JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 23, No. 10, pp. 2109-2115, Oct. 2022

Check for updates

금융거래 정보 및 모바일 GPS를 활용한 모바일 이상 거래 탐지 시스템 구현

윤 규 성[1] · 박 수 환[2] · 이 광 재[3*]
[1]쿠콘 금융보안클라우드센터 연구원
[2]윈스 보안관제팀 연구원
[3*]상명대학교 정보보안공학과 교수

An Implementation of Mobile Fraud Detection System using Financial Transactions and Mobile GPS Information

Gyu-Seong Yun[1] · Soo-Hwan Park[2] · Kwangjae Lee[3*]

[1]Researcher, Financial Security Cloud Center, COOCON, Seoul 07228, Korea
[2]Researcher, Department of Security Control Service, WINS, Seongnam 13487, Korea
[3*]Professor, Department of Information Security Engineering, Sangmyung University, Cheonan 31066, Korea

[요 약]

최근 모바일 결제의 사용자가 증가함에 따라 모바일 결제를 바탕으로 한 금전 탈취 사례도 같이 증가하는 추세이다. 현재 이상거래 탐지 시스템은 일반적인 거래에 최적화되어있어 모바일 거래에 적용하기에는 부족하다. 본 논문에서는 기계학습 기법을 활용한 모바일 이상거래 탐지 시스템을 제안한다. 추가로 모바일 결제는 공간제약 없이 사용 가능한 특성이 있으므로 이를 활용한 위치정보 분석을 제안한다. 평가지표는 Accuracy, Recall, Precision, F1 Score를 사용하였는데, 각각 0.99, 0.91, 0.95, 0.93으로 높은 성능을 보였다. 그리고 진단 방법에 대한 예측력을 평가하기 위해서 AUC(area under the ROC curve)를 사용하였고, 결과는 0.98이었다. 본 논문에서 제안한 방법으로 기계학습을 수행한 결과 모든 성능지표가 0.93 이상의 높은 탐지율을 보여주었으며, 모바일 기기의 위치정보를 활용으로 이상거래 탐지의 신뢰성이 높였다.

[Abstract]

Recently, as the number of users of mobile payment increases, cases of money theft based on mobile payment are also increasing. Currently, the anormal transaction detection system is optimized for general transactions, so it is insufficient to apply to mobile transactions. In this paper, we propose a mobile abnormal transaction detection system using machine learning techniques. In addition, since mobile payment can be used without space restrictions, location information analysis using this is proposed. The evaluation indicators were Accuracy, Recall, Precision, and F1 Score, which showed high performance with 0.99, 0.91, 0.95, and 0.93, respectively. And to evaluate the predictive power of the diagnostic method, AUC (area under the ROC curve) was used, and the result was 0.98. As a result of performing machine learning by the method proposed in this paper, all performance indicators showed a high detection rate of 0.93 or higher, and the reliability of anormal transaction detection was improved by using the location information of mobile devices.

색인어 : 이상거래 탐지 시스템, 모바일 금융거래, GPS 센서, 기계학습, 의사결정트리
Keyword : Fraud Detection System, Mobile Financial Transaction, GPS Sensor, Machine Learning, Decision Tree

http://dx.doi.org/10.9728/dcs.2022.23.10.2109

Received 12 October 2022; **Revised** 24 October 2022
Accepted 24 October 2022

***Corresponding Author; Kwangjae Lee**

Tel: +82-41-550-5269
E-mail: begleam@smu.ac.kr

Ⅰ. 서 론

국내외 금융 산업과 IT 기술의 발전에 따라 최근 금융거래 방식은 오프라인 거래보다 온라인 거래의 비중이 증가하고 있다[1]. 하지만 편리함을 강조한 모바일 금융거래는 낮은 보안성으로 해커들의 공격이 이어지고 있다[2]. 인터넷 은행인 카카오 뱅크를 예시로 들자면, 2017년 출범 이후 편의성을 이유로 이용 고객 수가 빠르게 증가하였지만 사기 이용 계좌가 2017년 199건에서 2020년 2,705건으로 13.6배 증가하였고 이는 전체 사기 이용 계좌 중 약 11%의 비율을 차지하는 수치이다[3], [4]. 앞선 사례로 모바일 금융거래의 취약성을 인식하였고, 정부는 금융 사고 방지를 위해 이상거래 탐지 시스템(FDS; Fraud Detection System)을 도입하고 있다. FDS는 전자금융거래에 사용되는 단말기의 정보, 접속정보, 거래내용, 결제 위치 등을 종합적으로 분석하여 평소 패턴과 다른 거래임이 탐지될 경우, 금융기관과 이용자에게 탐지 사실을 알리고 더 나아가 임의로 거래를 중단시키는 데 활용된다[5].

최근 스마트폰으로 결제하는 거래가 활성화되면서 기존에 구축된 일반적인 금융거래에 특화된 FDS는 모바일 금융거래의 특성을 잡아내지 못하는 경우가 빈번하다. 이에 모바일 환경에서 적용한 FDS들이 제안되었다. 하지만 이 FDS들은 정확성이 떨어지거나 나타내는 범죄를 해결하기에는 부족하였다[6]. 정확성을 높이기 위해서 기계학습을 사용한 연구가 제안되었으며, 계속하여 발전 중이다[7], [8]. 기계학습은 방대한 데이터셋이 필요하지만, 금융거래에서 사용하는 민감한 개인정보들 때문에 감춰지는 경우가 많아 어려움이 있다. 또한, 스마트폰에 탑재된 GPS 센서로 얻어온 사용자의 현재 위치 정보를 활용해 사용자가 직접 안전 구역을 구축하고 해당 지역에서 결제가 이루어지면 추가 인증 없이 결제가 진행되는 시스템도 제안되었다[5]. 이 시스템은 빠른 결제 서비스를 제공한다는 장점이 존재하지만, GPS 좌표 조작으로 추가 인증을 피해 가는 보안 취약점이 존재한다.

본 논문에서는 모바일 결제정보를 활용한 기계학습 기반 모바일 이상거래 탐지 시스템을 제안한다. 또한 신뢰성을 높이기 위해 모바일 GPS 및 비콘의 위치정보를 활용한 탐지 방법을 제안한다. 기계학습의 훈련을 위해서 PaySim 모바일 머니 시뮬레이터에서 생성된 합성 데이터셋을 사용하여 기계학습을 수행하였다[9]. 또한 결제 위치의 신뢰성을 줄 수 있도록 GPS 센서, 비콘 그리고 결제정보에 담긴 소매점 위치를 비교하는 방법을 사용하여 이상거래 탐지의 신뢰성이 높였다.

Ⅱ. 관련 연구

본 논문에서는 기계학습 기법을 활용한 모바일 이상거래 탐지 시스템을 제안하고 구현한다. 기계학습 기법을 활용할 때 시스템이나 데이터셋에서 얻을 수 있는 정보는 사기 거래가 0.13%인 불균형데이터를 사용해야 하기 때문에 사용하는 데이터셋의 특징에 적합한 알고리즘을 잘 선별해야한다. 따라서, 우리는 최적의 알고리즘 선별을 위해 의사결정트리(DT; Decision Tree), K-최근접 이웃(KNN; K-Nearest Neighbor), 서포트 벡터 머신(SVM; Support Vector Machine), 랜덤포레스트(RF; Random Forest) 총 4가지 알고리즘을 채택하여 각각의 알고리즘을 적용해 사용하였다.

DT 알고리즘은 불균형 데이터에서 다른 알고리즘보다 뛰어난 성능을 보이기 때문에 분류 알고리즘으로 널리 쓰인다. 데이터의 속성을 트리 구조로 나눠 정보이론을 이용해 데이터 집합을 분류하고 분류 규칙과 결과를 시각화한다. 이러한 과정을 쉽게 이해할 수 있어 탐지 결과를 사용자에게 명확하게 제시하고 설명할 수 있다[10]. 또한, 계산량이 적어 많은 컴퓨팅 작업을 필요로 하지 않기 때문에 빠른 속도로 분석을 수행할 수 있다. KNN 알고리즘은 분류나 회귀 알고리즘으로 사용되고 있으며, 기존 데이터 중 유사한 K개의 데이터를 이용해 값을 예측하는 알고리즘이다. 학습 과정이 빠르고 분류와 숫자 예측에 적용 가능하지만 각각의 데이터에 대한 거리를 구해야하기 때문에 분류에 많은 시간이 걸린다는 단점이 존재한다[11]. SVM 알고리즘은 지도 학습에서 사용되는 대표적인 알고리즘이다. SVM은 각 관찰값들을 선형 경계로 구별하는 방법을 사용하는 최대 마진 분류기이며 이는 훈련 관측치들로부터 거리가 가장 먼 최대 마진 초평면을 선택하여 분리하게 되는데 초평면은 데이터를 분리하기 위해 필요한 직선으로 초평면과 가장 가까이 있는 데이터와의 거리, 즉, 마진을 최대로 만드는 직선을 계산하여 데이터를 분류하는 방법을 SVM이라고 한다. 이 알고리즘은 예측의 정확도는 높지만, 모형 구축 시간이 오래 걸리고 신경망과 같이 해석이 어렵다는 단점이 존재한다[12]. Random Forest 알고리즘은 다수의 결정 이진 트리를 결합한 앙상블 학습 알고리즘이다. 학습과정에서 랜덤한 수의 노드를 생성시키고 각 트리의 노드마다 가장 적합한 판별식과 임계값을 결정한다. 생성된 임의의 수의 이진 결정트리들은 패턴 분류의 일반화율을 높이는 역할을 한다. 특히 불균형 데이터에서 다른 알고리즘보다 뛰어난 성능을 보이고 단일 의사결정 트리 모델이 특정한 학습 데이터에 대한 민감함과 불안정함을 보완해준다[13].

Ⅲ. 기계학습 기반 이상 금융거래탐지

본 논문은 위치정보 분석 및 기법을 활용한 모바일 이상 금융거래탐지 시스템을 제안한다. 이를 GFM(GPS-based Fraud detection method in Mobile payment) 시스템이라 지칭한다. 이 시스템은 기존에 있던 다른 연구의 이상 금융거래탐지 시스템과 달리 모바일 디바이스를 통한 온라인 또는 오프라인 결제에 대해서 이상거래와 정상거래를 판별하는 시스템이다. GFM은 위치정보를 활용해 탐지의 신뢰성을 높이고, 모바일 결제 데이터를 이용해 학습시킨 기계학습 모델로 정확한 정보를 사용자에게 제공한다.

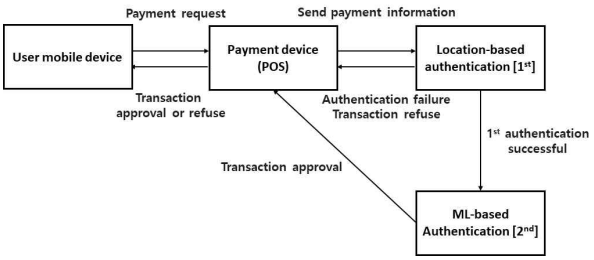

그림 1. 제안하는 2차 인증 기능의 이상거래 탐지 시스템
Fig. 1. Proposed second authentication function abnormal transaction detection system.

제안한 시스템의 구조는 그림 1과 같이 사용자 기기, 결제 기기, GPS 정보를 활용한 위치 기반 인증, 기계학습 기반 인증으로 구성된다. GFM 시스템은 Windows 운영체제를 기반으로 하여 웹 서버와 기계학습 서버가 통합된 서버를 구축하였다. 웹 서버에서는 사용자로부터 결제요청 데이터를 안드로이드 App을 통해 수집하고, Rest API 방식을 이용해 기계학습 서버로 전달한다. 그 데이터 중에서 우선 사용자의 모바일 기기의 위치와 판매자의 결제기기의 위치의 정보를 분석하여 1차적으로 정상 거래인지 이상 거래인지를 판단한다. 위치에 대한 정보만으로 이상이 없을 시에는 기계학습 서버로 데이터가 전달되고, 이상이 있다고 판단되면 해당 요청에 대한 거래 정보를 분석 및 탐지하게 된다. 사용자에게 정상 거래로 판단될 때는 결제승인, 이상 거래로 판단될 때는 결제를 거부하기 때문에 금융 사고로부터 사용자를 보호해주는 시스템이다. 또한, 사용자의 모바일 디바이스 위치와 판매자의 결제기기의 위치를 분석하는 정보를 제공하여 해당 거래에 대한 신뢰성을 높인다.

3-1 데이터 소개 및 피처 선별

GFM 시스템에서 사용되는 데이터셋은 PaySim 모바일 머니 시뮬레이터에서 생성된 합성 데이터셋(Synthetic Financial Datasets For Fraud Detection)을 사용한다[9]. 금융 관련 서비스의 데이터들은 개인정보 등 보안의 이유로 공개되어 있는 데이터셋을 구하기가 힘들다. 이러한 문제를 해결하기 위해 아프리카 국가에서 구현된 모바일 머니 서비스에서 1개월간 재무 로그에서 추출한 실제 거래 샘플이 기반인 합성 모바일 금융거래 데이터셋을 사용하였다. 이렇게 합성된 데이터셋은 원래 원본의 데이터셋의 1/4로 축소된 데이터셋으로 생성되었으며 다섯 가지 타입의 금융 데이터들을 도합해서 약 2,400만 개의 재무 기록이 있다.

PaySim 데이터셋의 특징들은 표 1과 같다. GFM 시스템에서는 기계학습에 훈련에 불필요한 특징인 발송인과 수령인의 ID, 수령인의 계좌 잔고 정보를 제외한 거래시간, 거래 유형, 거래 금액, 거래 전후의 잔고 총 다섯 가지의 특징들을 선별했다. 그리고 정답 데이터로 사기 여부를 보여주는 isFraud를 사용하였다.

표 1. 모델 구축에 사용된 PaySim 데이터셋 특징
Table 1. The PaySim dataset feature used to build the model

Label Name	Description
step	Maps a unit of time in the real world. In this case 1 step is 1 hour of time.
type	Transaction type – 1, 2 : CASH_IN, CASH_OUT (Transactions without sellers) – 3 : DEBIT (Transaction that pay by debit card) – 4 : PAYMENT (Transaction that pays the seller) – 5 : TRANSFER (Remittance)
amount	Amount of the transaction in local currency – the amount of transactions – most fraudulent transactions are small
oldbalanceOrg	Initial balance before the transaction
newbalanceOrig	Customer's balance after the transaction
isFraud	Identifies a fraudulent transaction (1) and non fraudulent (0)

2018년 전 세계 카드 사기율은 0.07%로 통상적으로 사기 거래와 정상 거래를 구분해주는 데이터셋은 불균형한 데이터셋의 형태를 가진다. 방대한 양을 가진(약 630만 개) PaySim 데이터셋에서도 사기 거래의 비율이 0.13%밖에 되지 않았다[14]. 또한 PaySim 데이터셋의 사기 거래가 발생한 데이터들을 분석해본 결과 대부분의 사기 거래가 연속적으로 이루어진다는 사실을 알게 되었고, 이러한 데이터들의 특징을 이용해 현재 거래 및 이전 거래들을 동시에 검사할 수 있도록 새로운 데이터셋을 제안하고, 기계학습 훈련을 진행해 보았다. 그림 2는 PaySim 데이터셋의 사기거 래 발생 데이터의 예시이다. 데이터 셋 내에 사기 거래만을 확인하기 위해 isFraud 항목이 1인 경우, TRANSFER와 CASH_OUT이 연속적으로 발생하는 것을 확인할 수 있었다.

3-2 기계학습

기계학습 분류 모델은 PaySim 데이터셋의 특성에 맞추어 불균형 데이터셋에서도 높은 정확도를 가질 수 있는 DT 알고리즘을 사용하였다. 이 알고리즘은 계산량이 적어 많은 컴퓨팅 작업을 필요로 하지 않기 때문에 빠른 속도로 분석을 수행할 수 있는 장점이 있어 앞서 설명한 다른 알고리즘보다 더 좋은 결과를 얻을 것이라 예상했다. 또한 빠르게 동작하기 때문에 실시간으로 거래의 사기 여부를 판단하기에 좋은 모델이다.

step	type	amount	nameOrig	oldbalanceOr	newbalanceO	nameDest	isFraud
1	TRANSFER	181	C130548614	181	0	C553264065	1
1	CASH_OUT	181	C840083671	181	0	C38997010	1
1	TRANSFER	2806	C142019642	2806	0	C972765878	1
1	CASH_OUT	2806	C210152707	2806	0	C100725173	1
1	TRANSFER	20128	C137533655	20128	0	C184841504	1
1	CASH_OUT	20128	C111843067	20128	0	C339924917	1

그림 2. PaySim 데이터셋의 사기 거래 발생 데이터 예시
Fig. 2. An example of fraudulent transaction occurrence data in PaySim dataset.

3-3 위치정보 및 비콘 기능설계 구현

GPS 정보를 이용한 이상 거래 탐지를 위해 결제 기기가 비콘(Beacon)을 탑재하고, 사용자 기기 애플리케이션을 통해 정보를 얻어오도록 구현하였다. 비콘이란 근거리에 있는 스마트 기기를 자동으로 인식하여 필요한 데이터를 전송할 수 있는 무선 통신 장치이다. 본 논문에서는 비콘에서 제공하는 정보인 Major, Minor, UUID 값을 활용하여 판매자의 고유 식별자를 만들었다. 거래를 시도하는 해당 결제 기기에서 결제 요청이 들어오면 사용자의 스마트폰에서 결제 기기의 ID가 인식된 시점으로 제공되는 RSSI(Receive Signal Strength Indicator)를 역계산 하여 나온 거리정보가 GFM 애플리케이션에 출력한다. 사용자는 거래가 실행된 장소와의 거리정보를 확인함으로써 해당 거래에 대한 신뢰에 도움을 준다. 식 (1)은 RSSI로 거리를 측정하는 수식이다. 여기서 Distance는 거리이며 미터 단위로 표시한다. 그리고 TXPOWER는 송신기의 전파 크기, RSSI는 수신된 전파 크기, n은 보정 상수(보통의 경우 2로 설정)이다.

$$Distance = 10^{((TX_{power} - RSSI) / (10 * n))} \quad (1)$$

사용자 모바일 기기에서 거래 요청하면 사용자의 위치정보 및 결제정보를 결제 기기로 전송한다. 결제 기기는 사용자 기기에서 받은 정보를 인증 서버로 전송하여 이상 거래 여부를 판단한다. 인증 서버의 1차 인증은 위치정보 기반탐지로, 이상 거래로 판단되면 거래 거부 결과를 결제 기기로 전송하며, 정상 거래라고 판단되면 2차 인증을 진행한다. 2차 인증은 미리 학습된 모델을 통해 이상거래 여부를 판단한다. 판단 결과값은 판매기기로 전송되게 되고 결과값에 따라 사용자 기기에 결제승인, 결제거부 응답을 한다.

Ⅳ. 시스템 구성

시스템은 클라이언트, 서버, 위치기반 인증(1차 인증), 기계학습기반 인증(2차 인증)으로 이루어진다. 클라이언트는 서버에게 결제요청을 보내고, 서버는 위치기반 인증과 기계학습기반 인증을 통해 결제요청이 올바른 지 검사한다. 2차 인증까지 수행한 결과에서 결제요청이 정상 거래였다면 서버는 클라이언트에게 결제요청을 승인한다. 반대로, 사기 거래였다면 서버는 클라이언트에게 결제요청을 거부한다.

4-1 클라이언트

클라이언트는 안드로이드 기반으로 제작된 애플리케이션으로, 사용자가 앱 실행 후 결제 요청을 진행한다.

그림 3. 결제정보 입력 후 이상거래 판단 결과 예시
Fig. 3. A example of results of normal and anormal transaction.

구현된 애플리케이션에서는 가상의 사용자 정보 및 결제정보를 입력하여 결제 요청하였다. 또한, GFM 시스템을 통해 사용자의 결제 요청시 보낸 정보에 따라 이상 거래 판단 결과를 가시화해주는 역할도 수행한다. 애플리케이션에서는 기계학습 과정에서 필요한 정보인 결제 시간, 결제유형, 금액, 거래 전 잔고, 거래 후 잔고를 입력하도록 되어 있다. 결제정보 입력 후 결제 실행 시 데이터값을 전환하고 Rest API 방식으로 인증 서버로 값을 전달한다. 그리고 다시 인증 서버는 애플리케이션으로 응답 정보인 거래 성공, 거래 실패 결과를 받고, 이상 거래 판단 결과 확인할 수 있도록 최하단에 텍스트로 출력하였다.

4-2 서버

서버 구축은 Windows 운영체제를 기반으로 하여 클라이언트에서 주입하는 데이터를 훈련된 기계학습 모델로 전달하고 판단 결과를 다시 클라이언트로 전달한다. 안드로이드 애플리케이션과 기계학습 모델을 연결하기 위해 Python에서 Flask 서버를 구축하였고, 모델에서 저장한 Pkl 파일을 Flask 서버에서 불러오는 방식을 사용하였다. Flask는 Python으로 작성된 웹 프레임워크이다. Flask는 자체에 구현된 것처럼 애플리케이션 기능을 추가할 수 있는 확장기능을 지원하기 때문에, 이 시스템에 사용하였다.

클라이언트(App)에서 JSON 형태의 데이터로 저장 후 Flask 서버에 전달하게 되면 List 방식으로 정보를 받아오게 구성된 기계학습 모델에 정보를 넣기 위해 JSON 형태의 데이터를 List 형태의 데이터로 바꾸는 작업을 진행하였다. Flask 서버 구축이 완료된다. 클라이언트에서 Flask 서버에 접속하여 데이터를 전달하기 위해서는 서버가 외부로 접속을 허용해야한다. 본 시스템에서는 Ngrok 프로그램을 사용하여 로컬 호스트를 외부에서 접근할 수 있도록 터널을 만들고 도메인을 할당하였다. 그림 4는 port를 7070으로 지정하고 Flask 서버, Ngrok 프로그램을 이용하여 안드로이드 클라이언트와 Flask 서버의 통신이 가능하도록 한 예시이다.

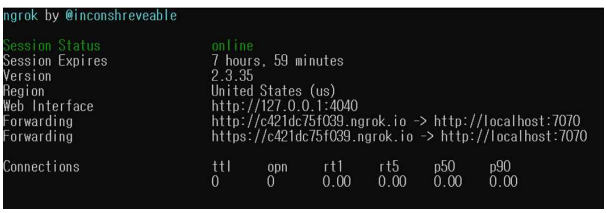

그림 4. 서버 구축 결과 화면(ngrok 실행)
Fig. 4. Server build result screen(run ngrok).

이후 클라이언트에서 거래 데이터를 주입하게 되면 거래 데이터는 Flask 서버를 거쳐 기계학습 모델로 들어가게 되고 모델에서 이상 거래 여부를 판단한 결과를 다시 Flask 서버를 통해 안드로이드 애플리케이션 화면에 출력한다.

4-3 위치정보

위치정보를 분석하여 이상 거래 탐지를 위한 실험을 위해서 스마트폰과 비콘을 사용하였다. 스마트폰은 사용자가 모바일 결제를 하는 기기이며, 비콘은 판매자의 결제 기기 POS 기기와 같은 결제 담당 기기로 가정한다. 비콘에서 제공하는 정보인 Major, Minor, UUID 값을 활용하여 판매자의 고유 식별자를 만들었다. 비콘은 라즈베리파이 3 B+ 모델을 사용하여 전파 송출을 만들었으며, UUID를 포함한 값을 수정하여 고유 식별자로서 판매자를 구분할 수 있도록 하였다. 그리고 스마트폰 애플리케이션을 통해 비콘의 정보를 읽어와 얼마나 떨어져 있는지, Major, Minor, UUID값을 갖는지 확인할 수 있다. 그림 5는 비콘 기능을 사용했을 때 POS기기와 사용자의 거리를 보여준다. UUID로 개인 고유의 ID를 지정하고 아래의 RSSI 값을 이용해 거리를 계산 후에 거리정보를 출력하게 된다.

V. 이상거래 탐지 실험 결과 및 고찰

DT로 구축된 이상 거래 탐지 모델에 대한 예측 성능을 확인하기 위해 다양한 평가지표를 통해 모델의 예측 성능을 비교/평가한다. 정확성(Accuracy), 재현율(Recall), 정밀도(Precision), F1 score, AUC, ROC curve와 같이 여섯 개의 평가지표를 사용해 예측 성능을 평가했다[15]. 정확성은 전체 대비 정확하게 예측한 개수의 비율이다. 하지만 본 논문의 데이터셋처럼 데이터가 불균형하면, 무조건 높은 성능을 도출하게 되므로 성능을 왜곡할 수 있는 문제점을 가지고 있다. 재현율은 실제 정상 거래인 것 중에서 얼마나 잘 예측하였는지의 비율이다. 정밀도는 정상 거래라고 예측한 비율 중 진짜 정상 거래인 비율이다. F1 score은 정밀도와 재현율을 통합한 측정 지표로 두 지표 값의 조화평균이다. F1 score은 불균형 데이터셋을 이용한 모델일 때 좋은 지표이다.

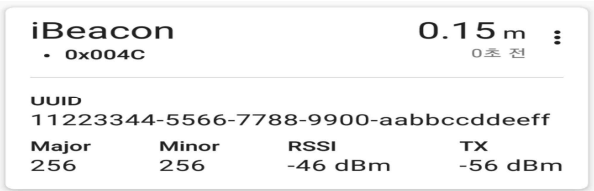

그림 5. 라즈베리파이를 이용한 Becon 구현
Fig. 5. Becon implementation using Raspberry Pi (Distance, Major, Minor, UUID).

ROC curve는 진짜 정상 양성 비율에 대한 거짓 양성 비율을 나타내며, 불균형한 데이터셋에서 정확도보다 훨씬 더 좋은 지표이다. AUC는 ROC 그래프를 명확한 수치로써 비교하기가 어려워 그래프 아래의 면적값을 이용한다. 최대값은 1이며 좋은 모델은 1에 가까운 값이 나온다[16].

본 논문의 아이디어는 이전 거래를 활용하여 예측의 정확성을 높이는 방법이다. 활용한 PaySim 데이터셋을 이전거래도 함께 테스트할 수 있도록 변형을 하고 각각을 비교해 보았다. PaySim 데이터셋의 사기 거래가 발생한 데이터들을 분석한 결과 대부분의 사기 거래가 연속적으로 이루어진다는 사실을 알게 되었고, 이러한 데이터들의 특징을 이용해 현재 거래 데이터셋과 이전거래 데이터셋을 모델에 동시에 넣어 평가지표를 비교했다.

표 2. GFM 시스템의 알고리즘 성능 평가
Table 2. A Performance evaluation of GFM system's algorithm.

	Accuracy	Recall	Precision	F1 Score	AUC
1-Tran	0.99	0.84	0.90	0.87	0.97
2-Tran	0.99	0.91	0.95	0.93	0.98
3-Tran	0.99	0.90	0.94	0.92	0.97
4-Tran	0.99	0.89	0.94	0.91	0.97
5-Tran	0.99	0.88	0.92	0.90	0.97

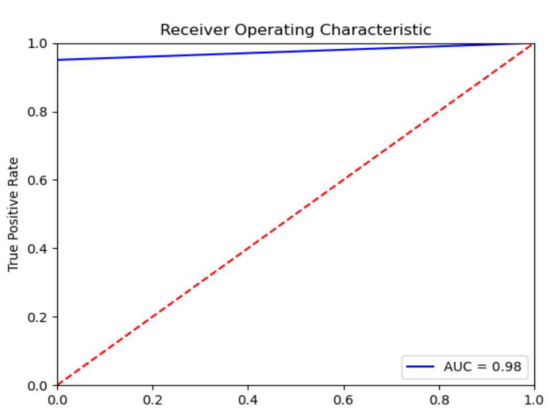

그림 6. AUC-ROC 결과 (2-Tran 기준)
Fig. 6. a AUC-ROC result based on 2-Tran.

총 5개의 모델을 비교했는데 현재 거래와 이전거래들을 포함해 2개, 3개, 4개, 5개의 데이터를 비교하는 모델이며, 현재 거래만을 표현한 데이터셋은 1-Tran, 이전 거래를 포함한 데이터셋은 개수에 따라서 2-Tran, 3-Tran, 4-Tran, 5-Tran으로 표현한다. 표 2는 제안한 GFM 시스템의 성능 평가 지표이다. 앞서 언급한 대로 사기의 패턴이 연이어 사기가 발생하였으므로, 현재 거래와 바로 직전 거래를 데이터셋으로 사용한 2-Tran이 가장 좋은 지표를 가진다. 그림 6은 AUC-ROC 곡선으로 분석한 성능 평가 지표로 0.98로 매우 높은 수치를 갖는다.

Ⅵ. 결 론

본 논문에서는 웹과 기계학습 서버, GPS 정보를 이용한 기계학습 모델 구축, 사용자의 결제정보를 바탕으로 한 GFM 시스템을 설계했다. 제안한 현재 거래 및 이전 거래를 합쳐 기계학습을 수행한 결과로 모든 성능지표가 0.91 이상임을 확인할 수 있었고, 라즈베리파이를 통해 판매자의 결제 기기를 모델링하여 스마트폰과의 위치정보를 확인해 이상거래 탐지의 신뢰성을 높였다. 모바일 결제시장의 규모가 점차 확대됨에 따라 그에 따른 결제 사기 등의 사례가 더 증가할 것으로 판단되므로, 본 논문에서 제안하는 GFM 시스템의 연구 및 개발을 통해 사용자의 모바일 결제 이상거래 탐지에 도움이 되는 시스템이 되기를 기대한다.

참고문헌

[1] K. H. Seo. In finance, the era of digital and smart transactions is now wide open [Internet]. Available: https://news.imaeil.com/page/view/2019050918340441728.

[2] J. W. Lee. Convenient finance, how to ensure ‘safety’ [Internet]. Available: https://www.hani.co.kr/arti/economy/economy_general/969264.html.

[3] K. M. Lee. Kakao Bank fraudulent accounts soar... 13.6 times in 3 years [Internet]. Available: https://www.hani.co.kr/arti/economy/finance/1013579.html.

[4] H. W. Lim. When will Kakao Bank's fraudulent payment, FDS improve? [Internet]. Available: https://www.ekoreanews.co.kr/news/articleView.html?idxno=44677.

[5] M. K. Lee, H. J. Shon, B. M. Sung, and J. B. Kim, “Fraud Detection System applying GPS in Fintech-based environment,” *Multimedia Journal of Convergence of Arts, Humanities and Social Sciences*, Vol. 5, No. 4, pp. 659-666, 2015.

[6] H. Zhou, H. F. Chai, and M. L. Qiu, “Fraud detection within bankcard enrollment on mobile device based payment using machine learning,” *Frontiers of Information Technology & Electronic Engineering*, Vol. 19, No. 12, pp. 1537-1545, 2018.

[7] C. M. R. Haider, A. Iqbal, A. H. Rahman, and M. S. Rahman, “An ensemble learning based approach for impression fraud detection in mobile advertising,” *Journal of Network and Computer Applications*, Vol.112, pp. 126-141, 2018.

[8] J. Hu, T. Li, Y. Zhuang, S. Huang, and S. Dong, “GFD: A Weighted Heterogeneous Graph Embedding Based Approach for Fraud Detection in Mobile Advertising,” *Security & Communication Networks*, pp. 1-12, 2020.

[9] E. A. Lopez-Rojas, A. Elmir, and S. Axelsson, “PaySim: A financial mobile money simulator for fraud detection,” In *28th European Modeling and Simulation Symposium, EMSS*, Larnaca, pp. 249-255, 2016.

[10] H. C. Han, H. Kim, and H. K. Kim, “Anormal transaction detection system using data mining in mobile payment environment,” *Journal of the Information Security Society*, Vol. 26, No. 6, pp. 1527-1537, 2016.

[11] T. Park, “Shifted Sorting-based k-Approximate Nearest Neighbor Searching Algorithm with Extra Loops Based on Permutation for Better Accuracy,” *Journal of Digital Contents Society*, Vol. 22, No. 2, pp. 325-330, 2021.

[12] M. Kim and J. Seo, “A Control Method of ASMR Contents through Attention and Meditation Detection Based on Internet of Things,” *Journal of Digital Contents Society*, Vol. 19, No. 9, pp. 1819-1824, 2018.

[13] N. H. Lee and T. M. Lee, “A study on sentiment analysis from photos based on machine learning model,” *Journal of Digital Contents Society*, Vol. 22, No. 8, pp. 1295-1302, 2021.

[14] E. A Lopez-Rojas, S. Axelsson, and D. Baca, “Analysis of fraud controls using the PaySim financial simulator,” *Int. J. Simul. Process. Model.*, Vol. 13, No. 4, pp. 377-386, 2018.

[15] Wikipedia. Confusion matrix [Internet]. Available: https://en.wikipedia.org/wiki/Confusion_matrix.

[16] S. Narkhede, “Understanding auc-roc curve,” *Towards Data Science*, Vol. 26, No. 1, pp. 220-227, 2018.

윤규성(Gyu-Seong Yun)

2021년 : 상명대학교 정보보호공학과 (공학학사)

2021년~2022년: SK쉴더스 보안관제팀
2022년~현 재: 쿠콘 금융보안클라우드센터 연구원
※관심분야 : 금융 보안(Financial Security), 기계 학습(Machine Learning)

박수환(Soo-Hwan Park)

2021년 : 상명대학교 정보보호공학과 (공학학사)

2021년~현 재: 윈스 보안관제팀 연구원
※관심분야 : 금융 보안(Financial Security), Internet of Things(IoT)

이광재(Kwangjae Lee)

2014년 : 고려대학교 전자컴퓨터공학과 석박사통합과정 (공학박사)

2014년~2017년: 한국전자기술연구원 정보통신미디어연구본부 연구원
2017년~현 재: 상명대학교 정보보안공학과 교수
※관심분야 : 디지털시스템 설계, 암호모듈 검증(CMVP), 임베디드시스템