1. 초록

선정한 데이터는 kaggle에서 가져온 데이터로, 카드 사기를 판별할 수 있는 데이터이다.   
<https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud>

본 실험에서는 이미 y 값이 labeling 되어 있기에 (정상 : 0, 사기 : 1), 이를 이상탐지 알고리즘을 활용하여 해당 데이터의 정상과 비정상을 탐지한다. 이를 비교하여 해당 학습 모델을 구축하고 이들의 효율성을 측정하려 한다. 활용된 알고리즘은 SWM과 SVDD이다.

실험에서는 A, B, C. .. 알고리즘을 적용하여 정확도 측면의 비교 실험을 수행했으며, 실험 결과 X가 TPR X.XX%, FPR X.XX%, 그리고 F1 X.XX%만큼 좋은 성능을 보였다. 본 실험을 통해서 이상탐지가 현장에 적용될 수 있음을 느꼈고, 본 데이터를 기준으로 봤을 때 현장에 어떻게 적용될 수 있을 것 같다는 생각이 들었다

선정한 데이터는 XXX로부터 XXX에 대한 문제를 담고 있는 데이터이다. 최근 이와 관련한 사회적 관심이 집중된 가운데, 기법적으로 이러한 문제를 풀기 위한 많은 노력들이 있다. 본 실험에서는 데이터의 기본적인 특징을 이해하고 정상을 정의한 후, 이상탐지 알고리즘을 활용하여 해당 데이터의 정상을 탐지하는 학습모델을 구축하고자 한다. 특히, 수업을 통해 학습한 알고리즘 외 XXX 알고리즘을 스스로 학습하고 실험에 추가하였다. 실험에서는 A, B, C. .. 알고리즘을 적용하여 정확도 측면의 비교 실험을 수행했으며, 실험 결과 X가 TPR X.XX%, FPR X.XX%, 그리고 F1 X.XX%만큼 좋은 성능을 보였다. 본 실험을 통해서 이상탐지가 현장에 적용될 수 있음을 느꼈고, 본 데이터를 기준으로 봤을 때 현장에 어떻게 적용될 수 있을 것 같다는 생각이 들었다

1. 문제 설명

이 데이터는 카드 관련 features들을 토대로 카드가

이 데이터는

1. 이상탐지를 적용하려는 적용하려는 데이터의 EDA

이 데이터는 7개의 독립변수와 (=x) 1개의 종속변수(= y)가 존재한다.   
이는 다음과 같다:

Distance\_from\_home: 거래가 일어난 장소와 집 사이의 거리를 나타낸다.

Distance\_from\_last\_transaction: 마지막 거래 장소와 집 사이의 거리를 나타낸다

Ratio\_to\_median\_purchase-price: 구매 값/ 중앙값 의 비율을 나타낸다.

Repeat\_retailer: 같은 소매인에게 반복적으로 거래가 일어났는지를 암시한다.

Used\_chip: 카드를 긁어서 계산했는지 확인한다.

Used\_pin\_number: 핀 번호를 입력하여 계산했는지 확인한다.

Online\_order: 거래가 온라인으로 이루어졌는지 확인한다.

Fraud: 거래가 사기였는 지 확인한다.

데이터

1. 본인이 선정한 새 이상탐지 알고리즘 간략 소개
2. 실험결과 및 결론