탁방지(STR)

V. 상시감사

VI. 내부통제 Digitalization “추진방향”

KB FDS AI Technology Universe

• KB FDS 인공지능 내부보유 기술 및 활용



KB 내부통제 Digitalization 핵심 Factor

- KB FDS는 AI기술 내재화로 100% 자체개발
- 지속적인 연구개발 및 PoC 통해 혁신적인 외부 AI기술 도입
- 내부데이터 가공, 모델링, 운영 아키텍처 구축 등 종합적인 역량을 갖춘 Data Scientist 역할 수행

전문성 강화

- AI 기술과 금융 융합 석박사 전문가 집단으로 구성
- 지속적인 연구개발과 POC를 통한 혁신적 AI기술 도입 및 사용
- FDS AI기술 선도와 경쟁우위 확보 위해 전문성 강화



KB FDS 기술 내재화

- KB-FDS는 내부기술력으로 100% 개발된 AI-Digitalization
- 이상거래 탐지 현업노하우 노출방지
- AI성능 모니터링을 통한 지속적 고도화



All-Around Data scientist

- FDS 특화 feature engineering 기술
- 데이터 원천 추출 및 가공 기술
- 데이터 마트 구성 및 feature 구현
- 모델링 및 AI 모델 운영 자동화 구축



KB 내부통제 Digitalization 현황

- 현업 도메인 지식과 내부 데이터를 반영한 맞춤형 Feature engineering 기술로 AI모델 성능 극대화 및 금융사고 예방에 기여
- 다양한 AI 기술 생성형 AI, 자연어 처리, 머신러닝 등 을 연속적 업무흐름에 따라 적용하여 중장기적 모델 개발가능
- XAI 결과를 현업 친화적인 용어로 변경하여 모델 설명력을 강화

KB FDS 분류



I 탐지성능 향상통해 금융사고 예방

- AI기술 접목된 FDS는 탐지 정확도 높이고 오탐을 줄이는데 기여
- 지속적인 성능 모니터링과 고도화를 통해 신규 패턴 학습하여 금융 사고 예방

II 업무 효율화

- 업무의 필요한 부분에 다양한 AI기술 적용하여 활용도 높은 모델 개발
- 생성형 AI, 자연어 처리, 머신러닝 등 AI기술을 복합적으로 활용해 업무 프로세스 최적화 확대

III 설명력 강화

- XAI를 통해 AI모델의 이상거래 탐지 근거 명확히 제시
- XAI결과를 이상패턴 유형별 recategorizing하고 고객 친화적 용어로 변경하여 모니터링 담당자 이해를 도움

Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

IV. 자금세탁방지(STR)

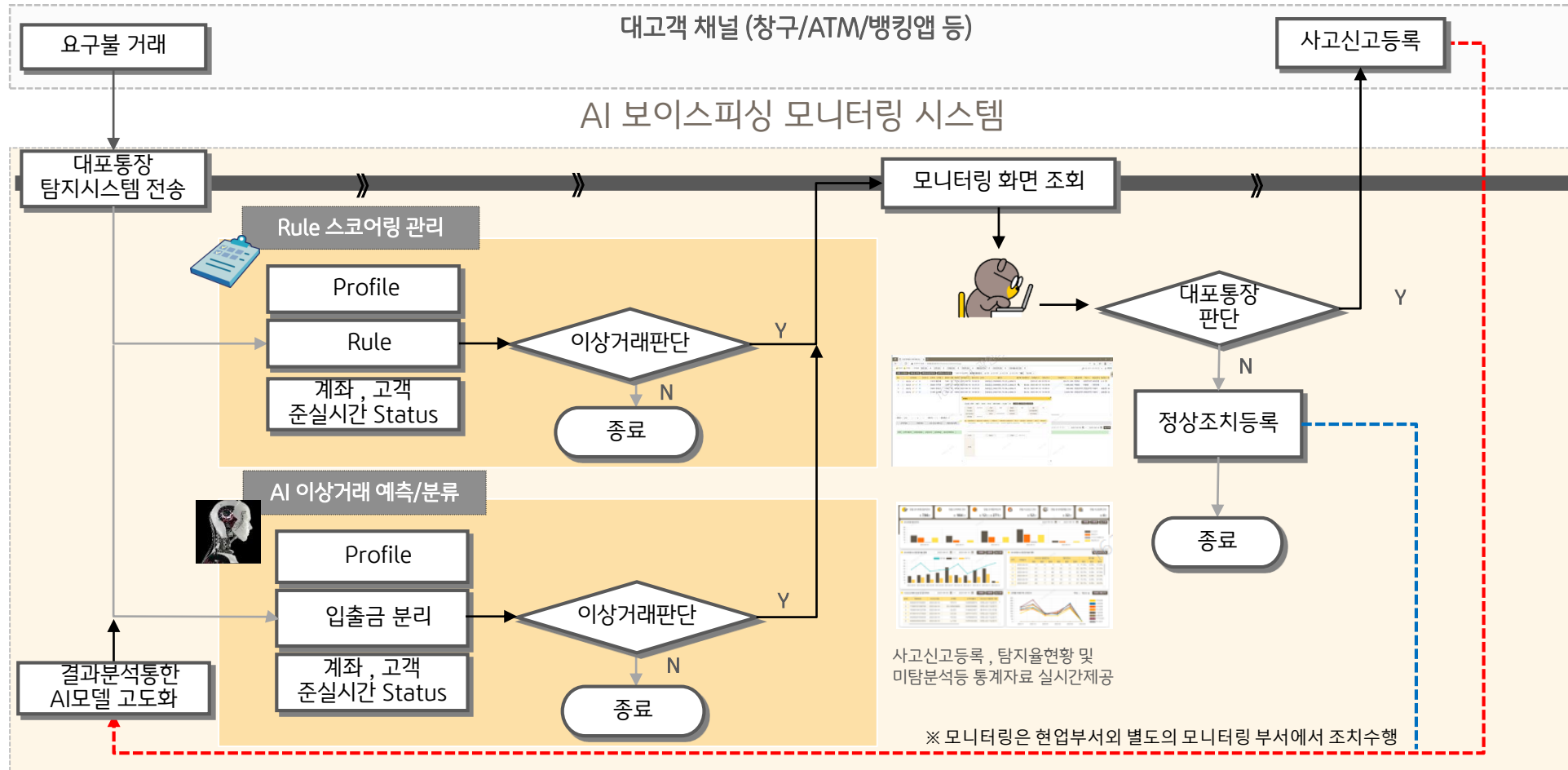
V. 상시감사

VI. 내부통제 Digitalization “추진방향”

업무 프로세스

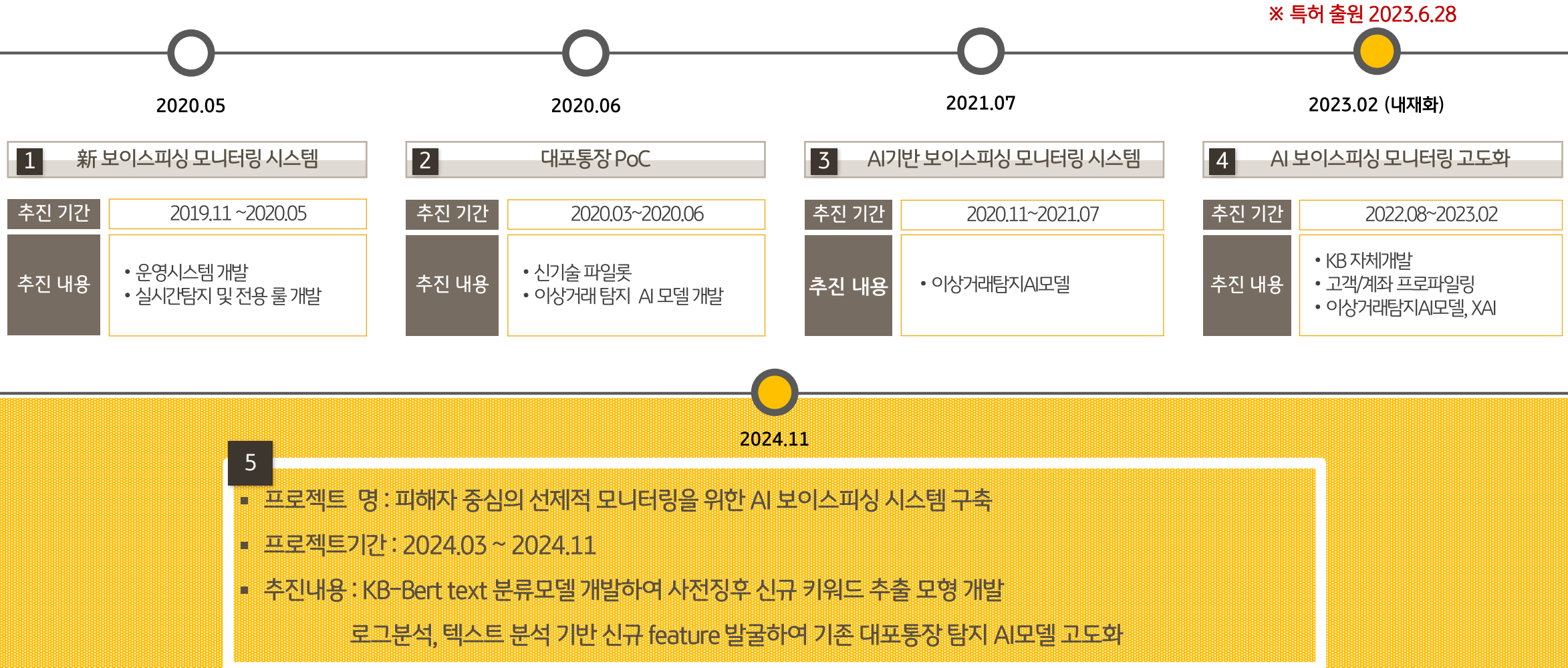
- 실시간 요구불 거래내역 기반으로 Rule과 AI 모델이 이상거래 여부 조기 탐지(예측)
- 이상거래 판단건은 모니터링 담당직원이 수기분석하여 대포통장여부 판단후 사고신고 등록 처리

※ 특허 등록(2025.2.3) : 제 10-2763990 호 출원번호 제 10-2023-0083531호 (출원일 2023.6.28)



프로젝트 추진 이력

다양한 수법으로 진화하는 금융사기에 “신속한 대응”과 실시간 대포통장 탐지로 “실질적 예방”을 목표로 AI 보이스피싱 모니터링 시스템 고도화 프로젝트를 지속적으로 추진



AI 기술

- 대포통장거래특성과 현업 도메인 배경의 맞춤 feature engineering 및 시나리오 검증 통해 최적의 알고리즘 선정, 모델학습 및 성능 최적화 진행
- KB-BERT기반 사전징후 text 분류 모델을 통해 신규 패턴 발굴 및 신규 룰 개발 Insight 제공
- XAI 모델의 결과를 금융용어 기반으로 재해석하여 설명력 강화

기술 개요

AI 보이스피싱 입출금 탐지 모델

AI 모델을 통한 사전 탐지 및 이상거래 예측

- 대포통장특성과 모니터링 노하우를 반영한 feature engineering기술로 신규지표발굴
- NLP, 로그분석 통해 feature 확장(KB-Bert)
- 데이터 불균형 처리 (SMOT, W-GAN, CT-GAN)
- 입출금 모델 개발 (LGBM, CATboost, DNN, TabNet 등)

목적

AS-IS 모델보다 Rule 미탐 탐지를 높이고, 탐지율(Precision) 향상시키는 것이 목표였음.

▶ 운영시 Precision: Max 13%

KB-Bert 기반 신규 사전징후 발굴

사전징후 신규 키워드 추출 모형 개발

- DB-SCAN Clustering Algorithm을 활용해 대포통장 탐지 이전의 text데이터를 유사한 속성으로 군집화
- 다양한 시나리오 검증으로 최적 학습데이터 구성 방안 도출
- Multiclass와 class별 binary Text 분류 모델을 비교 분석한 결과, Multi-Class Text 분류모델 선택 (KB-BERT, KoBert, TF-IDF+SVM)

1 Clustering 통해 사전이상징후 정의

2 KB-BERT Text 분류 모델 개발

3 키워드 추출

XAI

XAI통한 모델 설명력 강화

- AI feature 특성을 금융용어 기반으로 재해석하여 현업의 이해 용이
- AI 모델에 대해 XAI 구현하여 탐지된 거래에 대해 feature의 positive/negative effect(shapley value) 예측하여 ranking화함
- XAI통해 도출된 중요도 높은 feature의 카테고리 내용을 화면에 표시하여 AI 모델 탐지 근거 제시

① AI feature 특성별로 카테고리화

② XAI feature Shapley value 예측

③ 중요 feature 카테고리 화면 표시



Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

IV. 자금세탁방지(STR)

V. 상시감사

VI. 내부통제 Digitalization “추진방향”

업무 프로세스

- STR Rule에서 탐지된 의심거래건을 고위험과 중·저위험 거래로 분류
- 지속적으로 신규 AI 적용 범위를 확대하여 본부직원의 업무 효율성 증대

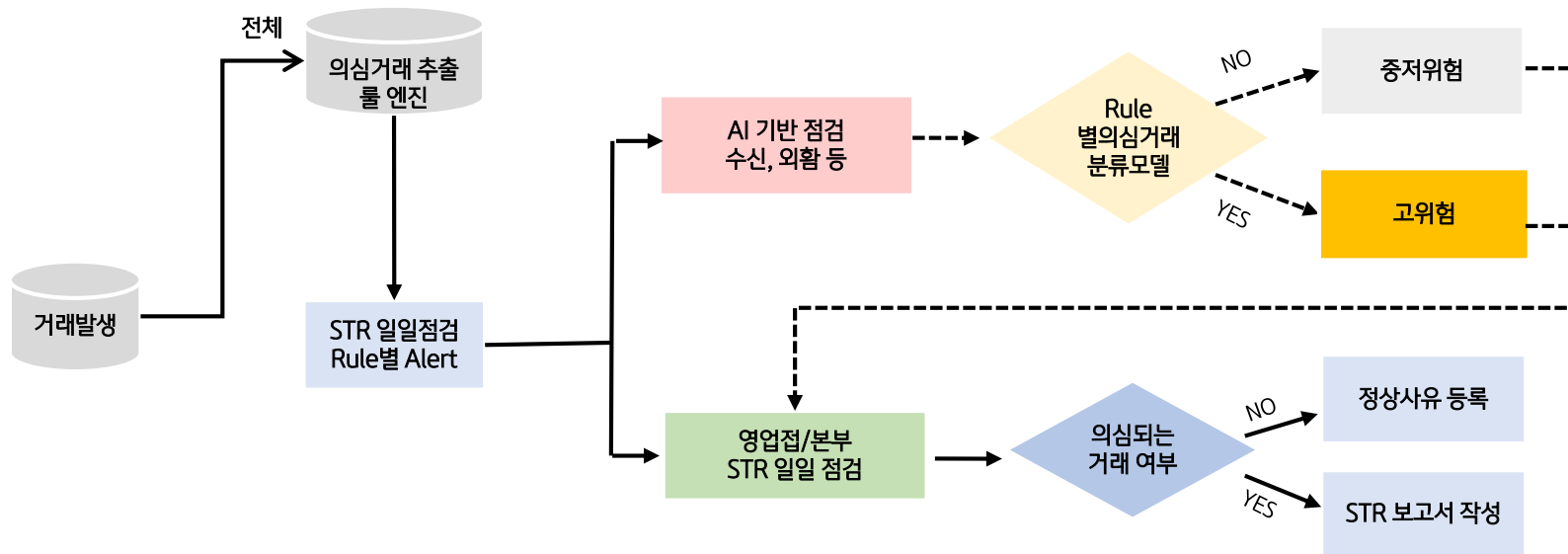
STR 이란

「STR : Suspicious Transaction Report (의심되는거래보고, 이하 ‘STR’)」
불법재산이라고 의심되는 등 자금세탁행위를 하고 있다고 의심되는 합당한 근거가 있는 거래를 금융정보분석원에 보고하는 제도

거래발생

STR 일일 점검

본부 심의 및 FIU 보고



STR 보고서 분석·심의 및 FIU 보고

자금세탁방지부 심의자

AI기술

- 수신, 외환 대상 자금세탁 의심 고위험 자동 분류 모델 개발(LGBM)
- AI 분류모델의 고위험 예측건에 대해 생성형 AI기반 보고서 생성하여 'STR 일일점검' 보고서 초안 제공

기술개요

고위험 자동 분류 모델

AI 모델을 통한 자금세탁 의심 고위험 예측

- 수신, 외환 대상 고위험 분류 모델 개발(LGBM)
- 거래 주체(개인,법인)와 추출 단위(고객, 계좌)에 따라 4 그룹으로 분리하여 단계별로 모델링 진행
- 그룹별 룰특성 반영한 feature engineering 수행
- feature selection 집중적으로 수행 (permutation importance, XAI, feature importance, correlation check등)

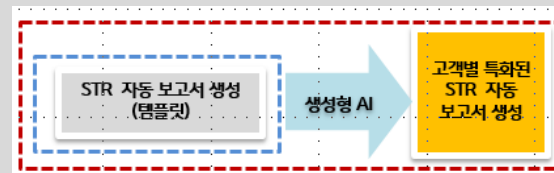
개인/고객 ↔ 개인/계좌 ↔ 법인/고객 ↔ 법인/계좌

- ▶ 고위험 예측 임계치는 오탐을 최소화하는 범위로 선택
- ▶ 임계치 선정후 Precision이 100%에 가까움

생성형 AI기반 보고서 생성

LLM 기반 보고서 생성하여 초안 제공

- 다양한 LLM 모델을 사용
- LLM을 통한 예시문장 증강
- Hallucination 방지위해 case 세분화하여 샘플 및 prompt 정교화
- 데이터 정합성 검증 agent 추가



I 고위험 자동 분류 모델

- 외환, 수신 대상 고위험 분류 모델 개발
- feature selection에 집중 (permutation importance, XAI, feature importance, correlation 등)

II 생성형 AI기반 보고서 생성

- 다양한 LLM 모델 사용
- LLM 사용해서 샘플 증강작업
- 데이터 정합성 검증 Agent 개발
- Hallucination 방지를 위한 케이스 세분화하여 샘플 및 prompt 정교화

Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

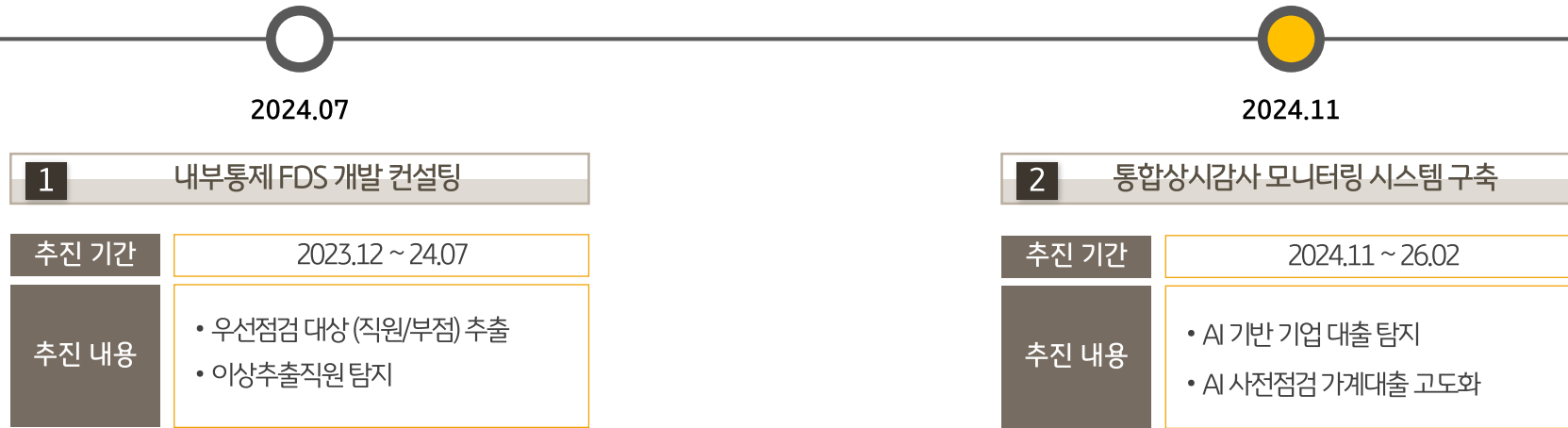
IV. 자금세탁방지(STR)

V. 상시감사

VI. 내부통제 Digitalization “추진방향”

프로젝트 추진 이력

- AS-IS 상시감사 시스템 진단하여 사고 위험의 조기 예방을 위한 디지털 기반 내부통제 FDS 개발 컨설팅 프로젝트 추진
- AI기술로 금융 사고의 조기탐지 및 모니터링 업무를 지원할 수 있는 통합상시감사 모니터링 시스템 구축 프로젝트 진행



“AI기반 금융사고 조기 통제“

AI기술을 접목하여 모니터링 업무 효율화

금융사고 발생시 AI기반 신속한 대응 및 의사결정을 통한 조기 탐지 역량 강화

AI 기술

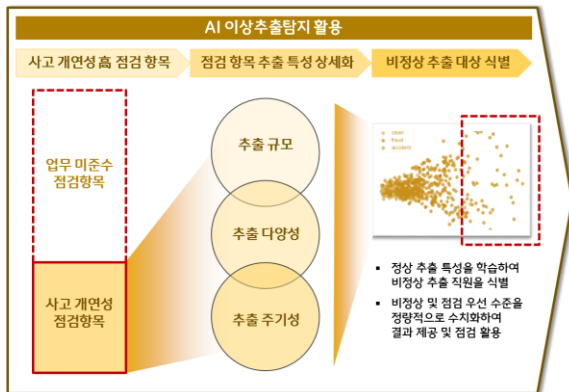
- Supervising/Unsupervising learning 기법으로 우선점검 대상제공 및 이상추출 직원 탐지하여 누적 score 순위 제공
- AI 사전점검 가계대출 모델 고도화하고 기업대출 사전점검 AI모델 개발하여 AI score와 XAI 결과 제공

기술 개요

AI기반 우선점검대상/이상추출직원탐지 모델

AI기반 우선점검대상 추출 및 이상추출직원탐지

- 부정 분류 모형 (KNN, LGBM, Decision Tree, Xgboost 등) 활용하여 우선순위 점검 대상 추출하고 XAI 결과 전달
- Unsupervising 기법을 활용해 이상추출직원 탐지(Isolation Forest, Auto Encoder)



AI사전점검 가계/기업 대출 모델

기업/가계 대출 사전 점검

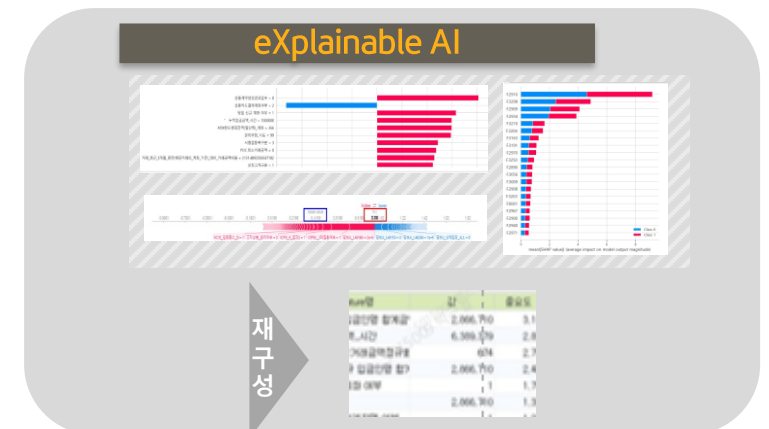
- 패턴 조합하여 AI모델 학습(supervising learning)
- 비대면거래 사전 이상징후 반영
- 패턴(룰)과 AI hybrid로 부당대출 탐지



XAI

XAI통한 모델 설명력 강화

- 이상추출직원 XAI기반 feature 영향도 표출
- XAI결과를 이상패턴 유형과 융합하여 재구성
- 대출 유형별 중요변수를 현업 친화적인 message로 변환



Contents

I. 금융AI센터소개

II. 내부통제 Digitalization

III. 보이스 피싱

IV. 자금세탁방지(STR)

V. 상시감사

VI. 내부통제 Digitalization “추진방향”

KB 내부통제 Digitalization "추진방향"

1

FDS AI Ecosystem

통합 DB구축, 상호 노하우 공유, 다양한 AI기술 결합 등 FDS 업무간 AI기술 유기적으로 연계하여 상호 통합 진화 발전 가능한 복합적 Network구축

2

AI Orchestration

ML/DL, LLM등의 AI기술을 복합적으로 활용해 거래유입부터 이상거래 탐지, 자동결과 보고서 생성까지의 전체 업무 프로세스 자동화

3

MLOps

AI 성능, feature drift, 정합성 모니터링 기반으로 ML모델이 지속적이고 안정적으로 학습/ 배포되도록 유지, 관리, 모니터링 하는 프로세스 자동화

감사합니다.

End of Document

Q&A