

# 부스팅 알고리즘 기반 신용 카드 이상 거래 탐지

이하랑, 김신, 윤경로  
 건국대학교 스마트ICT융합공학과  
 harang8069@somewhere.ac.kr, new.xin22@gmail.com, yoonk@konkuk.ac.kr

## Credit Card Fraud Detection based on Boosting Algorithm

Lee Harang, Kim Shin, Yoon Kyoungro  
 Dept. of Smart ICT Convergence Engineering, Konkuk University

### 요 약

전자금융거래 시장이 활발해지며 이에 따라 신용 카드 이상 거래가 증가하고 있다. 따라서 많은 금융 기관은 신용 카드 이상 거래 탐지 시스템을 사용하여 신용 카드 이상 거래를 탐지하고 개인 피해를 줄이는 등 소비자를 보호하기 위해 큰 노력을 하고 있으며, 이에 따라 높은 정확도로 신용 카드 이상 거래를 탐지할 수 있는 실시간 자동화 시스템에 대한 개발이 요구되었다. 이에 본 논문에서는 머신러닝 기법 중 부스팅 알고리즘을 사용하여 더욱 정확한 신용 카드 이상 거래 탐지 시스템을 제안하고자 한다. XGBoost, LightGBM, CatBoost 부스팅 알고리즘을 사용하여 보다 정확한 신용 카드 이상 거래 탐지 시스템을 개발하였으며, 실험 결과 평균적으로 정밀도 99.95%, 재현율 99.99%, F1-스코어 99.97%를 취득하여 높은 신용 카드 이상 거래 탐지 성능을 보여주는 것을 확인하였다.

이상 거래 탐지 시스템을 제안하고자 한다.

### 1. 서론

전자금융거래 시장의 규모가 커지며 이와 함께 신용 카드 이상 거래가 증가하고 있다. 2020년 2월 한국은행이 공개한 ‘주요국의 지급수단 사기 동향 및 시사점’ 보고서[1]에 따르면 2018년 전 세계 카드 사기 금액이 278.5억 달러에 달하는 것으로 나타났다. 이 수치는 2013년 137억 달러였던 전 세계 카드 사기 규모에 비해 약 141억이나 늘어난 수치이다. 이에 많은 금융 기관들은 신용 카드 이상 거래로부터 소비자를 보호하기 위해 신용 카드 이상 거래 탐지 시스템을 사용하여 실시간으로 신용 카드 이상 거래를 분석하고 신용 카드 이상 거래의 경우, 이를 이용자에게 알리고 거래를 중단시킨다. 또한 신용 카드 이상 거래 탐지 시스템은 이상 패턴을 분석하여 새로운 신용 카드 이상 거래들에 대한 모니터링을 가능하게 한다. 이때, 큰 손실을 줄 수 있는 신용 카드 이상 거래를 정확하고 신속하게 분석 및 탐지하는 것이 중요하다.

따라서 본 논문에서는 머신러닝 기법 중 부스팅 알고리즘을 사용하여 보다 높은 이상 탐지 성능을 위한 신용 카드 이상 거래 탐지를 위한 신용 카드

### 2. 관련 연구

#### 2-1. 신용 카드 이상 거래 탐지 시스템

신용 카드 이상 거래 탐지 시스템이란 신용 카드 결제로 인한 전자 금융 거래 시 접속 정보, 거래 정보 등을 분석하여 신용 카드 이상 거래를 탐지하고 거래를 중단시키는 시스템이다. 해당 시스템은 거래가 일정한 패턴에서 벗어났을 때 신용 카드 이상 거래로 판단할 수 있으며, 이는 소비자를 보호할 수 있는 장치가 될 수 있기에 신용 카드 거래에 있어 중요한 시스템이라 할 수 있다.

#### 2-2. 부스팅 알고리즘

##### 1. XGBoost

XGBoost[2]는 과적합 방지를 위해 그래디언트 트리 부스트에 파라미터가 추가된 지도 학습 알고리즘이다. 의사결정나무를 기본 학습기로 가지며 균형 트리 분할 방식으로 트리를 생성하고 잔차를 이용해 이전 모형의 약점을 보완한다.

## 2. LightGBM

LightGBM[3]은 예측 오류 손실을 최소화하기 위해 리프 중심 트리 분할 방식을 채택한 지도 학습 알고리즘이다. 트리의 균형을 고려하지 않고 최대 손실 값을 가지는 리프 노드를 분할하여 비대칭적인 트리를 생성한다.

## 3. CatBoost

CatBoost[4]는 부스팅 기반 알고리즘의 과적합 문제를 해결하기 위해 제안된 지도학습 알고리즘이다. 균형 트리 분할 방식으로 트리를 만들어 나가며 순서형 부스팅, 무작위 순열 등의 방법을 통해 과적합을 방지한다.

### 2-3. SMOTE

SMOTE(Synthetic Minority Oversampling Technique)란 오버샘플링 기법의 하나로, 불균형 데이터의 문제를 해결하기 위한 기법이다. 낮은 비율로 존재하는 클래스의 데이터에 대해 k-NN 알고리즘을 사용해 데이터를 새롭게 생성하며, 단순 무작위 추출에 비해 과적합이 발생할 가능성이 적다는 장점이 있다.

## 3. 제안 신용 카드 이상 거래 탐지 시스템

본 논문에서는 XGBoost, LightGBM, CatBoost 알고리즘 기반의 신용 카드 이상 거래 탐지 시스템을 제안하고자 한다. XGBoost, LightGBM, CatBoost 알고리즘을 사용하여 실험을 진행하였고, 세 가지 모델 모두 훈련 횟수를 1,000회로 설정하였다. 추가적으로 CatBoost 모델의 경우 정수형 데이터를 다루는 모델이므로, 훈련 및 분류 평가를 위해 데이터 형식을 실수형에서 정수형으로 전처리를 수행하였으며, PCA 기반으로 데이터 전처리를 수행하였다. 훈련을 위해 전체 데이터에서 학습 데이터 70%, 테스트 데이터 세트를 분할하여 실험을 진행하였다.

## 4. 신용 카드 이상 거래 탐지 시스템 실험 결과

본 논문에서는 캐글에서 제공되는 신용 카드 이상 거래 데이터 세트를 사용하였다. 해당 데이터 세트는 2013년 9월 유럽 카드 소유자들의 거래 정보로, 이틀간 총 284,807개의 거래 내역으로 구성되어 있으며 신용 카드 이상 거래 데이터가 전체의 0.172%를 차지, 즉 정상 데이터 위주의 매우 불균형한 데

이터 세트이다. 따라서 이러한 불균형 문제를 해결하기 위해 SMOTE 기법을 이용하여 이상 데이터를 증가시켰다.

실험 결과로서, 신용 카드 이상 거래 탐지 및 분류의 정확도 판단을 위하여 정밀도, 재현율, F1-스코어를 산출하였다. 표1은 3가지 부스팅 알고리즘 기반 모델의 정밀도, 재현율, F1-스코어를 보여주며, 정밀도는 평균 99.95%, 재현율은 평균 99.99%, F1-스코어는 평균 99.97%를 보여줌으로써 이상 탐지 성능이 매우 월등한 것을 확인할 수 있었다.

model	정밀도	재현율	F1-스코어
XGBoost	99.96%	100%	99.98%
LightGBM	99.97%	100%	99.99%
CatBoost	99.92%	99.98%	99.95%

<표 1> 모델별 성능 비교 결과

model	unsampled	oversampled
XGBoost	90.84%	99.98%
LightGBM	23.87%	99.99%
CatBoost	82.58%	99.95%

<표 2> 오버샘플링 유무에 따른 F1-score

표 2는 오버샘플링 유무에 따른 F1-스코어 결과를 보여준다. 오버샘플링 기법을 적용하였을 경우 모든 부스팅 알고리즘 기반 신용 카드 이상 거래 탐지 성능이 매우 좋아지는 것을 확인할 수 있었다.

## 5. 결론

본 논문에서는 부스팅 기반 알고리즘인 XGBoost, LightGBM, CatBoost을 이용하여 신용 카드 이상 거래 탐지 시스템을 학습하였고, 세 가지 알고리즘 기반 모델 모두 평균 정확도 99% 이상을 보여주며 우수한 신용 카드 이상 거래 탐지 능력을 보여주었으며 그 중 LightGBM 알고리즘 기반 모델의 경우 오버샘플링 기법을 적용하였을 시 F1-스코어 기준 99.99%에 도달함에 따라 매우 우수한 신용 카드 이상 거래 탐지가 가능한 것을 확인하였다.

## 참고문헌

- [1] 금융결제국 결제연구팀, “주요국의 지급수단 사기[fraud] 동향 및 시사점”, 한국은행 금융결제국, 2020, <http://edu.copykiller.com/edu-source/faq/?mod=d>

ocument&uid=156.

- [2] Tianqi, C. and G. Carlos, "XGBoost: A Scalable Tree Boosting System", KDD'16, 13p, 2016.
- [3] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", NIPS, 9p, 2017.
- [4] Liudmila Prokhorenkova, Gleb Gusev , Aleksandr Vorobev , Anna Veronika Dorogush , Andrey Gulin "CatBoost: unbiased boosting with categorical features", NIPS, 23p, 2017.