

# 핀테크 2021 조별 프로젝트 설명

## 프로젝트 설명

### 프로젝트 목표

- 기계학습 수업 시간에 배운 여러 알고리즘이 실제 데이터에 어떻게 적용될 수 있는지 알아보고, 각 모델 별 차이를 알아보는 시간을 가지는 데 목표가 있습니다.
- 수업시간에 배웠던 Ensemble 방법론처럼 다양한 분야의 학생들이 모였으니, 팀원들간의 충분한 토의를 통해 더 나은 결과물을 보여주세요!

### 프로젝트 기간

- 02.09(화): 프로젝트 설명
- 02.22(월): 프로젝트 발표 (발표 + Q&A 30분)
- 03.01(월): 프로젝트 보고서 제출

### 프로젝트 평가

- 교수님, 조교, 상호 평가

## 프로젝트 구성

아래는 예시일 뿐이며, 다른 순서와 내용으로 프로젝트를 구상하셔도 무방합니다.

### Data Selection

- kaggle, 기타 public에 공개된 데이터 사용 (크롤링 가능)
- 데이터 설명과 택한 이유 등 설명
- visualization 등등

### Preprocessing

- NaN 제거
- 이상치 제거
- 원하는 데이터 추출
- 데이터의 개수가 맞지 않을때 불균형 해소
  - RUS: undersampling (majority class에서 random하게 데이터를 지워주는 방법)
  - SMOTE: oversampling (minority class 데이터에서 주변 k개의 데이터 간의 간격 내에서 데이터를 새로 생성하는 방법)

### Modeling

- 수업 시간에 배운 linear regression, logistic regression, SVM, Tree, Ensemble, Random Forest, boosting (XGBoosting) 등을 사용해 데이터에 적합한 모델 선택
- generalization 성능이 좋은 모델이 보편적으로 좋은 모델

### Result & Discussion

- 얻은 결과와 이에 대한 해석 공유

## 참고 자료

### 작년 프로젝트

2020 핀테크 기계학습 프로젝트 중 선정된 3가지 프로젝트를 github로 공유하니 참고하시기 바랍니다.

- <https://github.com/SLCFLAB/FinTech2021/tree/main/Projects>

프로젝트를 하다 궁금한 점이나 해결되지 않는 부분이 있으면 조교 박진성(jinseong@snu.ac.kr)로 보내주세요!