

# 방학세미나

즐거운 학회장팀  
염예빈  
이지연

# INDEX

---

1. 주제 및 팀 소개
2. 평가 요소
3. 제출 방법
4. 상품

# 1

주제 및 팀 소개

## | 주제 소개

데이터 : Customer Transaction Prediction data

주제 : 주어진 변수들을 통해 성능이 좋은 예측모델 만들기(이진분류)

- 고객이 거래를 할 것인지 파악함으로써 기업의 의사결정을 도움
- 데이터 불균형 상황. 1과 0이 1:9
- 데이터 변수가 매우 많음
- 데이터 변수들은 모두 익명처리 되어있음

## 데이터 소개

Train : 28000 \* 201, Masked Variables

Test : 2000 \* 200, Masked Variables

200000 \* 201개였는데 사이즈 많.이. 줄였습니다~!

컴퓨터 파워를 고려하여,, 열리지 않은 컴퓨터도 있을까봐,, ㅎㅎ  
모두들 파이팅><

## | 변수 소개

변수의 의미는 아무도 몰라요!!  
처음부터 익명처리된 변수였어요!!

## 평가지표

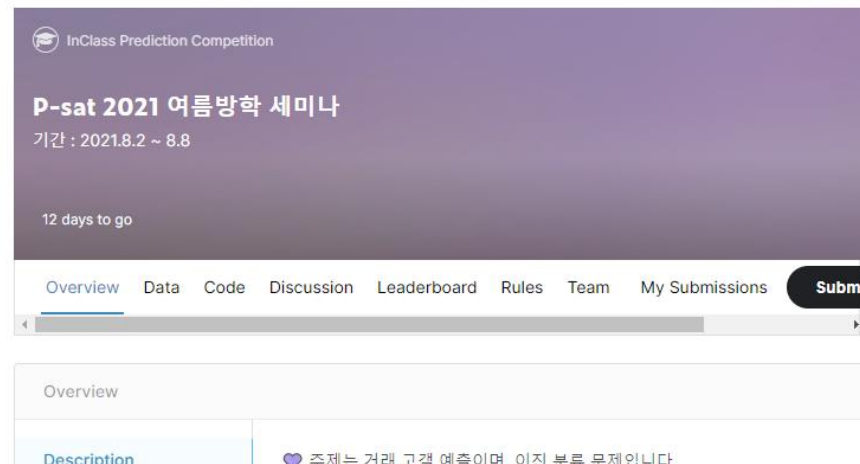
F1 스코어 :

그래도 모른다면 범주팀 클린업 참고!

정밀도와 재현율을 결합한 지표이며, 정밀도와 재현율이 어느 한쪽으로 치우치지 않는 수치를 나타낼 때 상대적으로 높은 값을 가짐

$$F1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = 2 * \frac{precision * recall}{precision + recall}$$

## | 캐글 Kaggle



캐글에 점수를 산출할 수 있도록 컴피티션을 열어졌습니다.

F1-score로 50% test셋에 대하여 리더보드 채점할 수 있습니다.

최종 점수는 제출된 Submission.csv에 대하여 채점될 예정입니다.

리더보드는 참고용으로 사용하세요!

<https://www.kaggle.com/t/34f675f93ea44a48a02b09447bd1521e>



## | 팀 소개

## Team 1

이진모  
문병철  
이수정

조혜현  
홍현경

## Team 2

주혜인  
이정우

반경림  
심예진

## Team 3

이승우  
위재성  
임주은

고경현  
박세령

## Team 4

조장희  
김지현  
전효림

이재현  
이은서

최대한 기존팀원들이 겹치지 않게 배정했습니다!  
코로나때문에 서로의 교류가 많이 없어서 서로 알아가는 시간이 되었으면 좋겠습니다><

# 2

평가 요소

평가기준	항목	상세	배점
CODE	성능	F1 스코어	5
	재현성	코드가 재현되는가?	3
	가독성	코드가 깔끔히 정리되어있는가?	3
	시간	Predict 시간	2
Analysis	EDA	변수 의미 파악 및 해석 정도	3
	명확성	분석흐름에 대한 논리와 명확성	5
	Data Leakage	테스트에 대한 정보 사용불가	3
PPT	논리성	분석 흐름에 대한 논리적 명확성	2
	심미성	디자인, 미적구성 ( 보는 사람이 불편하지않게 PPT를 잘 구성하였는가?)	2
제출기한	지각	제출시간 준수(기본 2점)	2
총점		30	

## 성능

Test Set에 대한 평가지표  
(F1스코어)를 계산

클래스 1에대한 F1스코어로 평가할거예요!

## 재현성

전처리 및 모델링 코드 재현성

## 가독성

코드 가독성 및 효율성

코드를 읽었을 때 가동성 및 깔끔함  
주석 잘 달아주셔도 좋아요!-!

## 시간

모델 부르기 및  
Predict 시간 측정

## EDA

변수간의 관계 및 의미파악  
시각화 및 해석

변수가 익명처리 되어 있기에 한계가 많음을  
고려하여 채점할 예정입니다! 너무 힘쓰지 마세요!

## 명확성

분석 흐름에 대한 논리와  
전체적인 흐름의 명확성을 평가

- 데이터 불균형을 위한 적절한 방법을 사용하여 분석하였는지 평가
- 매우 많은 변수의 모델링을 위하여 적절한 방법을 사용하였는지 평가

## Data Leakage

Test 데이터에 대한 정보 유출X

Ex) 전처리/ 모델링 과성에서 train/test셋을 합쳐서 전체 데이터에 동일한 PCA X

## 논리성

분석흐름에 대한  
논리적 명확성

## 지각

제출시간을 잘 지켰는지 평가할 예정입니다.

1시간에 1점씩 감점

## 심미성

PPT에 대한 디자인, 미적구성을 평가

예쁜 PPT가 아닌  
보는 사람이 불편하지 않게 PPT를 잘 구성하고  
만들었는지 평가할 예정입니다.

언제든 어렵거나 궁금한 것은  
물어보세요! ~0~

# 3

## 제출 방법

## Leader Board

### 리더보드

팀	점수
1팀	0.2406
2팀	0.2507
3팀	0.2888

제출 가능 횟수 : 2회 / 1일

제출 양식 : id(1:2000)와 target 예측값

### 제출 방법

학회장에게카톡 제출!

Process/Modeling/Predict코드 + 최종 예측 결과 (csv) + PPT 를  
묶어서 zip파일로 제출하시면 됩니다!



## Leader Board

### 리더보드

id	target
1	0
2	0
3	0
4	1
5	0
6	0
7	0
8	0
9	1
10	0

점수

0.2406

0.2507

0.2888

제출 양식: Submission.csv 형식

다음 날 랜덤 시간 중으로 업데이트

id와 target 예측값

### 제출

학회 홈페이지 '드라이브' 에 파일 제출

Ex ) 08<예시>1팀\_1회.csv

## | 최종 제출, 세미나 및 결과 발표

### 최종 제출

2021년 8월 8일, 23시 59분까지 제출

### 세미나

2021년 8월 9일 오후 6시  
발표자는 5시에 랜덤으로 뽑을 예정  
발표시간 : 10분

### 결과 발표

2021년 8월 9일 : 방학세미나 (오후 6시:온라인)에 발표

4

상품

## | 상품

1등

힘난한 채점을 거쳐 1등을 한 팀에게는 소정의 기프티콘을 드립니다!

모두 파이팅!!  
더위, 코로나 조심하세요오!!



THANK YOU

