

방학세미나



불꽃 학회장팀

박이현 홍지우

INDEX

1. 주제 및 팀 소개
2. 평가 기준
3. 제출
4. 발표

1

주제 및 팀 소개

TOPIC

주어진 데이터를 활용하여 성능이 좋은 이진 분류 모델 만들기

DATA

Train : 416648*59

Test : 178564*58

변수명은 확인 불가하며, 자료형만 확인 가능합니다!

bin : Binary / cat : Categorical / 그 외는 Numeric 또는 Ordinal

평가지표

F1 Score

더 자세하게 알고 싶다면 2022-1학기 범주팀 클린업 3주차 참고!

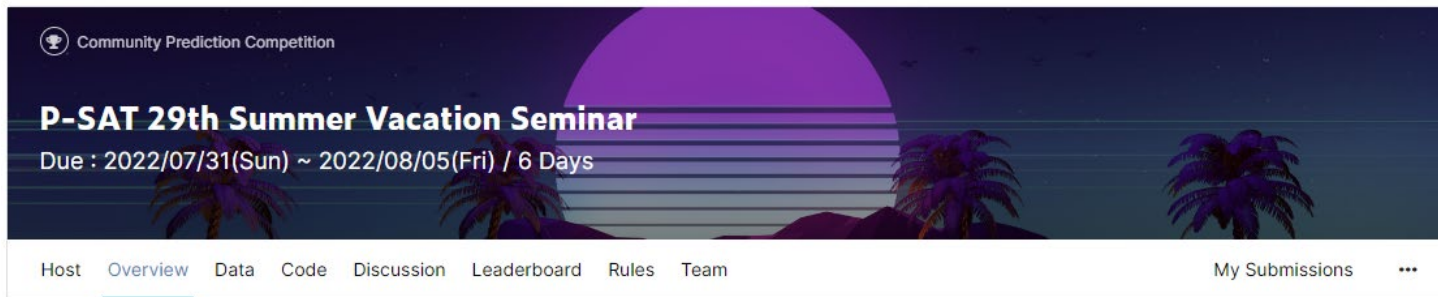
: 정밀도와 재현율의 조화평균

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

분류 모델의 성능을 평가하는 지표로 사용

진행방식

kaggle



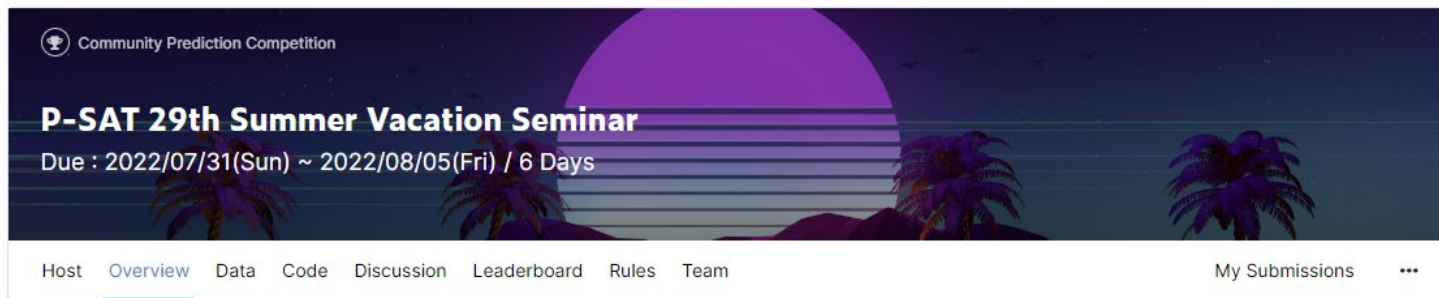
<https://www.kaggle.com/t/9ec2e19bedc44a8eaf607502c8d1b0c3>

캐글에서 모델의 성능을 파악할 수 있도록 컴피티션 개최
Public leader board에서 50%의 test set에 대해 채점 가능
최종 점수는 제출된 submission.csv의 F1 Score로 평가

계정은 1인당 1개만 사용 가능합니다!

진행방식

kaggle



<https://www.kaggle.com/t/9ec2e19bedc44a8eaf607502c8d1b0c3>

제출 가능 횟수 : 1일 3회

Leaderboard를 통해 Score를 실시간으로 확인 가능
가장 Score가 높은 팀원이 속해있는 팀이 Competition 1등

팀 소개

1팀

조웅빈 윤지영 이지윤 박시언 김민서

2팀

김예찬 유종석 김준서 윤경선 김영호

3팀

정희철 김현우 이주형 채소연 서희나 박윤아

2

평가 기준

평가기준	항목	상세	배점
CODE	성능	F1 Score	5
	재현성	코드가 재현되는가?	3
	가독성	코드가 깔끔히 정리되어 있는가?	3
	시간	Predict 시간	2
Analysis	EDA	시각화를 통한 데이터의 해석 및 인사이트 도출	6
	명확성	분석흐름에 대한 논리와 명확성	5
	참신성	데이터 처리 과정 및 사용 모델의 참신성	3
	Data Leakage	테스트 셋에 대한 정보 사용불가	2
PPT	논리성	분석 흐름에 대한 논리적 명확성	2
	심미성	디자인, 미적구성 (PPT의 가독성과 내용 구성이 적절한가)	2
제출기한	지각	제출시간 준수(기본 2점)	2
총점		35	

✓ 성능

Test Set에 대한 평가지표
F1 Score 평가

✓ 재현성

전처리 및 모델링 코드 전반의 재현성
모든 과정이 동일하게 재현되어야 합니다.
랜덤씨드 설정에 유의해주세요

✓ 가독성

코드의 가독성 및 효율성
단계별로 다른 코드 파일을 생성해도 좋으며,
한 코드 파일에 모두 담아도 좋습니다.

✓ 시간

모델 학습 / 모델 예측 시간 측정
코드 파일 내에 아래 2가지가 명시적으로 측정되어야 합니다.
① Train 데이터에 대한 학습 시간
② Test 데이터에 대한 예측 시간

✓ EDA

시각화를 통한 변수 특성 파악,
변수와 모델의 해석을 통한
인사이트 도출 과정 전반

✓ 명확성

분석 전개 과정에서
전반적인 흐름의 논리와 방향 평가

✓ 참신성

변수 선택, NA 처리, 모델 선정
등에서의 참신성 평가

✓ Data Leakage

분석 과정에서 Test Data 관련
정보 사용 금지

✓ EDA

시각화를 통한 변수 특성 파악,
변수와 모델의 해석을 통한
인사이트 도출 과정 전반

변수명이 익명 처리되어 있으므로

변수의 의미를 파악하기보다는, 아래 3가지에 집중해 주세요!

- ① EDA가 논리적으로 사용 및 전개되었는지
- ② 얼마나 다양한 정보를 EDA에 담아낼 수 있는지
- ③ 새로운 인사이트를 도출해 냈는지

✓ 참신성

변수 선택, NA 처리, 모델 선정
등에서의 참신성 평가

✓ Data Leakage

분석 과정에서 Test Data 관련
정보 사용 금지

✓ 명확성

명확성은 아래 2가지를 기준으로 평가할 예정입니다.

- ① EDA에서 도출된 인사이트가 모델링에서 적절히 활용되었는지
- ② 데이터의 특성을 반영한 모델을 적절하게 선정했는지

분석 전개 과정에서
전반적인 흐름의 논리와 방향 평가

✓ 참신성

변수 선택, NA 처리, 모델 선정
등에서의 참신성 평가

✓ Data Leakage

분석 과정에서 Test Data 관련
정보 사용 금지

✓ EDA

시각화를 통한 변수 특성 파악,
변수와 모델의 해석을 통한
인사이트 도출 과정 전반

✓ 명확성

분석 전개 과정에서
전반적인 흐름의 논리와 방향 평가

✓ 참신성

변수 선택, NA 처리, 모델 선정
등에서의 참신성 평가

1학기 각 팀의 클린업과 주제분석에서 다뤄진
전처리 과정, 모델, 변수 선택법, NA 처리 방법 등을
얼마나 잘 활용했는지를 위주로 평가할 예정입니다.

✓ EDA

시각화를 통한 변수 특성 파악,
변수와 모델의 해석을 통한
인사이트 도출 과정 전반

매우 당연하게도, 전처리부터 모델링 과정 전반에서
Test 데이터와 Test 데이터로부터 도출된 정보를
절대 사용해서는 안 됩니다.
Ex) Scaling, PCA, ...

✓ 명확성

분석 전개 과정에서
전반적인 흐름의 논리와 방향 평가

✓ Data Leakage

분석 과정에서 Test Data 관련
정보 사용 금지

✓ 논리성

분석 흐름 소개 과정의
논리적 명확성

✓ 심미성

PPT의 가독성 및 구성

✓ 지각

제출 기한 엄수

3

제출

제출 방법

1) Leader board

- 1일 최대 제출 가능 횟수 : 3회
- 제출 양식 : id(1:178564)와 target 예측값

id	target
1	0
2	0
3	1
...	...
178562	0
178563	1
178564	0

2) 학회장 제출

- 팀원 중 1인이 대표로 학회장에게 카톡으로 파일 제출
- 제출 파일 : PREPROCESSING/MODELLING/PREDICT 코드 파일
+ 최종예측결과 csv 파일

<R, Python 다 가능>

제출 방법

1) Leader board

- 1일 최대 제출 가능 횟수 : 3회
- 제출 양식 : id(1:178564)와 target 예측값

id	target
1	0
2	0
3	1
...	...
178562	0
178563	1
178564	0

2) 학회장 제출

- 팀원 중 1인이 대표로 학회장에게 카톡으로 파일 제출

<R, Python 다 가능>

- 제출 파일 : PREPROCESSING/MODELLING/PREDICT 코드 파일
+ 최종예측결과 csv 파일

하나의 zip파일로 제출

제출 기한

<div>8</div> <div>2022 August</div>						
SUN	MON	TUE	WED	THU	FRI	SAT
	01 음7.4	02 음7.5	03 음7.6	04 음7.7	05 음7.8	06 음7.9
07 음7.10	08 음7.11	09 음7.12	10 음7.13	11 음7.14	12 음7.15	13 음7.16
14 음7.17	15 음7.18	16 음7.19	17 음7.20	18 음7.21	19 음7.22	20 음7.23
	광복절					
21 음7.24	22 음7.25	23 음7.26	24 음7.27	25 음7.28	26 음7.29	27 음8.1
28 음8.2	29 음8.3	30 음8.4	31 음8.5			

2022년 8월 4일 (목), 11시 59분까지 제출

<상대적으로 짧은 방세기간 고려...>

4

발표

세미나

1) 일시

- 2022년 8월 5일(금) 오후 6시 오프라인 진행
- 경영관 33201호로 시간에 맞춰서 와주시면 됩니다.
- 오후 4시부터 대여를 해 났으니 미리 오셔서 준비하셔도 됩니다.

2) 발표자 선정

- 발표자 및 발표 순서는 5시에 랜덤으로 선정 후 공지
- 발표시간 : 팀당 10분

3) 결과 발표 및 시상

- 2022년 8월 5일 : 각 팀 발표 종료 후 시상
- 1등 팀은 소소한 상품과 방세 1등 팀이라는 큰 명예를 얻을 수 있습니다!



THANK YOU

