<실패 원인 보고서>

## FDS 프로젝트 데이터 분석 및 전처리 1차 및 2차

| 목차 | 제목 | 페이지 |
| --- | --- | --- |
| 1 | FDS 데이터 탐색 및 전처리 1차 | 2 |
| 2 | FDS 데이터 탐색 및 전처리 2차 | 3 ~ 4 |

| 과제명 | FDS 데이터 탐색 및 전처리 1차 | | |
| --- | --- | --- | --- |
| 책임자 | 김동훈 | 보고서 작성 일자 | 2024/09/05 |

| 평가 결과 및 사유 | 원본 데이터셋에 이미 자체적인 이상거래 탐지 기준을 가지고 이상 거래로 판단한 is\_fraud 컬럼이 존재하나, 해당 데이터셋이 삭제된 상태로 데이터 탐색, 분석, 및 처리가 이루어졌습니다. 또한, 위치기반 이상 거래 탐지를 하기에는 위치를 특정할 수 있는 데이터 컬럼이 삭제되어 있습니다. 이로인하여 현재 진행하는 프로젝트의 테스트 시나리오 기준이 임의적으로 거래 데이터를 이상거래로 판단하기에는 한계가 존재합니다. |
| --- | --- |
| 실패 사유 | 이상거래 탐지를 위한 데이터 분석 및 전처리 과정에서 사용한 데이터셋의 컬럼 중 생략된 부분이 존재하고, 원본 데이터셋이 자체적인 이상거래를 분류하는 기준이 존재하여 해당 기준을 무시한 채 프로젝트 자체적인 이상거래 탐지 기준을 적용하기에 부적합 합니다. 뿐만 아니라, 삭제된 데이터들의 존재로 명확한 상관 관계 분석이 용이하지 않아 이상 거래를 판단하는 임계치 설정에 어려움이 존재합니다. |
| 향후 계획 | 1. 데이터셋 수정 2. 데이터셋 변경 3. 테스트 시나리오 변경 4. 데이터간 상관 관계 분석 후 자체적인 이상 거래 탐지 기준 적용 |
| 비고 |  |

| 과제명 | FDS 데이터 탐색 및 전처리 2차 | | |
| --- | --- | --- | --- |
| 책임자 | 김동훈 | 보고서 작성 일자 | 2024/09/06 |

| 평가 결과 및 사유 | 낮은 정확도: 전체 예측 중 정확하게 예측한 비율이 27%에 불과합니다.  매우 낮은 정밀도: 사기 거래로 예측한 것 중 실제 사기 거래인 비율이 0.3%에 불과합니다.  재현율: 실제 사기 거래 중 모델이 사기 거래로 예측한 비율이 73%입니다. 하지만, 정상 거래도 사기 거래로 오탐하는 경우가 많습니다.  매우 낮은 F1-score: 정밀도와 재현율의 조화 평균이 매우 낮다는 의미로, 모델의 전반적인 성능이 좋지 않다는 것을 의미합니다. |
| --- | --- |
| 실패 사유 | 기존 FDS 프로젝트의 테스트 시나리오 목표 중 하나인 거리 기반(위도/경도) 이상 거래 탐지 시나리오를 적용하기에는 해당 데이터셋이 적합하지 않습니다. |
| 향후 계획 | 1. 데이터 셋 수정 또는 변경 2. 위도/경도 임계값(기준점) 수정 3. 거리 기반 단위 변경 (위도/경도 -> km) 4. 변수 추가(거래 금액, 시간, 요일 등을 추가) 5. 데이터 뷸균형 문제 처리: 오버샘플링 또는 언더 샘플링 기법을 적용 6. 이상치 처리 알고리즘 변경  * Isolation Forest: 이상치를 격리하는 방식으로 탐지하는 머신러닝 기반 알고리즘 * One-Class SVM: 정상 데이터를 학습하고, 학습 데이터와 다른 패턴을 가진 데이터를 이상치로 탐지하는 머신러닝 기반 알고리즘 |
| 비고 | 혼동 행렬:  [[235147 637595]  [ 813 2190]]  정확도: 0.2710115387470097  정밀도: 0.0034230249224348804  재현율: 0.7292707292707292  F1-score: 0.00681406622401165 |