Child Mind Institute — Problematic Internet Use

Description

In today's digital age, problematic internet use among children and adolescents is a growing concern. Better understanding this issue is crucial for addressing mental health problems such as depression and anxiety.

Current methods for measuring problematic internet use in children and adolescents are often complex and require professional assessments. This creates access, cultural, and linguistic barriers for many families. Due to these limitations, problematic internet use is often not measured directly, but is instead associated with issues such as depression and anxiety in youth.

Conversely, physical & fitness measures are extremely accessible and widely available with minimal intervention or clinical expertise. Changes in physical habits, such as poorer posture, irregular diet, and reduced physical activity, are common in excessive technology users. We propose using these easily obtainable physical fitness indicators as proxies for identifying problematic internet use, especially in contexts lacking clinical expertise or suitable assessment tools.

This competition challenges you to develop a predictive model capable of analyzing children's physical activity data to detect early indicators of problematic internet and technology use. This will enable prompt interventions aimed at promoting healthier digital habits.

Your work will contribute to a healthier, happier future where children are better equipped to navigate the digital landscape responsibly.

Dataset Description

The Healthy Brain Network (HBN) dataset is a clinical sample of about five-thousand 5-22 year-olds who have undergone both clinical and research screenings. The objective of the HBN study

is to find biological markers that will improve the diagnosis and treatment of mental health and learning disorders from an objective biological perspective. Two elements of this study are being used for this competition: physical activity data (wrist-worn accelerometer data, fitness assessments and questionnaires) and internet usage behavior data. The goal of this competition is to predict from this data a participant's **Severity Impairment Index** (`sii`), a standard measure of problematic internet use.

Note that this is a **Code Competition**, in which the actual test set is hidden. In this public version, we give some sample data in the correct format to help you author your solutions. The full test set comprises about 3800 instances.

The competition data is compiled into two sources, `parquet` files containing the accelerometer (actigraphy) series and `csv` files containing the remaining tabular data. The majority of measures are missing for most participants. In particular, the target `sii` is missing for a portion of the participants in the training set. You may wish to apply non-supervised learning techniques to this data. The `sii` value is present for all instances in the test set.

HBN Instruments

The tabular data in **train.csv** and **test.csv** comprises measurements from a variety of instruments. The fields within each instrument are described in **data_dictionary.csv**. These instruments are:

- `Demographics` - Information about age and sex of participants.
- `Internet Use` - Number of hours of using computer/internet per day.
- `Children's Global Assessment Scale` - Numeric scale used by mental health clinicians to rate the general functioning of youths under the age of 18.
- `Physical Measures` - Collection of blood pressure, heart rate, height, weight and waist, and hip measurements.
- `FitnessGram Vitals and Treadmill` - Measurements of cardiovascular fitness assessed using the NHANES treadmill protocol.
- `FitnessGram Child` - Health related physical fitness assessment measuring five different parameters including aerobic capacity, muscular strength, muscular endurance, flexibility, and body composition.
- `Bio-electric Impedance Analysis` - Measure of key body composition elements, including BMI, fat, muscle, and water content.
- `Physical Activity Questionnaire` - Information about children's participation in vigorous activities over the last 7 days.
- `Sleep Disturbance Scale` - Scale to categorize sleep disorders in children.
- `Actigraphy` - Objective measure of ecological - physical activity through a research-

grade biotracker.

- `Parent-Child Internet Addiction Test` - 20-item scale that measures characteristics and behaviors associated with compulsive use of the Internet including compulsivity, escapism, and dependency.

Note in particular the field `PCIAT-PCIAT_Total`. The target `sii` for this competition is derived from this field as described in the data dictionary: `0` for `None`, `1` for `Mild`, `2` for `Moderate`, and `3` for `Severe`. Additionally, each participant has been assigned a unique identifier `id`.

Actigraphy Files and Field Descriptions

During their participation in the HBN study, some participants were given an accelerometer to wear for up to 30 days continually while at home and going about their regular daily lives.

- **series_{train|test}.parquet/id={id}** - Series to be used as training data, partitioned by `id`. Each series is a continuous recording of accelerometer data for a single subject spanning many days.

  - `id` - The patient identifier corresponding to the id field in **train/test.csv**.
  - `step` - An integer timestep for each observation within a series.
  - `X`, `Y`, `Z` - Measure of acceleration, in g, - experienced by the wrist-worn watch along each standard axis.
  - `enmo` - As calculated and described by the wristpy package, ENMO is the Euclidean Norm Minus One of - all accelerometer signals (along each of the x-, - y-, and z-axis, measured in g-force) with negative values rounded to zero. Zero values are indicative of periods of no motion. While no standard measure of acceleration exists in this space, this is one of the several commonly computed features.
  - `anglez` - As calculated and described by the - wristpy package, Angle-Z is a metric derived from individual accelerometer components and refers to the angle of the arm relative to the horizontal plane.
  - `non-wear_flag` - A flag (0: watch is being worn, 1: the watch is not worn) to help determine periods when the watch has been removed, based on the GGIR definition, which uses the standard deviation and range of the accelerometer data.
  - `light` - Measure of ambient light in lux. See here for details.
  - `battery_voltage` - A measure of the battery voltage in mV.
  - `time_of_day` - Time of day representing the start of a 5s window that the data has been sampled over, with format `%H:%M:%S.%9f`.
  - `weekday` - The day of the week, coded as an integer with 1 being Monday and 7 being Sunday.

- quarter - The quarter of the year, an integer from 1 to 4.
  - `relative_date_PCIAT` - The number of days (integer) since the PCIAT test was administered (negative days indicate that the actigraphy data has been collected before the test was administered).

- **sample_submission.csv** - A sample submission file in the correct format. See the Evaluation page for more details.

Acknowledgments

Sponsorship

Dell Technologies and NVIDIA are thrilled to partner with the Child Mind Institute, recognizing the profound impact this collaboration will have on advancing mental health support for children and adolescents. This partnership aligns perfectly with our commitment to leveraging technology for social good and fostering a healthier, more inclusive future.

Dell Technologies AI solutions from desktop to datacenter to cloud. NVIDIA pioneered accelerated computing to tackle challenges no one else can solve. Our work in AI and digital twins is transforming the world's largest industries and profoundly impacting society.

Evaluation

Submissions are scored based on the quadratic weighted kappa, which measures the agreement between two outcomes. This metric typically varies from 0 (random agreement) to 1 (complete agreement). In the event that there is less agreement than expected by chance, the metric may go below 0.

To compute the quadratic weighted kappa, we construct three matrices, O, W, and E, with N the number of distinct labels.

The matrix O is an $N \times N$ histogram matrix such that $O_{i,j}$ corresponds to the number of

instances that have an actual value i and a predicted value j.

The matrix W is an $N \times N$ matrix of weights, calculated based on the squared difference between actual and predicted values:

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

The matrix E is an $N \times N$ histogram matrix of expected outcomes, calculated assuming that there is no correlation between values. This is calculated as the outer product between the actual histogram vector of outcomes and the predicted histogram vector, normalized such that E and O have the same sum.

From these three matrices, the quadratic weighted kappa is calculated as:

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

Submission File

For each `id` in the test set, you must predict the corresponding `sii` (described on the Data page). The file should contain a header and have the following format:

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

Timeline

- **September 19, 2024** - Start Date.

- **December 12, 2024** - Entry Deadline. You must accept the competition rules before this date in order to compete.

- **December 12, 2024** - Team Merger Deadline. This is the last day participants may join or merge teams.

- **December 19, 2024** - Final Submission Deadline.

All deadlines are at 11:59 PM UTC on the corresponding day unless otherwise noted. The competition organizers reserve the right to update the contest timeline if they deem it necessary.

Prizes

- 1st Place - $ 15,000
- 2nd Place - $ 10,000
- 3rd Place - $ 8,000
- 4th Place - $ 7,000
- 5th Place - $ 5,000
- 6th Place - $ 5,000
- 7th Place - $ 5,000
- 8th Place - $ 5,000

Code Requirements

This is a Code Competition Submissions to this competition must be made through Notebooks. In order for the "Submit" button to be active after a commit, the following conditions must be met:

- CPU Notebook <= 9 hours run-time
- GPU Notebook <= 9 hours run-time
- Internet access disabled
- Freely & publicly available external data is allowed, including pre-trained models
- Submission file must be named `submission.csv`

Please see the Code Competition FAQ for more information on how to submit. And review the code debugging doc if you are encountering submission errors.

Citation

Adam Santorelli, Arianna Zuanazzi, Michael Leyden, Logan Lawler, Maggie Devkin, Yuki Kotani, and Gregory Kiar. Child Mind Institute — Problematic Internet Use. https://kaggle.com/competitions/child-mind-institute-problematic-internet-use, 2024. Kaggle.