

금융데이터 ML 실습

산업정보시스템 한다운
산업정보시스템 최도한
산업정보시스템 황수은

Contents

01 클론코딩

02 DAICON 실습

클론코딩

Kaggle의 Credit Card Fraud Detection

- 약 3주간 진행
- 공유된 코드 중 most votes 기준으로 정렬하여 상위 코드 클론코딩



JANIO MARTINEZ BACHMANN · 5Y AGO · 877,994 VIEWS

▲ 5234

Copy & Edit 11156

Credit Fraud | | Dealing with Imbalanced Datasets

Python · Credit Card Fraud Detection

<https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets>

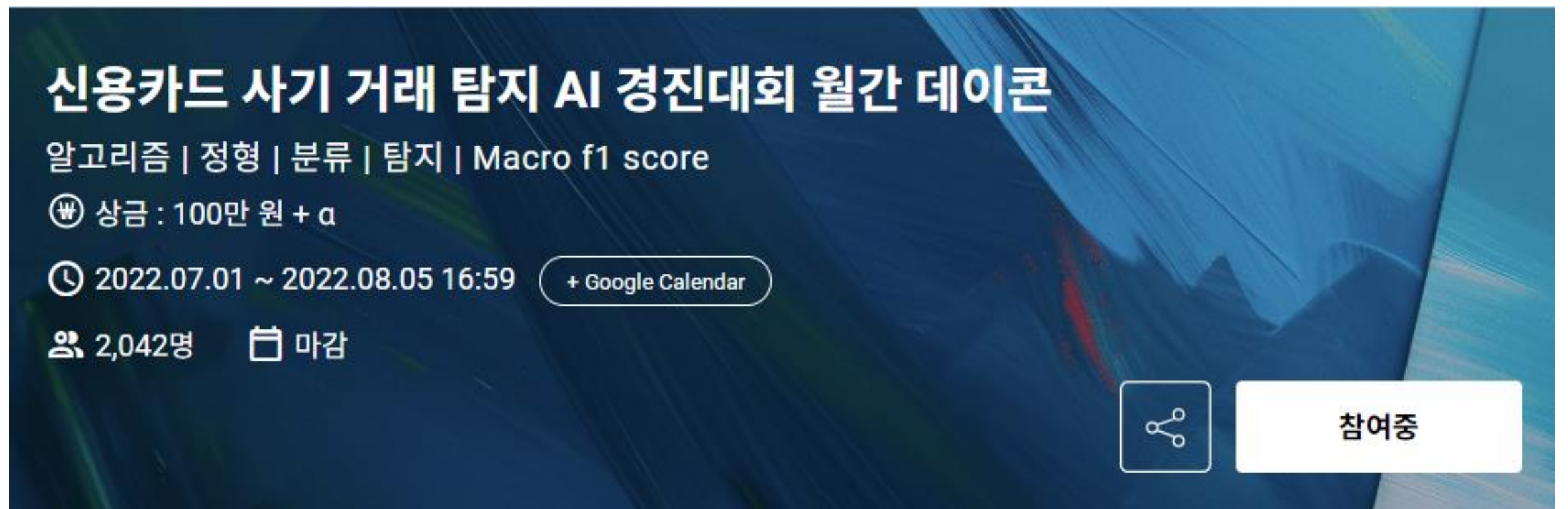
(+) velog 정리

1. <https://velog.io/@ceromi04/FDS이상거래탐지-kaggle-클론코딩01>
2. <https://velog.io/@ceromi04/클론코딩>

DACON 실습

DACON의 신용카드 사기 거래 탐지 AI 경진대회 월간 데이콘

- 4주차~
- 매주 정해진 단계까지 분석 진행해온 다음 모여서 코드 공유, 피드백하는 방식으로 진행



신용카드 사기 거래 탐지 AI 경진대회 월간 데이콘

알고리즘 | 정형 | 분류 | 탐지 | Macro f1 score

₩ 상금 : 100만 원 + α

🕒 2022.07.01 ~ 2022.08.05 16:59 [+ Google Calendar](#)

👤 2,042명 📅 마감

[🔗](#) [참여중](#)

대회 소개

[사용 데이터]

- **train.csv(113842, 30)**

정상, 사기 거래의 여부를 알 수 없는(Unlabeled) 신용 카드 데이터

- **val.csv(28462, 31)**

정상, 사기 거래의 여부가 포함된(labeled) 신용 카드 데이터

- **test.csv(142503, 30)**

정상, 사기 거래의 여부를 알 수 없는(Unlabeled) 신용 카드 데이터

대회 소개

[규칙]

- 심사 기준: macro f1-score
- 1차 평가(public score): 테스트 데이터 중 랜덤 샘플링된 30%로 채점
- 2차 평가(private score): 테스트 데이터 중 나머지 70%로 채점
- labeled data(validation set)은 모델 학습에 사용은 불가능하지만 데이터 통계 정보와 모델 평가에는 사용 가능

DACON

실습

분석 방법 - 1

Labeled data인 valid dataset과 unlabeled data인 train dataset를 합쳐서 하나의 데이터셋으로 만들.

이후 Fancyimputer 사용하여 train 데이터에 라벨 생성 후 XGBoost 모델 적용

Public: 0.925

Private: 0.909

<https://velog.io/@ceromi04/DACON신용카드-사기-거래-탐지-AI-경진대회-월간-데이콘>

분석 방법 - 2

군집 알고리즘을 사용해 이상치를 탐색하고 그걸 기반으로 train 데이터에 레이블을 부여. 로지스틱 회귀, 트리 모델, SVM을 앙상블한 모델을 사용.

Public: 0.686
Private: 0.687

<https://velog.io/@ehgks904/금융데이터ML-1-데이콘-신용카드-사기-거래-탐지>

분석 방법 - 3

Valid data를 기반으로 KNN 모델을 이용하여 train 라벨 생성. XGBoost로 분류

Public: 0.932

Private: 0.902

<https://velog.io/@hgs/ppstudyfrauddetection1>