

kaggle CMI  
178th solution



**Child Mind  
Institute**

村瀬晃

# コンペ概要

装着型デバイスをつけた被験者（10代~20代前半）の事前調査とセンサーから得られる日々の活動量やセンサーに入り込む光の強度からインターネットの依存度合いを推定し、その精度を競うコンペティション

# データ概要

1被験者につき、デバイスから取得されたセンサーデータとメタデータが提供されていた

ラベルはParent-Child Internet Addiction Testというインターネットの中毒性を測るテストから算出されるsiiスコア

sii : 0は「なし」、1は「軽度」、2は「中度」、3は「重度」となる

# 評価指標

quadratic weighted kappa

2 者の一致度合いを測る指標

正解ラベルと予測ラベルの一致度合いを計算する

# データの具体的な中身

全被験者が3960人

ラベルがつけられている被験者が1224人

センサーデータがある被験者が996人

# 仮説

インターネットの依存度合いを推定するためには、その人がどれだけ長期的にインターネットに接続しているデバイスを扱っているかを定量化する必要がある。

センサーデータをもとに、「その明るさの下で取られる行動の統計量」がインターネット中毒者かどうかを識別するのに役立つのではないか

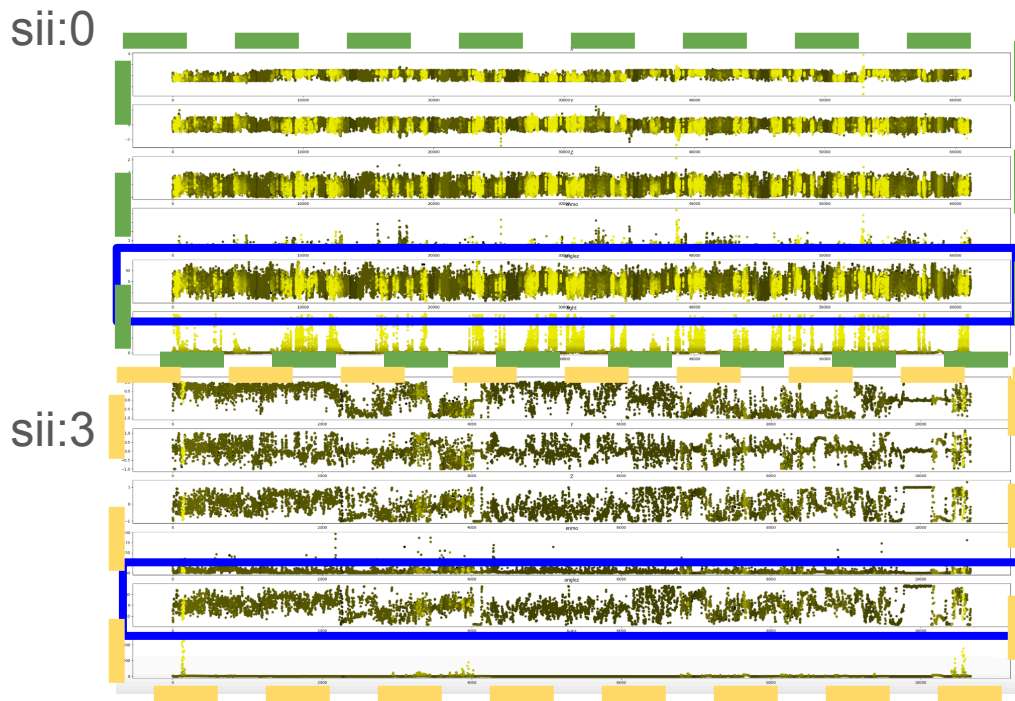
→明るいのに活動量が小さいと、インターネットやテレビなどの電子機器に没頭している可能性が高くスコアが高い傾向になるのではないか。

→明るい時に活動量が大きいと、屋外での活動をしており、少なくともインターネットに依存する確率が低くなるのではないか

# 仮説を裏付ける可視化 1

ある適当なsiiが0の被験者とsiiが3の被験者の2人とを比べたところ以下のことが確認された

- sii3のユーザは比較的、暗い場所で活動している
- 加速度で見比べた時にsii=0のユーザは明るい場所で値が分散している

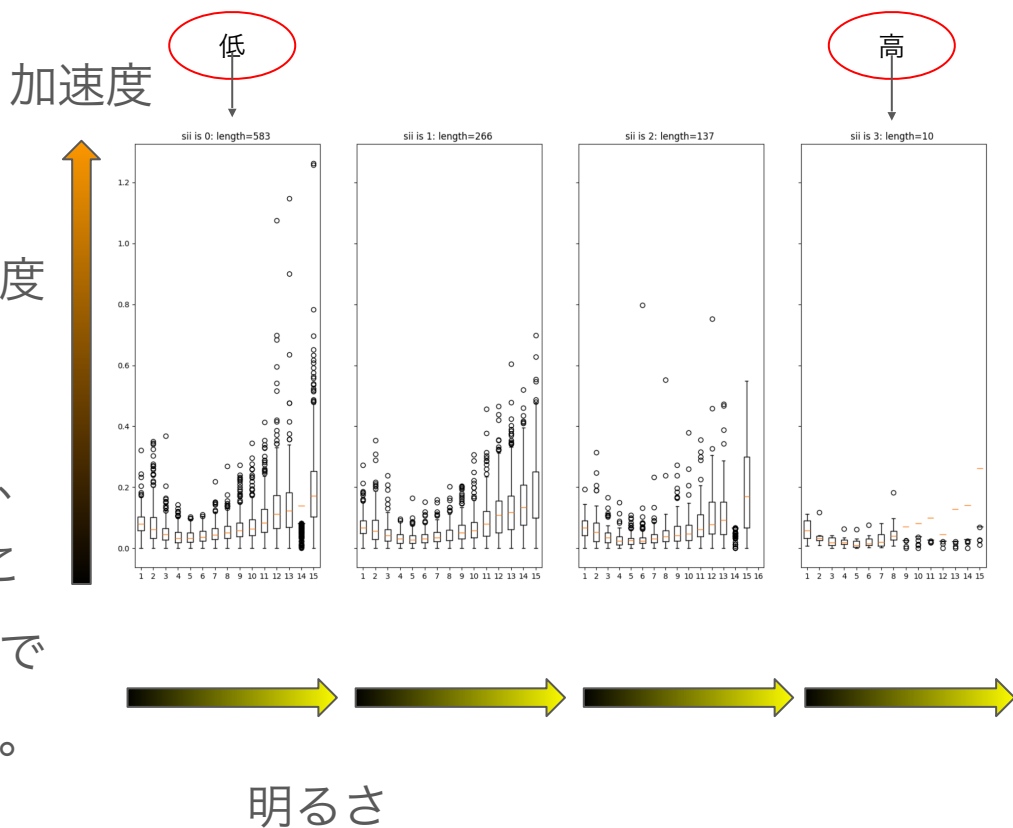


黄色の濃さが強いほど明るい場所にいることを表している  
青枠は加速度の時系列プロット  
上からx,y,z,enmo(加速度),lightを時系列に並んでいる

## 仮説を裏付ける可視化 2

明るさを16カテゴリに分類し、  
各ラベル群ごとにその明るさでの加速度  
(enmo)の平均値を並べた図

siiが高くなるに従って明るい場所では、  
enmoの平均値の平均が下がっているこ  
とから、siiが高い被験者は明るい場所  
では活動量が低下していると考えられる。



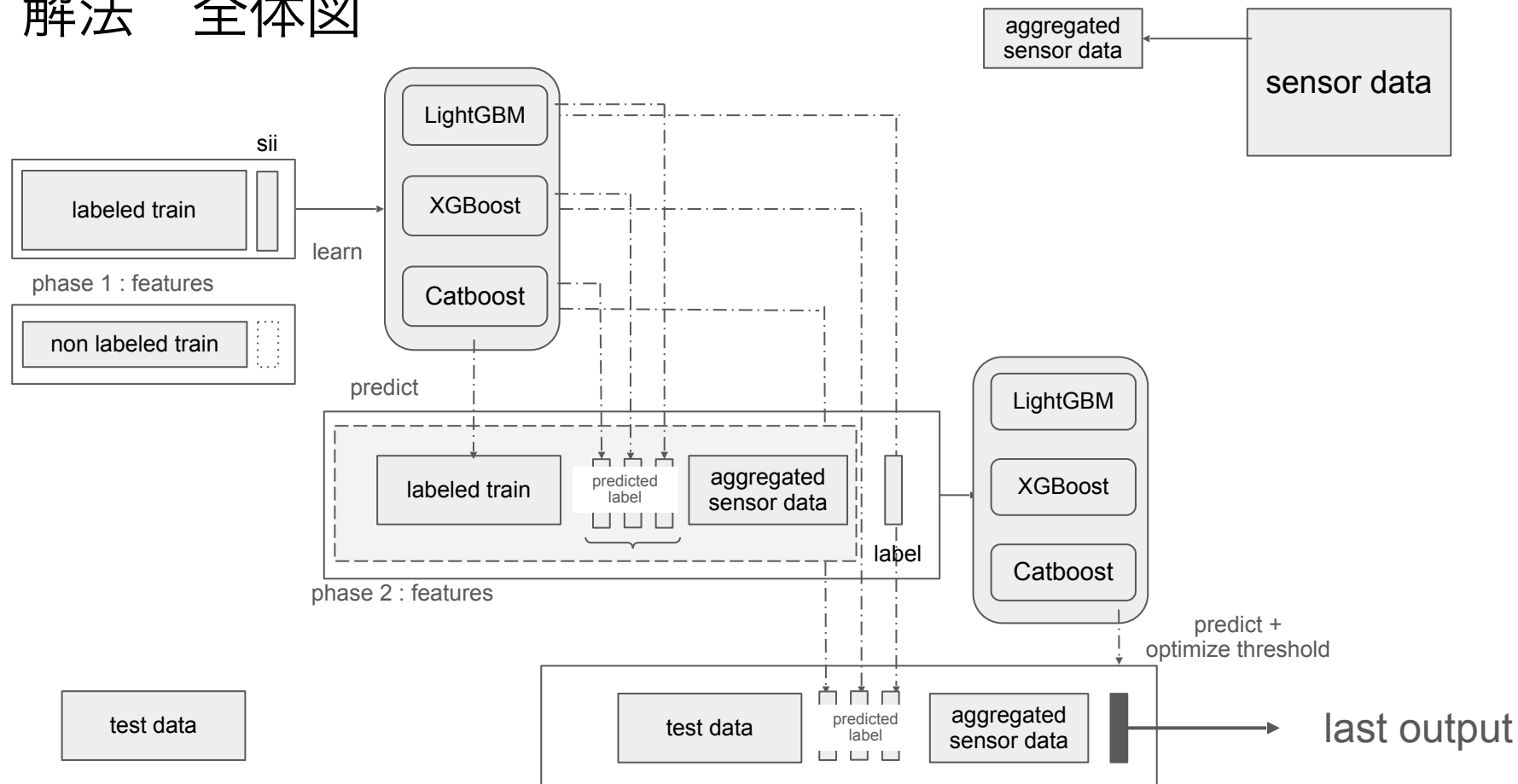


# アプローチ

センシングされた明るさ（lux）から18段階のlux categoryを割り出し、各lux category内での活動量の統計量を算出し、モデルの特徴量とした

xgboost, catboost, lightgbmを用いた2階層のstacking modelingで予測

# 解法 全体図



# 結果

スコア 0.448で178位だった

seedを変更して最終アプローチの精度も確認した