

프로젝트 과정

< 파일럿 프로젝트 기획서 >

주 제	American Express – 카드사 고객의 채무불이행 여부 예측		
분 야	금융 데이터 분석	일 자	2022년 10월 06일
팀 장	임지윤 <pucca4881@naver.com >	팀 원	임지윤

1. 과제 개요

집 앞 편의점을 가든 음식점을 가든 편리함과 효율성을 추구하는 현대인의 삶에 있어서 신용카드 없는 삶은 안될 생활 필수품으로 자리잡았다. 이에 따라 청구한 금액을 갚을지 여부(신용 불량)를 예측하는 것이 카드사들에게 있어 매우 중요한 솔루션이다. 신용 불량 예측을 통해 대출 결정을 최적화할 수 있으며, 이는 더 나은 고객 경험과 비즈니스 경제로 이어진다.

America Express(Amex)는 신용카드, 여행자 수표 등을 다루는 미국계 다국적 금융 서비스 기업이다. 기존의 신용불량 예측은 대출 위험관리에 초점을 두고 있다. 기존의 예측모델보다 성능이 좋은 머신러닝 기법을 적용한 신용 불량 예측 모델을 통해 더 나은 고객 경험을 제공하는 것이 이 대회의 목표이다.

2. 데이터 수집 및 분석

- 데이터 셋 URL

[American Express - Default Prediction | Kaggle](#)

- 데이터 셋 출처

[American Express - Default Prediction | Kaggle](#)

3. 프로젝트 계획(일정, 수행과제, 목표 등)

- [예] 데이터 수집/탐색/시각화, 인사이트 얻기/데이터모델링/모델 평가/결론

[목표]

Kaggle (American Express)을 처음부터 끝까지 따라가면서 대용량 데이터를 압축 및 처리하는 방법을 배우고, 다양한 머신러닝 기법을 적용 및 학습하여 향후 고객의 채무불이행(신용 불량) 여부를 예측하는 모델을 구축한다.

[역할 분담 및 일정 계획]

• 22/10/07: 데이터 이해, 데이터 압축 개념 및 코드 이해

데이터 타입 변환 및 파일 형식(Feder, Parquet, Pickle) 선택 후 청크 단위로 처리.

• 22/10/08 ~ 22/10/15: 데이터 탐색 (EDA) 및 전처리

결측값 확인 및 처리, 시각화.

변수들간 상관관계 확인 및 분포 시각화.

- 22/10/16 ~ 22/10/23 : 모델 구축 및 개선을 통해 최적의 모델 선정, ppt
Decision Tree > Random Forest > Gradient Boosting > XGBoost > CatBoost
변수 중요도 시각화.

- 22/10/24 : 최종 발표