데이터 세트

1. <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

2013년 유럽의 카드 소지자의 신용카드 거래 내역을 담은 데이터 세트

이상거래 탐지에 가장 많이 사용되는 데이터 세트

Feature

-V1부터 V28: 은 PCA를 통해 얻은 주성분으로 익명화된 변환 특징 정보

-Time: 데이터 셋의 첫 번째 거래 이후 경과된 시간을 초단위로 나타냄(거래시간 정보)

-Amount: 거래 금액을 나타내며, 비용민감 학습에 활용할 수 있음(거래금액 정보)

-Class: 응답 변수로, 이상 거래인 경우 1, 정상거래인 경우 0의 값을 가짐(사기 여부, 모델이 예측하고자 하는 대상)

2. [https://www.kaggle.com/datasets/dhan shnarayananr/credit-card-fraud](https://www.kaggle.com/datasets/dhan%20shnarayananr/credit-card-fraud)

Feature

-distaance\_from\_home: 카드 주인의 집과 거래가 발생한 장소와의 거리

-distance\_from\_last\_transaction: 마지막 거래가 발생한 장소와 현재 분석하고자 하는 거래가 발생한 장소와의 공간적 거리 차이

-ratio\_to\_median\_purchase\_price: (해당 거래의 구매 가격)/(중앙값 구매 가격)

중앙값 구매 가격: 지금까지 발생했던 모든 거래가격을 순서대로 정렬했을 때 가운데 위치하는 값

-repeat\_retailer: 동일한 소매점에서 과거에 소비가 발생했는지 여부

-online\_order: 거래가 온라인에서 이루어졌는지 여부

-fraud: 거래가 이상거래인지 여부(1=이상거래, 0=정상거래)

3. <https://www.kaggle.com/datasets/bhadramohit/credit-card-fraud-detection>

Feature

-Transaction ID: 각 거래를 식별하는 고유 ID

-TransactionDate: 거래가 발생한 날짜와 시간

-Amount: 거래 금액

-MerchantID: 거래가 발생한 상점의 고유한 ID

-TransactionType: 거래 유형(예: 구매, 환불)

-Location: 거래가 발생한 위치

-IsFraud: 거래가 이상거래인지 정상거래인지 여부

4. <https://www.kaggle.com/datasets/anurag629/credit-card-fraud-transaction-data/data>

Feature

-Transaction ID:각 거래를 식별하는 ID

-Date: 거래가 발생한 날짜

-Day of week: 거래가 발생한 요일

-Time: 거래가 발생한 시간

-Type of Card: 거래의 사용된 신용카드의 종류(예: visa, master card)

-Entry Mode: 카드가 어떻게 사용되었는지(예: tap, pin입력)

-Amount: 거래 금액

-Type of Transaction: 거래가 이루어진 방식(예: online, pos, atm)

-Merchant Group**:** 거래가 이루어진 가맹점의 그룹 또는 카테고리

-Transaction Country: 거래가 발생한 국가

-Shipping Address**:** 상품 배송 주소 (온라인 거래의 경우)

-Billing Address: 카드 청구 주소

-Gender of Cardholder: 카드 소지자의 성별

-Age of Cardholder**:** 카드 소지자의 나이

-Issuing bank: 카드를 발급한 은행

-Fraud: 거래가 이상거래인지 정상거래인지 여부

5 <https://www.kaggle.com/datasets/youssefismail20/fraudsynth-credit-fraud-detection-dataset>

Feature

-timestamp: 거래의 날짜와 시간

-user\_id: 각 사용자의 고유 식별자

-amount: 통화로 표시된 거래 금액

-location: 거래가 발생한 위치(도시)

-device\_type: 거래에 사용된 기기 유형(데스크톱, 모바일, 태블릿)

-age: 사용자의 나이

-income: 사용자의 소득

-debt: 사용자와 관련된 부채

-credit\_score: 사용자의 신용점수

-is\_fraud: 거래가 이상거래인지 여부

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 데이터 셋 | 전체거래 수 | 정상거래 수 | 이상거래 수 | 비율(이상/전체) | 컬럼의 종류 |
| 1 | 284,807 | 284,315 | 492 | 0.0017 | 31 |
| 2 | 1,000,000 | 912,597 | 87,403 | 0.0874 | 8 |
| 3 | 100,000 | 99,000 | 1,000 | 0.01 | 7 |
| 4 | 100,000 | 92,805 | 7,195 | 0.0719 | 16 |
| 5 | 1,048,575 | 1,000,000 | 48,575 | 0.0463 | 10 |
| 계 | 2,533,382 | 2,388,717 | 144,665 | 약 5.7% |  |

공통 feature

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 거래가 사기인지 여부 | 거래 금액 | 거래 날짜(시간) | 거래가 발생한 장소 |  |
| 1 | Class | Amount | Time(경과 시간) |  |  |
| 2 | Fraud | ratio\_to\_median\_purchase\_price |  | distaance\_from\_home, distance\_from\_last\_transaction(공간적 거리 차이) |  |
| 3 | IsFraud | Amount | TransactionDate | Locatiion |  |
| 4 | Fraud | Amount | Date, Time | Transaction Country |  |
| 5 | Is\_fraud | amount | timestamp | location |  |