Systems Analysis & Design

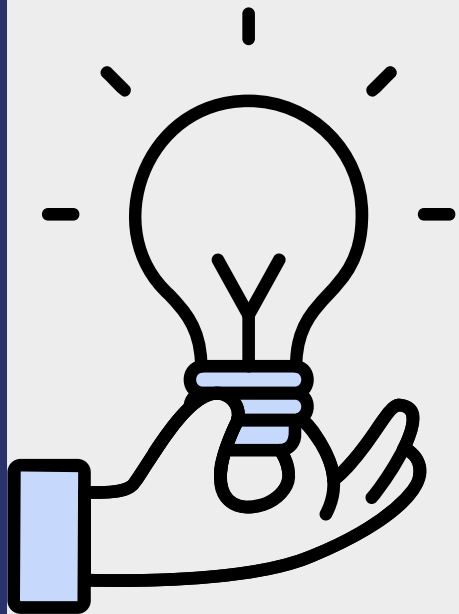# PROBLEMATIC INTERNET USE

CHILD MIND INSTITUTE

Members:
- Bettsy garses
- Isabel sanchez
- Luis rojas

INTRODUCTION

The Problematic Internet Use project originated from the Child Mind Institute competition on Kaggle, the challenge lies in identifying behavioral patterns that can predict problematic Internet use based on physiological and psychological data.

This project applies the principles of systems analysis and design to understand and model this complex phenomenon, employing **systems theory and technological tools** to design a predictive and analytical tool.
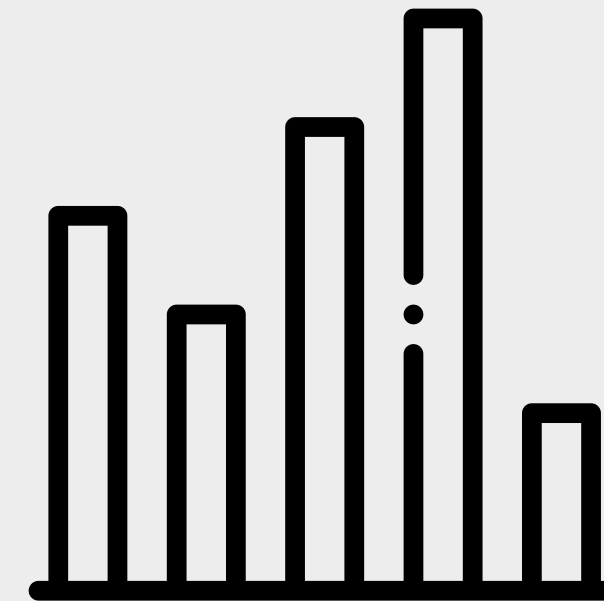
DATASET AND COMPETITION METRIC

TARGET VARIABLE

The target variable is the **Severity of Internet Use (SII)**, categorized into four ordinal levels representing increasing degrees of problematic Internet behavior.
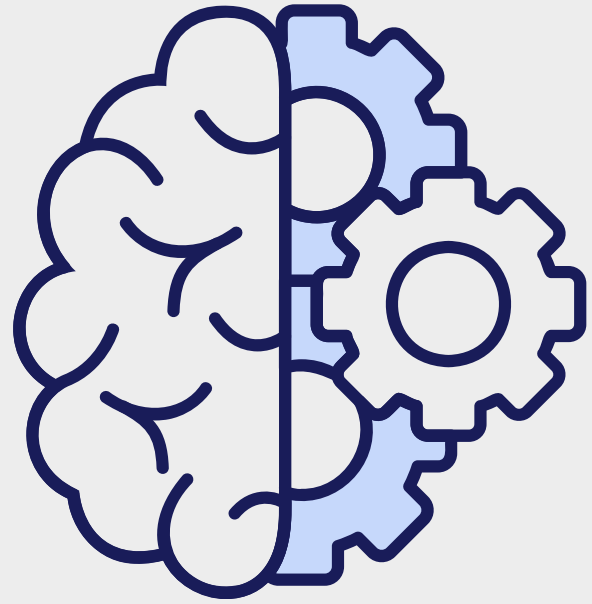
SEMI-SUPERVISED CHALLENGE

The competition constitutes a semi-supervised challenge, characterized by **missing data** and **experimental noise**, requiring the application of **robust preprocessing and modeling techniques** to ensure reliable predictions.

EVALUATION METRIC

The evaluation metric used is the **Quadratic Weighted Kappa (QWK)**, which measures the agreement between predicted and actual classifications. Higher QWK values indicate greater predictive accuracy and consistency.

SUBSYSTEMS WITHIN THE PREDICTIVE FRAMEWORK

Data Collection Subsystem

Involves multiple domains:
- **Biological**: sleep patterns, BMI, blood pressure.
- **Psychological**: anxiety, impulsivity, emotional regulation.
- **Social/Environmental**: family dynamics, school context, accessibility.
- **Technological**: device usage time, type of digital activity.

Modeling Subsystem

It trains and optimizes machine learning models such as CatBoost, or neural networks to predict the severity level of Internet use disorder (IUD).

The subsystem focuses on balancing accuracy, interpretability, and computational efficiency.

INTEGRATION

The project implements a functional REST API in Flask that allows you to:

Automatically train or load a CatBoost model when the application initializes.

Retrain the model each time the user uploads a new file.

Expose endpoints that display metrics, feature importance, server status, and predictions.
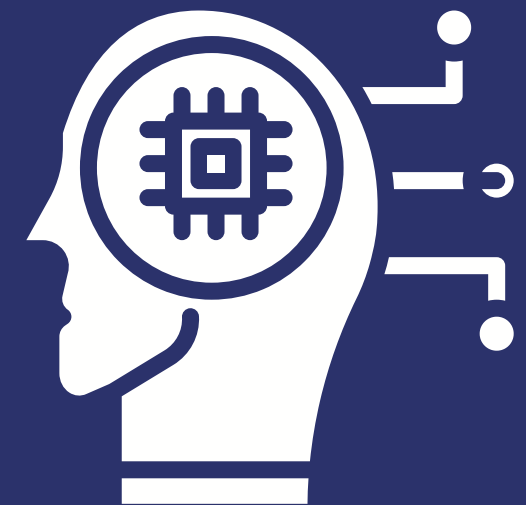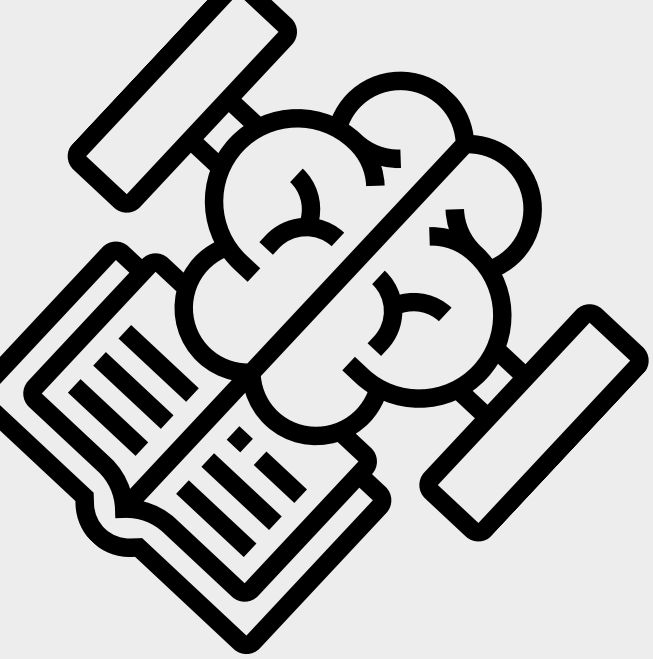
ARCHITECTURE

- Train_model.py → Model training and loading

- Preprocess.py → Consistent data preprocessing

- App.py → API and user interaction logic

- Config.py → Routing and global configuration

PREDICTIONS

- The system provides detailed predictions for each record, including both multiclass (0–3) and binary risk outcomes.
- Each prediction includes:
- ID – original identifier or index
- SII prediction (0–3) – severity level
- Binary risk score (0 = normal, 1 = problematic)
- Normal vs. problematic probability
- Overall confidence level
- This structure enables accurate decision-making, risk categorization, and seamless integration with dashboards or reporting tools.

PRETRAINED MODEL

• Training Data:

CatBoost model trained on a curated dataset of 2,736 validated samples.

• Predictive Performance:

- TAccuracy: 85%
- Recall: 79%
- F1-score: 81%
- AUC: 0.91, indicating strong class discrimination.

• Key Features:

- PCIAT problematic internet use scale
- Daytime sleepiness (SDS)
- Physical activity indices
- BMI
- Heart rate

WE LOOK FORWARD TO YOUR FEEDBACK!