

## CUAI 프로젝트 DA 1팀

2024.10.29

발표자 : 이해원

## 스터디원 소개 및 만남 인증

사진을 못 찍었습니다..!

김민하

성산해

안상우

이승현

이혜원

## 목차

프로젝트 소개

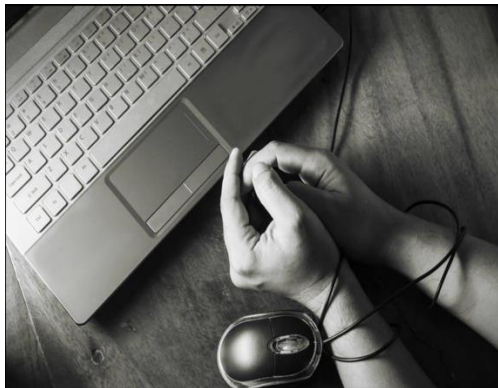
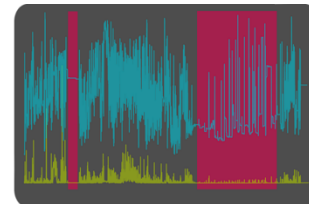
프로젝트 진행계획

## 프로젝트 진행 계획

DAICON

### Child Mind Institute — Problematic Internet Use

Relating Physical Activity to Problematic Internet Use



Start

a month ago

Close

2 months to go

Deadline: Dec 13, 2024

## 프로젝트 소개(내용)

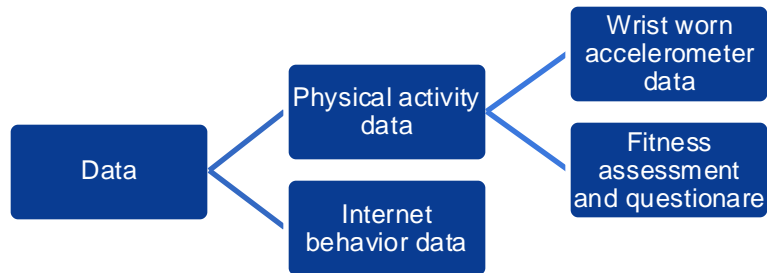
어린이의 신체 활동 및 체력 데이터를 활용한 문제적 인터넷 사용 초기 징후 식별 AI 모델 개발







Target Variable : SSI(Severity Impairment Index)

- 문제적 인터넷 사용의 표준 지표
- 0: 없음, 1: 경미, 2: 중간, 3: 심각

```
id, sii
000046df, 0
000089ff, 1
00012558, 2
00017ccd, 3
...
```

## 데이터 구조



- ▶  series\_test.parquet
  - ▶  series\_train.parquet
  -  data\_dictionary.csv
  -  sample\_submission.csv
  -  test.csv
  -  train.csv
- 단일 참가자의 가속도계로 측정한 시계열 데이터
- Train, test 피쳐들의 추가적인 설명, 및 정보
- 참가자의 나이, 성별, 일일 인터넷 사용 시간, 신체 정보(혈압, 심박수 등), 체력 및 수면 관련 평가, 그리고 인터넷 중독 성향(PCIAT 점수) 등의 임상 데이터가 포함

## 평가 지표

### quadratic weighted kappa

예측 값과 실제 값 간의 **일치도를 측정**하는 지표  
QWK의 값은 -1에서 1 사이

- 1: 완벽한 일치 (예측 값과 실제 값이 완전히 같다)
- 0: 무작위적인 일치 (우연에 의한 일치와 같음)
- 0보다 작음: 무작위보다도 더 낮은 일치도 (예측이 실제와 거의 정반대라고 볼 수 있음)

Formula

$$W_{i,j} = \frac{(i-j)^2}{(N-1)^2}$$





$$\kappa = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}}$$

## 주목할(유의할 점)













1. 1000명 가량의 시계열 데이터 처리 방법
  - 6.3gb이므로 어떻게 전처리, 사용할 지 고민해야
2. Target variable인 SII가 train 데이터에 일부 누락됨
  - 비지도학습, 반지도학습 고려

### Data Explorer

6.73 GB

- ▼  series\_test.parquet
  - ▶  id=00115b9f
  - ▶  id=001f3379
- ▶  series\_train.parquet

series\_train.parquet (996 directories)

 id=00115b9f 1 files	 id=001f3379 1 files	 id=00f332d1 1 files	 id=01085eb3 1 files
 id=012cadd8 1 files	 id=012e3669 1 files	 id=029a19c9 1 files	 id=02cebf33 1 files
 id=02cf7384 1 files	 id=035c96dd 1 files	 id=03a9019b 1 files	 id=0417c91e 1 files



## 프로젝트 진행 계획

~ 11/3 데이터 탐색

~ 11/19 EDA(preprocessing)

~ 모델링 및 모델 수정

감사합니다