

프로젝트 과정

< 파일럿 프로젝트 기획서 >

주 제	Amex-default-prediction		
분 야		일 자	2022년 10월
팀 장	오동익	팀 원	

1. 과제 개요

캐글 대회에 참해서 실전 데이터 분석을 해봄으로써 데이터 분석을 구체적으로 익히고 머신러닝 딥러닝 기본 능력을 향상시킬 수 있다.

American Express는 결제를 대신해주고 고객에게 결제액을 빌려주는 대신 수수료를 받는 BM을 가지고 있을 텐데, 고객이 결제를 한 후 추후 채무 불이행을 한다면 금전적 손해로 이어지기 때문에 머신러닝 기술을 이용하여 고객이 신용 채무 불이행을 저지르지 여부를 예측하는 모델을 만든다.

target(binary)은 최근 명세서 이후 18개월 간 관찰 후 산출되고 최근 명세서 이후 120일 이내에 납부 금액을 납부하지 않으면 디폴트로 간주한다..

2. 데이터 수집 및 분석

American Express – Default Prediction

<https://www.kaggle.com/competitions/amex-default-prediction>

train_data.csv - customer_ID 당 여러 명세서 날짜가 있는 학습 데이터

train_labels.csv - 각 customer_ID에 대한 대상 레이블

test_data.csv - 해당 테스트 데이터; 목표는 각 customer_ID에 대한 대상 레이블을 예측하는 것

sample_submission.csv - 올바른 형식의 샘플 제출 파일

3. 프로젝트 계획(일정, 수행과제, 목표 등)

데이터 크기와 변수 분석

데이터 시각화하기, 변수들간의 관계 확인

다양한 모델 만들고 feature 추가해보기