

# 우리은행 조별 프로젝트 설명

## 프로젝트 설명

### 프로젝트 개요

- 6월 13일, 6월 14일에 PPT 발표를 목표로 조별로 PPT, 코드를 제출합니다.
- 주제는 머신러닝/딥러닝 모두 가능하나, 금융과 관련있는 주제와 데이터셋을 활용해야 합니다. (비전 불가능)
- 딥러닝 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!
- Colab에서 런타임을 GPU로 바꾸어서 사용하시길 권장드리며 개인 GPU가 있으시다면 GPU를 사용하셔도 무방합니다.

### 프로젝트 기간

- 5월 중순 : 수업 및 실습 마무리, 프로젝트 설명 & 프로젝트 주제 정하기
- 6월 13일, 14일: 프로젝트 발표 및 코멘트 (성적 평가해서 분석방법론 성적 최종 마무리)
- 발표 이후: 코멘트를 바탕으로 업그레이드 해서 제출, 수료식 때 몇 조 뽑아서 발표 예정

### 프로젝트 평가

- 발표: 발표자료(5), 발표자의 태도(3), 질의 답변 및 발표시간 준수(2) 에 따라 교수님 , 조교 평가 (상호평가 포함 가능)

## 프로젝트 구성 (머신러닝)

아래는 예시일 뿐이며, 다른 순서와 내용으로 프로젝트를 구상하셔도 무방합니다.

### Data Selection

- kaggle, 기타 public에 공개된 데이터, 제공된 개인신용정보 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

## 프로젝트 구성 (딥러닝)

아래는 예시일 뿐이며, 다른 순서와 내용으로 프로젝트를 구상하셔도 무방합니다.

## Data Selection

- kaggle, 기타 public에 공개된 데이터, 제공된 개인신용정보 데이터 사용 (크롤링 가능)
- 데이터 설명과 택한 이유 등 설명
- visualization 등등
- 딥러닝은 기계학습에 비해 training time이 훨씬 오래 걸립니다. 따라서, 너무 큰 데이터를 고르면 학습에 너무 많은 시간이 걸릴 수 있기 때문에 적당한 데이터를 고르길 바랍니다.

## Preprocessing

- NaN 제거
- 이상치 제거
- 원하는 데이터 추출
- 데이터의 개수가 맞지 않을때 불균형 해소

## Modeling

- CNN(이미지), RNN(시계열), GAN, Autoencoder 등등 사용 무방합니다.
- 여러 모델을 사용해보거나, 만일 이미지 데이터를 골라 CNN을 고른다면 LeNet, ResNet 등을 여러 모델을 바꿔가며 실험을 진행하여 결과를 확인해도 무방합니다.

## Result & Discussion(\*\*)

- 얻은 결과와 이에 대한 해석 공유
- 어떤 모델이 잘 작동하는지, 왜 잘 작동하는지에 대한 해석
- 딥러닝의 결과가 데이터에 대해서 어떤 부분을 설명해주는 지에 대한 해석

## 참고 자료

- 참고자료 설명