

Classifying Credit Card Users :

A Machine Learning Analysis

Zero-base 1팀 / 김경훈, 목해민, 안선경, 윤세종, 이선명

프로젝트 배경

신용카드 사용률 증가

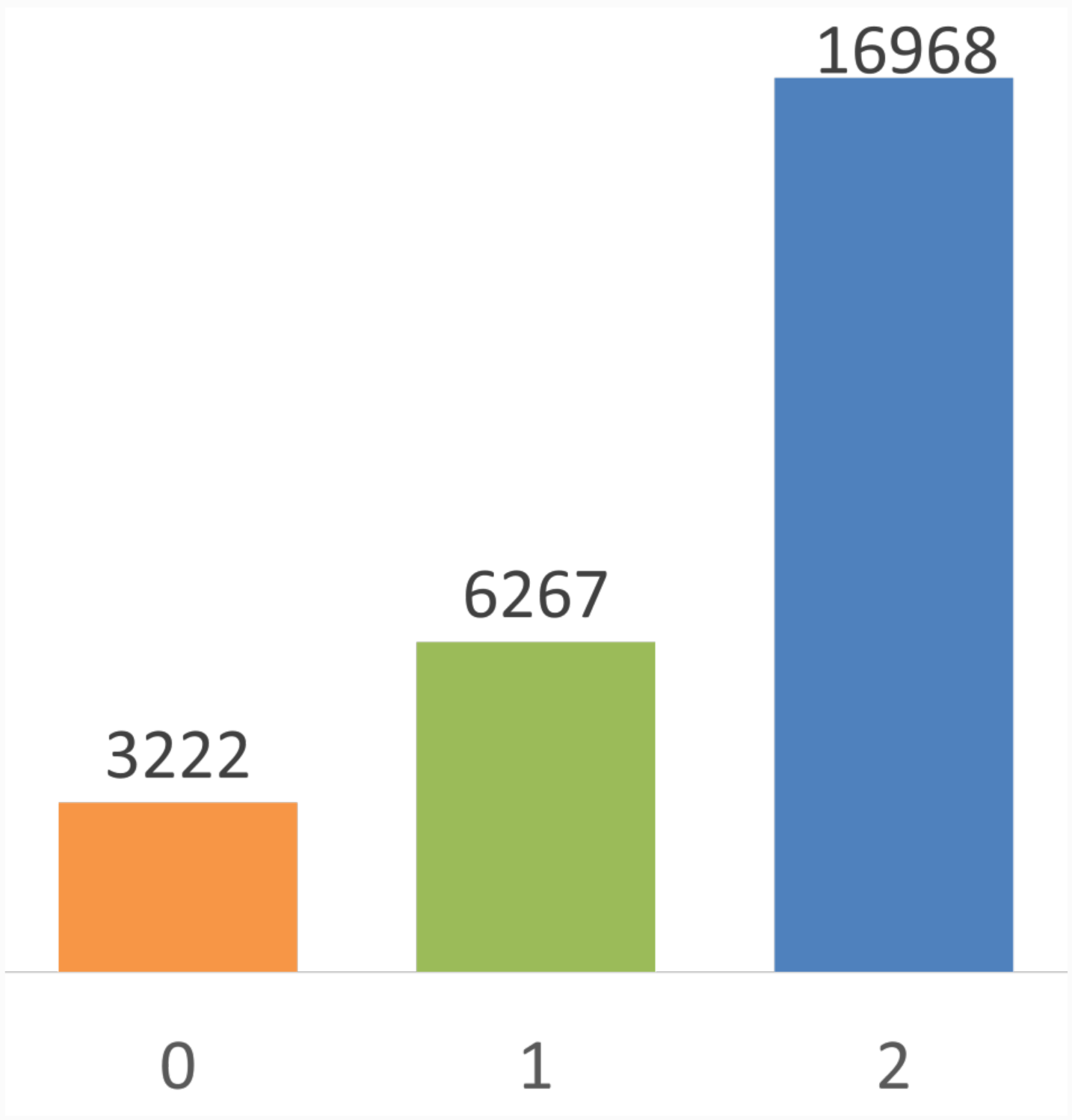
금리 상승, 신용카드 연체율 급증으로 인한

신용카드사의 자산건전성 약화와 위험부담 증가

후불 결제 시장의 성장으로 신용카드 연체율 관리의 중요성 증가

프로젝트 목표

카드 대금 연체 집단의 정보를 통해 연체 정도 예측 알고리즘을 개발하고
건전한 금융시장 유지에 도움이 되는 인사이트 제공



신용등급 분포

낮을수록 높은 신용

Classifier

CatBoost

주요 처리 사항

- ▶ occpy_type NaN 처리
- ▶ 중복데이터 제거
- ▶ Clustering / PCA 적용
- ▶ 연속형 변수 단위 변환 미적용

Feature Engineering

Income_total, edu_type,
family_type, house_type,
day_birth, day_employed,
begin_month, income_occpy,
car_reality,
car_begin_before_employed

최종모델				
Recall_0	Recall_1	Recall_2	Accuracy	Log_Loss
0.25	0.44	0.95	0.77	0.60

시행착오

파생변수 생성 : 성능 하락

XGBoost, LightGBM, CatBoost, RandomForest, DecisionTree 중
CatBoost 에서 제일 좋은 성능 확인

중복 데이터 제거 : 성능 향상

PCA 를 통해 차원을 축소하고 Clustering 후 성능 확인

SMOTE 적용 : 신용도 0과 1에 대한 recall은 상승했으나
가장 중요하게 고려하는 신용도 2에 대한 recall은 하락