

효과적인 신용카드 이상거래 탐지 프로세스 구축을 위한 비교실험

(Comparative Study for Establishing an Effective Credit Card Fraud Detection Process)

Team JH – 정보융합학부 송준현, 정보융합학부 김주혁

요약 신용카드의 사용이 늘며 방대한 양의 신용카드 결제 데이터가 생성되고 있으며, 신용카드 사용량 증가에 따라 이상 거래에 의한 피해가 증가하고 있다. 이상 거래에 의한 소비자의 피해를 최소화하기 위해 이상 거래를 신속하게 탐지하는 프로세스를 구축하는 것이 중요하다. 생성되는 결제 데이터에는 극소수의 비율만이 이상 결제에 해당하는 극심한 클래스 불균형 문제와 데이터 공간상에서 대부분의 이상 거래와 정상 거래의 구분이 매우 어려운 클래스 중복 문제가 존재하여 이상 결제 탐지에 어려움이 있다. 본 논문에서는 이와 같은 문제상황에 적합한 이상 거래 탐지 모델을 정의하고, 이를 해결하기 위해 여러 이상 탐지 기법들과 트리 기반의 이진 분류 모델을 공정한 조건에서 비교한다. 특히, 오버샘플링 시행 후 XGBoost 분류기를 활용하여 이상거래를 탐지하는 것이 해당 문제상황 내에서 가장 뛰어난 이상 거래 탐지 방법임을 제안한다. 또한, SHAP value 와 XGBoost 의 Feature Importance 기능을 활용하여 각각의 방식으로 추출한 Global Explanation 의 신뢰도를 검증하는 Ablation Study 를 진행한다. 본 연구에서 진행한 모든 실험은 [1]에서 확인이 가능하다.

Key words: 신용카드 이상결제 탐지, OverSampling, XGBoost, AutoEncoder, LOF, OCSVM, Isolation Forest, comparative analysis, Shapley additive explanations, Ablation Study

1. 서론

신용카드는 1920년대 처음 도입된 이후로 현금과 함께 대표적인 결제 수단으로 자리 잡았다. 한국은행이 조사한 <2021년 지급수단 및 모바일금융서비스 이용행태 조사>에 따르면 지급수단 별 이용 비중 통계(금액 기준)에서 신용카드와 체크/직불카드가 각각 49.5%와 16.9%로 1위와 2위를 차지하였고, 지급수단 별 종합만족도 평가에서는 두 지급수단이 모두 편리성에서 가장 높은 점수를 얻었다. 이처럼 카드 결제는 소비자들에게 높은 편의성을 제공해 주었지만, 이에 대한 부작용으로 카드 위/변조에 따른 수많은 이상 거래 피해가 발생하기도 하였다. 이러한 이상 거래를 탐지하고 예방하는 것은 현재 금융권의 중요한 이슈이며, 이상 탐지 분야에서도 활발히 연구되고 있는 주제이다.

이상 거래를 실시간으로 탐지하고 원인을 파악하여 소비자들의 피해를 줄일 수 있는 프로세스를 구축하는 것은 필수적이나, 신용카드 거래 데이터가 가지고 있는 후술할 여러 특징으로 인해 적절한 이상 탐지 모델을 구축하는데 다소 어려움이 있다. 첫째, 신용카드 데이터는 실시간으로 방대한 양의 데이터가 생성되며, 이상 거래의 피해를 최소화하기 위해 이상 탐지는 최대한 빠르게 수행되어야 한다. 둘째, 실시간으로 생성되는 방대한 거래 데이터 중 이상 거래는 극히 일부분에 해당하므로, 극심한 클래스 불균형 문제가 존재한다. 셋째, 다른 도메인의 이상 탐지 문제와 달리, 이상 거래의 경우 거래 주체자는 거래 내역이 이상 거래로 탐지되는 것을 인위적으로 회피하며, 결과적으로 이상 거래가 정상 거래와 구분이 어려운 클래스 중복 문제를 야기한다 [2].

위와 같은 신용카드 이상 거래 탐지의 특성을 고려했을 때, 좋은 이상 거래 탐지 모델의 조건을 아래와 같이 정의할 수 있다. 첫째, 방대한 양의 거래 데이터 중 이상 거래를 실시간으로 탐지하기 위하여 모델의 예측 소요 시간이 짧아야 한다. 둘째, 모델의 Precision 과 Recall 의 성능이 모두 우수해야 한다. Precision 이 지나치게 낮으면 많은 정상 거래가 이상 거래로 탐지되기에 소비자들에게 불편을 야기할 수 있고, 반면 지나치게 낮은 Recall 은 실제 이상 거래를 놓칠 수 있다. 셋째, 이상 탐지 모델을 해석 모델로 활용하여 이상 거래의 원인을 파악하는 데 이용하기 위해서는 예측력뿐만 아니라 분산이 낮은 안정적인 이상 탐지 모델을 구축해야 한다. 넷째, 실제 산업환경에 맞는 유연한 이상 탐지 프로세스를 구축하기 위해서는 상황에 맞는 임계값을 설정할 수 있어야 한다. 예를 들어, 결제 금액이 높은 거래의 이상 거래 여부를 판별하는 상황에서는 상대적으로 높은 비율의 False Alarm 을 감수하더라도 임계값을 낮추는 것이 적절하다.

기존에도 지도와 비지도 학습 관점에서 해당 문제 상황을 해결하는 데 적합한 이상 탐지 모델을 찾기 위한 비교 실험을 진행하거나, 새로운 아키텍처를 제안한 연구들이 존재하였다. 그러나 해당 연구들은 테스트

데이터셋의 비율을 임의로 조정하여 실제 현장의 극심한 클래스 불균형 문제상황을 제대로 반영하지 못해 활용성이 떨어진다는 한계가 존재한다 [3, 4].

따라서, 본 연구에서는 공정한 비교를 위해 여러 이상 탐지 모델 및 분류 모델을 실제 산업의 문제상황과 최대한 유사한 조건에서 반복적으로 비교하는 실험을 진행하였다. 실험 결과와 더불어 앞서 정의했던 좋은 이상 거래 모델의 조건을 종합적으로 고려하여, 해당 문제상황에서 최선의 이상 거래 탐지 모델을 제안한다. 또한, 모델 불특정 기법(model agnostic method)인 SHAP(Shapley Additive exPlanations)와 XGBoost 의 내장 기능인 Feature Importance 추출 기능을 활용하여 주요 변수를 추출하고, 이때 추출된 변수들의 해석능력을 비교 및 검증하는 Ablation Study 를 진행한다 [5].

2. 관련 연구

2.1 Credit Card Fraud Detection Techniques

심각한 클래스 불균형 문제가 존재할 때, 다수의 정상 데이터와 소수의 이상 데이터를 구분하는 방식은 전처리 방식에서 데이터를 조정하는 샘플링 기법(Sampling method)과 모델링 과정에서 알고리즘을 변경하는 단일 범주 분류기법(Novelty detection) 등으로 나눌 수 있다. 각각의 예로는 SMOTE, Over/Undersampling 과 AutoEncoder, LOF, OCSVM, Isolation Forest 등이 있다 [6-9]. 이상 거래 탐지를 위해 샘플링 기법을 시행하고 트리 기반의 이진 분류 모델을 이용하는 방법은 꾸준히 연구되어 왔으며, 그 중에서도 정형 데이터에서 뛰어난 성능을 보여주는 XGBoost(Chen et al) 알고리즘을 이용한 연구들이 활발히 진행되었다 [10,11].

2.2 Isolation Forest

Isolation Forest 는 결정 트리 기반의 이상 탐지 알고리즘이다. 주어진 데이터 샘플을 고립시켜 가며, 각 데이터 샘플의 고립 정도 여부에 따라 이상치를 판별한다. Information gain 을 최대화하는 결정 트리과 달리, Isolation Forest 는 특성을 무작위로 선택한 후, 특성의 최솟값과 최댓값 사이의 값을 무작위로 선택하여 샘플을 분할한다.

직관적으로 생각했을 때 정상 샘플은 밀집된 클러스터를 형성하며, 데이터를 고립시키기까지 여러 번의 분할이 필요하다. 반대로, 이상 샘플의 경우 더 적은 수의 분할로 샘플을 고립시킬 수 있다. 즉, 이상 샘플은 트리의 root node 와 근접하게 되고, root 부터의 경로 길이가 짧을 수록 anomaly score 를 높게 측정한다.

2.3 Ablation Study

Ablation Study는 주로 의학이나 심리학에서 주로 사용되는 연구 방법으로, 유기체의 장기 혹은 특정 조직을 제거한 뒤, 해당 유기체의 행동을 관찰함으로써 제거된 장기 혹은 특정 조직의 역할을 실험하는 방식으로 이루어진다. 이러한 연구는 의학이나 심리학에서 뿐만 아니라, 머신러닝 분야에서도 제안하는 시스템의 요소가 결과에 미치는 인과관계를 추론하는데 유용하여 널리 활용되고 있다 [5].

본 논문에서는 [5]에서 제시한 Ablation study의 정의를 바탕으로 다양한 방식으로 추출한 주요 변수들을 제거하여 이에 따른 결과차이를 비교하는 실험을 진행한다. 이를 통해 해당 문제상황 내에서 가장 신뢰도가 높은 Global Explanation 제공 방식과 중요도가 높은 변수들을 확인한다.

3. 제안 방법론 및 실험설정

3.1 방법론 개요

본 연구에서 진행한 이상 거래 탐지모델 비교 실험은 크게 두 단계로 나눌 수 있다. 첫째로, 여러 이상 탐지 모델과 오버샘플링을 진행한 데이터로 훈련된 XGBoost 이진 분류 모델의 이상 거래 탐지 성능을 알아보는 비교실험을 진행한다. 각 모델의 안정성을 확인하기 위해 50 회씩 반복 실험을 진행하였고, 테스트 데이터의 정상 데이터와 이상 데이터의 비율을 1:1 부터 원본 데이터의 극심한 불균형 상황까지 점진적으로 늘려가며 지표를 측정하였다. 또한, XGBoost 분류기의 수렴을 위한 최소 이상치 레이블의 개수를 확인하는 추가 실험도 진행한다.

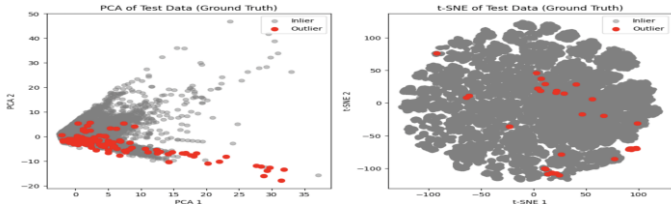
둘째로, SHAP과 XGBoost의 자체 해석 기능인 Feature Importance 추출을 통해 각각의 방식에서 제공하는 Global Importance의 신뢰성을 확인하는 Ablation Study를 진행한다. 구체적으로는, 각각의 방법으로 Global Feature Importance를 추출한 뒤, 상위 8개의 주요 특성을 제거하고 50회의 이상 거래 탐지 실험을 진행하여 이에 따른 지표변화를 확인한다. 데이터셋 내에 심각한 클래스 불균형 문제가 존재하므로 두 실험 모두 주요 지표로 F1 Score를 사용하고, 보조지표로 AUROC와 AUPRC를 사용한다. 신뢰성 있는 예측 소요 시간 측정을 위해 실험 환경을 Intel(R) Core(TM) i5-9400F CPU @ 2.90GHz, 16GB DDR RAM으로 통일하였다.

3.2 데이터 설명 및 전처리 과정

이상 거래 탐지에 활용한 데이터는 2013년 9월, 이를 동안 유럽에서 수집한 신용카드 거래 데이터이다. 한 개의 클래스 레이블 변수 외에 총 30개의 독립변수로 이루어져 있으며, 이 중 거래시간을 의미하는 "Time"과 거래 금액을 의미하는 "Amount" 2개 변수를 제외한 나머지 28개의 변수에는 개인 정보 보호를 위해 주성분 분석이 적용되었다. 각 변수의 분포를 살펴보면 V4, V9, V11, V13, V18, V22, V26 총 7개의 변수를 제외한 대부분의 변수에서 왜도(skewness)가 심함을 확인할 수 있었다. 또한 클래스별 분포 차이를 확인해 본 결과, V4 변수를 제외하고는 이상 거래와 정상 거래 사이의 변수 별 분포 차이는 거의 존재하지 않았다.

해당 데이터셋은 총 284,807건의 거래 데이터로 구성되어 있다. 그중 정상 거래는 284,315건, 이상 거래는 492건으로, 정상 데이터와 이상 데이터의 비율이 약 1 : 0.00173으로 매우 극심한 클래스 불균형 문제가 존재하였다. 또한 그림 1을 통해 각각 PCA와 t-SNE로 2차원으로 투영 후 시각화를 진행하였을 때, 데이터 공간상에서 정상 거래와 이상 거래가 명확히 구분되지 않는 클래스 중복 문제가 존재함을 확인하였다.

그림 1. PCA와 t-SNE를 활용한 2차원 투영 결과



총 30개의 독립변수 중 "Time" 변수는 데이터셋의 첫 결계 데이터를 시작으로 각각의 거래가 이루어짐에 따라 경과된 시간을 의미하므로 이상 거래 탐지와는 관계가 없는 변수라고 판단하여 제거하였다. "Amount" 변수에는 StandardScaler를 적용하였으며, 나머지 변수의 경우 각 변수의 값이 0을 중심으로 분포하고 결측치가 존재하지 않았기에 별도의 전처리 과정은 수행하지 않았다.

3.3 이상탐지 모델 구축

본 실험에서 사용한 데이터는 클래스 불균형이 매우 극심한 정형 데이터이다. 이처럼 이상 데이터가 매우 적거나, 데이터의 레이블이 아예 존재하지 않는 상황에서는 학습 시 정상 데이터만을 사용하여 정상 데이터에 대한 boundary를 설정하는 One-Class Classification을 사용하는 것이 일반적이다.

이상 탐지 분류 방식은 밀도 기반, 경계선 기반, 재구축 기반, 트리 기반 방법 등으로 나눌 수 있다. 이 외에도 K-Nearest Neighborhood와 K-Means Clustering과 같은 거리 기반 이상 탐지 모델도 존재하지만, 해당 모델들은 각각 예측 시간이 오래 소요된다는 단점과 초기 군집 중심점에 따라 분석 결과의 일관성이 떨어진다는 단점이 존재하여 본 연구의 문제 상황과는 맞지 않다고 판단하여 사용하지 않았다.

본 비교실험에서는 이상 탐지 모델로 밀도 기반 알고리즘인 LOF(Local Outlier Factor), 경계선 기반 알고리즘인 OCSVM(One class Support Vector Machine), 재구축 기반 알고리즘인 AE(AutoEncoder), 마지막으로 정형 데이터에서 트리 기반 알고리즘이 보여주는 뛰어난 성능을 고려하여 트리 기반의 이상 탐지 알고리즘인 Isolation Forest를 사용하였다.

3.4 XGBoost 이상거래 분류모델

XGBoost는 결정 트리 기반의 앙상블 알고리즘으로, 일반적으로 정형데이터에서 강건한 성능을 보이는 것으로 알려져 있다. 앞서 설명한 이상 탐지 모델과는 다르게 XGBoost는 분류 알고리즘이므로, 학습 시 반드시 클래스 레이블이 필요하다. 본 실험에서는 XGBoost를 활용하여 이상거래 탐지 성능을 확인해본 뒤, 이진 분류 모델이 수렴되는데 필요한 최소 이상치 레이블의 개수를 확인해 보고자 훈련 데이터에서 이상레이블 데이터의 비율을 점점 낮추고 그에 따른 탐지결과를 비교해 보는 실험을 추가로 진행하였다.

3.5 SHAP의 Global Explanation과 XGBoost의 Feature Importance Ablation

XGBoost에서 특성 중요도를 평가하는 기준으로는 특정 특성이 예측 과정에서 데이터를 분리하는데 사용된 횟수를 기준으로 하는 'Weight', 특정 특성으로 인해 분리된 데이터 수의 평균을 기준으로 하는 'Cover', 특성이 사용되었을 때 평균적으로 감소하는 Information Gain을 기준으로 하는 'Gain' 등이 있다. 이와 같이, XGBoost의 특성 중요도 추출 기능은 평가하는 기준에 따라 다른 결과가 계산될 수 있다.

SHAP(Shapley Additive Explanations)은 게임이론에 기반하여 모델의 예측에 대한 각 특성의 기여도를 SHAPely 값을 통해 계산하는 해석 모델이다 [12]. SHAPely 값 계산 과정에서, 모든 가능한 특성 조합에 대해 평균적인 기여도를 계산하여 일관성 있는 해석 결과를 제공한다. 또한 전역 변수 중요도 뿐만 아니라, instance-wise 한 지역적인 해석 능력도 제공한다.

4. 실험결과 및 결론

4.1 평가지표

학습 모델의 성능 비교를 위해 클래스 불균형 문제를 고려할 수 있는 F1 Score, AUROC, AUPRC 평가지표를 사용한다. F1 Score는 정밀도와 재현율의 조화 평균을 의미한다. 해당 지표는 위음성과 위양성 모두를 고려하기 때문에 클래스 불균형 문제에 유용하다. AUPRC는 정밀도-재현율 AUC(Area Under Curve)를 의미하며, 마찬가지로 정밀도와 재현율 간의 균형에 중점을 둔다. F1 Score와 같이 위음성과 위양성 모두를 고려하여 클래스 불균형 문제에 유용한 지표이다. AUROC는 모델의 True Positive Rate와 False Positive Rate를 고려하며, 다양한 확률 임계값에 걸쳐 모델을 평가한다. 해당 지표는 단일 임계값을 선택하는 것이 간단하지 않은 클래스 불균형 데이터셋 평가에 유용하다.

$$F1 - score = 2 \frac{precision \times recall}{precision + recall} \quad (1)$$

$$AUROC = \int_{t=-\infty}^{\infty} \frac{tp(t)}{tp(t)+fn(t)} \frac{d}{dt} \left(\frac{fp}{fp+tn} \right) |_t dt. \quad (2)$$

$$AUPRC = \int_{t=-\infty}^{\infty} precision(t) \frac{d}{dt} (recall) |_t dt. \quad (3)$$

4.2 이상탐지 모델 및 XGBoost 이진분류기의 비교실험 결과

표 1, 표 2는 신용카드 이상 거래 탐지에 대한 모델 별 50 회 반복 실험 결과를 보여주고 있다. 1:1, 1:10 등 상대적으로 클래스 불균형이 심각하지 않은 테스트 데이터셋에서는 LOF를 제외한 나머지 모델에서 공통적으로 좋은 성능을 확인할 수 있다. 그러나 클래스 불균형이 극심한 원본 비율의 데이터셋에서는 XGBoost가 3가지 지표에서 모두 압도적으로 뛰어난 이상거래 탐지 성능을 보여주었다. 또한, 이상 탐지 기법인 OCSVM과 AE 알고리즘이 그 다음으로 뛰어난 성능을 보여주었지만, XGBoost에 비해 현저히 많은 소요 시간과 낮은 탐지 성능을 보여주었다. 표 1, 2의 1:1, 1:10, 1:50은 테스트 데이터셋의 이상:정상 샘플 비율을 의미하며, Original의 경우 정상 샘플 비율을 조정하지 않은 원본 테스트 데이터셋이다.

표 1. 이상 거래 탐지 모델의 비교 실험 결과 및 소요 시간

Test dataset (AUROC / AUPRC) - 평균					
Model	1:1	1:10	1:50	Original	Time(s)
AE	0.91/0.94	0.92/0.76	0.91/0.57	0.92/0.46	19.75
XGBoost	0.91/0.95	0.91/0.92	0.91/0.91	0.91/0.86	2.69
LOF	0.60/0.71	0.60/0.30	0.59/0.17	0.59/0.13	61.21
OCSVM	0.91/0.95	0.91/0.85	0.91/0.70	0.91/0.47	110.52
Isolation Forest	0.89/0.93	0.89/0.76	0.89/0.57	0.89/0.43	0.92

표 2. 이상 거래 탐지 모델의 비교 실험 결과 및 소요 시간

Test dataset (F1 Score) - 평균(표준편차)					
Model	1:1	1:10	1:50	Original	Time(s)
AE	0.91(0.02)	0.74(0.02)	0.40(0.02)	0.06(0.00)	19.75
XGBoost	0.90(0.02)	0.90(0.02)	0.89(0.02)	0.86(0.02)	2.69
LOF	0.38(0.06)	0.27(0.04)	0.11(0.02)	0.01(0.00)	61.21
OCSVM	0.90(0.02)	0.84(0.02)	0.68(0.03)	0.19(0.01)	110.52
Isolation Forest	0.88(0.02)	0.75(0.02)	0.44(0.03)	0.07(0.01)	0.92

그림 2. PCA와 t-SNE를 활용한 OCSVM 예측 값 2차원 투영 결과

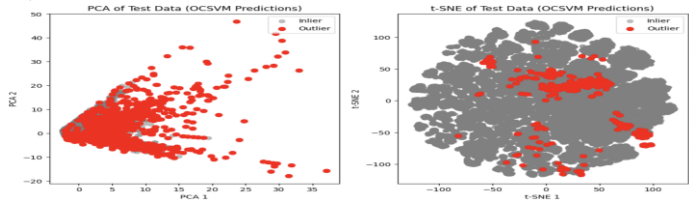


그림 3. PCA와 t-SNE를 활용한 XGBoost 예측 값 2차원 투영 결과

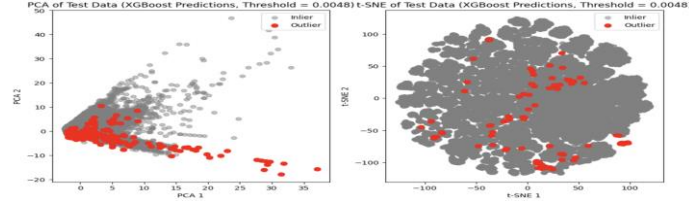


그림 2와 그림 3은 앞선 실험에서 가장 뛰어난 이상 거래 탐지 성능을 보여주었던 OCSVM과 XGBoost 두 개의 알고리즘의 예측 결과를 PCA와 t-SNE를 활용하여 2차원으로 투영한 뒤 시각화한 Plot이다. 그림 2에서 사용한 OCSVM 모델의 성능은 F1 Score와 Recall이 각각 0.19과 0.83이며, 이 예측 결과를 그림 1의 Ground Truth와 비교해 보면, 지나치게 많은 위양성이 존재하는 것을 육안으로 쉽게 확인할 수 있다. 이는 해당 이상 탐지 모델이 이상 거래 패턴을 제대로 감지하지 못하고 있음을 시사한다.

반면, 그림 3에서 사용한 XGBoost 이진 분류기의 F1 Score와 Recall은 각각 0.63과 0.90이다. 해당 모델은 임의로 임계값을 0.0048의 매우 낮은 값으로 조정하여, F1 Score의 손해를 감수하고 재현율을 높이도록 XGBoost를 튜닝한 버전이다. 마찬가지로 그림 1의 Ground Truth와 비교해보면, OCSVM 모델에 비해 높은 재현율을 가짐에도 불구하고 위음성의 비율이 훨씬 낮은 것을 육안으로 확인할 수 있다. 이를 통해 XGBoost 이진 분류기가 OCSVM과 같은 다른 이상 탐지 모델보다 이상거래 패턴 감지에 훨씬 뛰어난 성능을 보여줄 수 확인하였다.

표 3. XGBoost 수렴을 위한 레이블 임계값 실험 결과

Test dataset (F1 Score) - 평균(표준편차)				
Ratio Model	1/2 (184)	1/4 (92)	1/8 (46)	1/16(23)
XGBoost	0.86(0.02)	0.84(0.02)	0.80(0.03)	0.68(0.06)

표 3은 원본 데이터셋에서 분할한 테스트 데이터셋은 유지한 채, XGBoost 이진 분류기가 수렴되기 위해 필요한 최소 이상 거래 레이블의 수가 몇 개인지 확인하기 위해 진행한 반복 실험의 결과이다. 표 3의 1/2 (184)는 훈련 데이터의 이상 샘플을 절반으로 줄였을 때, 이상치 레이블의 개수가 184개임을 의미한다. 훈련 데이터에 속한 이상 거래 데이터의 수를 절반씩 줄여나가며 그때의 F1 Score를 확인하였다.

그 결과, 약 21만 개의 훈련 데이터 중 23개의 이상 거래 레이블만 존재하여도 F1 Score 0.68의 뛰어난 성능을 얻을 수 있음을 확인하였다. 이를 통해 XGBoost 이진 분류기를 이용하면 극소수의 이상 거래 레이블 데이터만으로도 다른 이상 탐지 알고리즘에 비해 뛰어난 이상 거래 탐지 성능을 보여준다는 것을 확인할 수 있었다.

4.3 SHAP과 Feature Importance 기반 Ablation Study

표 4. XGBoost의 Feature Importance와 SHAP의 Ablation 실험 결과

TOP 8 - Test dataset (F1 Score) - 평균(표준편차)				
Original	Gain	Cover	Weight	SHAP
0.858 (0.019)	0.841 (0.018)	0.839 (0.021)	0.829 (0.020)	0.832 (0.018)

표 4는 다양한 Global Feature Importance 방법으로 추출된 중요 변수 상위 8개를 제거하고 만들어진 별도의 데이터셋에서 XGBoost 이진 분류기로 이상 탐지 반복 실험을 50회 진행한 결과이다. 각각 SHAP 모델의 Global Importance를 이용한 데이터셋 하나와 XGBoost Feature Importance의 Gain, Cover, Weight 3가지 기준을 이용한 데이터셋 3개, 총 4개의 데이터셋을 사용하여 평가를 진행하였다.

평가 결과에 대해 분산분석을 진행한 뒤 Tukey HSD 사후 검정을 진행해 본 결과, PCA가 이미 적용된 데이터임에도 불구하고 4개의 데이터셋에서의 성능이 원본 데이터셋에서의 성능과 통계적으로 유의미한 차이가 존재하는 것을 확인했다. 또한, 4가지 경우 상위 8개 주요 변수에 모두 V14, V4, V12 3개의 변수들이 공통으로 포함되어 있어 해당 변수들은 이상 거래 탐지 상황에서 가장 중요한 변수임을 확인하였다.

5. 결론

본 논문에서는 신용카드 이상 거래 탐지 문제상황에서 존재하는 특성과 데이터셋이 가지고 있는 특징을 고려하여 좋은 이상 거래 탐지 모델이 갖춰야 할 요구 조건에 대해 정의하였다. 여러 이상 탐지 알고리즘과 오버샘플링을 진행한 데이터로 훈련된 XGBoost 모델을 비교하는 실험을 반복적으로 수행하여, 이상 탐지 알고리즘의 비수렴성과 XGBoost의 뛰어난 성능 및 적은 레이블에서도 수렴되는 강건성을 확인할 수 있었다. 앞서 정의한 좋은 이상 탐지 모델의 요구 조건과 성능 및 안정성을 종합적으로 고려하여 신용카드 이상 거래 탐지 상황에서 XGBoost 이진 분류기가 최적의 Solver임을 제안하였고, 산업 환경에서 위음성에 대한 위험도가 높은 특수 상황에서 threshold를 낮추어 활용성을 높일 수 있음을 확인하였다.

다양한 중요 변수 추출 기능의 신뢰도를 검증하기 위해, 각각의 경우의 중요 변수를 Ablation 하고 지표의 변화를 확인하였다. 이후 통계 검정을 진행하여 4가지 방식으로 제공하는 중요 변수들이 통계적으로 유의한 결과의 차이를 야기하는 것을 확인하였다. 또한 실제 문제상황과 최대한 유사한 조건에서 공정한 실험을 진행하기 위해 모든 실험을 반복적으로 시행하였으며, 임의로 비율이 설정된 데이터셋이 아닌 원본의 클래스 불균형 상황이 그대로 존재하는 테스트 데이터셋에서 진행하였다.

References

- [1] <https://github.com/Kyyle2114/Comparative-Study-of-Building-Effective-Credit-Card-Fraud-Detection-Process>
- [2] Ronaldo C. Prati, Gustavo E. A. P. A. Batista & Maria Carolina Monard, "Class Imbalances versus Class Overlapping: An Analysis of a Learning System Behavior," in *Mexican International Conference on Artificial Intelligence (MICAI'04)*, pp. 312-321, 2004.
- [3] Wu, Tung-Yu, and You-Ting Wang. "Locally interpretable one-class anomaly detection for credit card fraud detection." 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI). IEEE, 2021.
- [4] Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare. "Credit card fraud detection using machine learning techniques: A comparative analysis." 2017 international conference on computing networking and informatics (ICCI). IEEE, 2017.
- [5] MEYES, Richard, et al. Ablation studies in artificial neural networks. arXiv preprint arXiv:1901.08644, 2019.
- [6] Japkowicz, N., Myers C., & Gluck M., "A Novelty Detection Approach to Classification", in Mellish, C. (ed.) *The International Joint Conference on Artificial Intelligence (IJCAI-95)*, Montreal, Canada. IJCAI & Morgan Kaufmann. San Mateo, CA. pp 518-523., 1995.
- [7] Bernhard Schölkopf, Robert C. Williamson, Alex Smola, John Shawe-Taylor, John Platt, "Support Vector Method for Novelty Detection," *Advances in Neural Information Processing Systems 12 (NIPS 1999)*, pp. 582-588, 1999.
- [8] Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J, " LOF: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 93-104, Feb. 2000.
- [9] F. T. Liu, K. M. Ting and Z. -H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422, 2008
- [10] Priscilla, C. Victoria, and D. Padma Prabha. "Influence of optimizing XGBoost to handle class imbalance in credit card fraud detection." 2020 third international conference on smart systems and inventive technology (ICSSIT). IEEE, 2020.
- [11] Meng, Cuizhu, Li Zhou, and Bisong Liu. "A case study in credit fraud detection with SMOTE and XGboost." *Journal of Physics: Conference Series*. Vol. 1601. No. 5. IOP Publishing, 2020.
- [12] Lundberg, S. M., & Lee, S. I, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, Vol. 30, 2017.
- [13] 이용현, 구해모, and 김형주. "오토인코더를 활용한 효율적인 신용카드 사기 탐지 지도 기법." *정보과학회 컴퓨팅의 실제 논문지* 25.1 (2019): 1-8.
- [14] 방성진, 고해지, and 이상민. "TabNet 기반의 반도체 제공재고 수준의 조기에측과 해석." *정보과학회논문지* 49.6 (2022): 466-474.