THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Child Mind Institute — Problematic Internet Use

DATS 6501 - Team 4

BHARATH GENJI MOHANA RANGA

PRADEEP SOMASUNDARAM

TABLE OF CONTENTS

INTRODUCTION

The internet has become an essential part of modern life, but excessive usage can lead to **Problematic Internet Use (PIU)**, negatively impacting mental health, productivity, and social interactions.

This study aims to develop a **machine learning model** that predicts an individual's PIU risk based on behavioral, demographic, and survey data.

By leveraging AI, we can enable **early detection and targeted interventions,**

PROBLEM STATEMENT

PIU is characterized by excessive engagement with digital content, leading to:

- Reduced academic or work performance
- Increased stress, anxiety, and social withdrawal
- Difficulty in controlling screen time

Traditional **self-reported assessments** are subjective. Machine learning can provide a **data-driven, objective** approach to PIU detection.

Why does this matter?

- Internet addiction is a growing concern.
- Behavioral factors like **screen time, gaming, and social media usage** are crucial indicators of PIU.
- **Predictive models** can offer real-time risk assessment, aiding in intervention planning.

PROBLEM SCOPE

Objective: Develop a model to classify individuals into PIU risk categories.
Data: Healthy Brain Network

Methodology: Data preprocessing, Exploratory Data Analysis (EDA), Feature engineering, Machine learning modeling & evaluation, Visual insights

Goal: A reliable model that can aid in early detection and prevention strategies.

LITERATURE REVIEW

Gap: Existing research focuses on *post-diagnosis* interventions, mental health correlations, and neurological changes, but lacks tools for *early*, *objective detection* of PIU using accessible physical biomarkers.
Our Address: We bridge this gap by leveraging wearable-derived activity data (sleep, posture, activity bursts) to predict PIU severity (**sii**), enabling proactive, non-stigmatizing interventions before clinical escalation.

RELEVANT RESEARCH

Scarcity of Research on PIU in U.S. Youth
(Problematic Internet Use Among US Youth: A Systematic Review)
Limited empirical studies exist on **PIU among U.S. adolescents and college students**, highlighting the need for more research.
Impact on Mental Health
(Relationship between Internet Addiction and Mental Health in Adolescents)
Excessive internet use is strongly linked to **mental health issues**, raising **public health concerns**.
Association with Psychiatric Disorders
(Problematic Internet Use in Children and Adolescents: Associations with Psychiatric Disorders and Impairment)
PIU is correlated with **various psychiatric disorders**, affecting emotional well-being and daily life.

METHODOLOGY

The project follows a structured pipeline:

Data Collection & Cleaning – Address missing values, remove inconsistencies.

Exploratory Data Analysis (EDA) – Identify trends and correlations.

Feature Engineering – Extract meaningful behavioral indicators.

Model Training & Evaluation – Use machine learning for PIU classification.

Visualization & Insights – Present findings in an interpretable manner.

DATA DESCRIPTION

Physical activity data (wrist-worn accelerometer readings, fitness assessments, and surveys).

Internet usage behavior data (daily screen time, social media engagement, gaming hours).

Target Variable: Severity Impairment Index (sii), derived from the **Parent-Child Internet Addiction Test (PCIAT-PCIAT_Total)** score, categorized as:

- **0:** None : **1:** Mild : **2:** Moderate : **3:** Severe

DATA DESCRIPTION

Dataset Structure:

- **Tabular Data:** Demographics, behavioral information, survey responses (CSV format).
- **Actigraphy Data:** Continuous wrist-worn accelerometer readings (Parquet format).

DATA DESCRIPTION

Key Measurement Instruments:	**Demographics:** Age, sex.
	Internet Use: Hours spent online daily.
	Children's Global Assessment Scale: Measures mental health conditions.
	Physical Measures: BMI, blood pressure, heart rate.
	Fitness Assessments: Treadmill and child-specific fitness scores.
	Sleep Disturbance Scale: Identifies sleep disorders.

DATA DESCRIPTION

Actigraphy Data Fields:

X, Y, Z Accelerations: Measures motion along standard axes.

ENMO: Euclidean Norm Minus One of accelerometer signals.

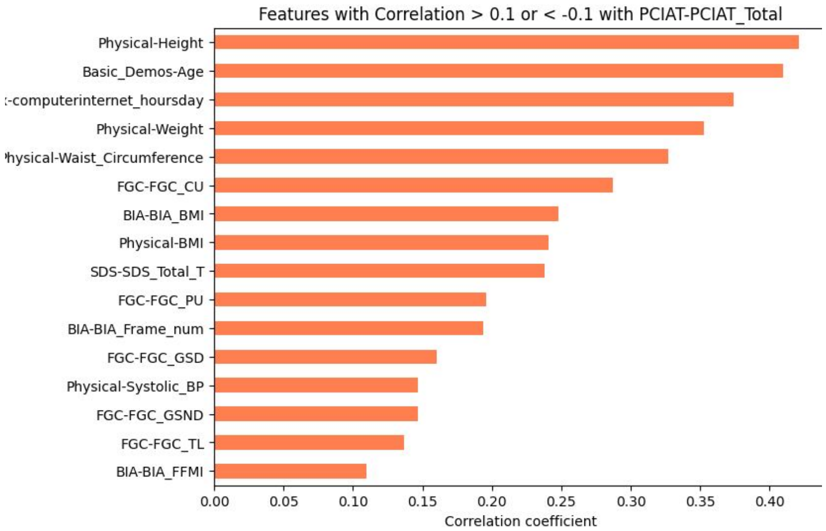
Angle-Z: Measures arm angle relative to horizontal plane.

Non-wear Flag: Detects when the device was removed.

Light Exposure: Captures ambient light levels in lux.

Battery Voltage: Indicates power status of the device.

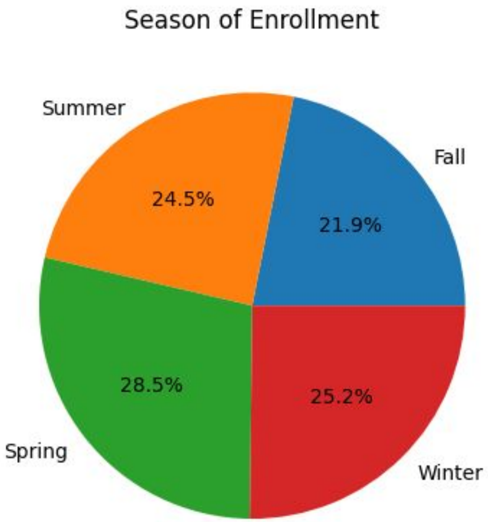
DATA PREPROCESSING AND FEATURE ENGINEERING

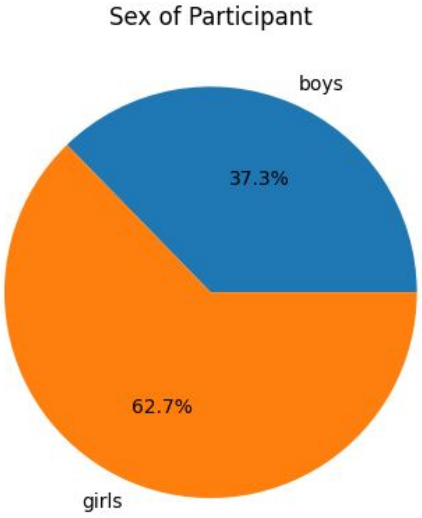


- Before training our model, we ensured data consistency:

- **Handled missing values** with median imputation.
- **Removed highly correlated features** (e.g., social media usage, r = 0.85).
- **Standardized numerical variables** for model compatibility.
- **Encoded categorical variables** like gender and region.

EXPLORATORY DATA ANALYSIS





Enrollment is relatively balanced across all seasons. This trend might indicate seasonal preferences or availability in the program.

A majority of participants are girls (62.7%) compared to boys (37.3%).This could indicate a higher interest or accessibility for girls in the program.

FEATURE ENGINEERING

Feature Selection: The dataset contains features related to physical characteristics (e.g., BMI, Height, Weight), behavioral aspects (e.g., internet usage), and fitness data (e.g., endurance time).

Categorical Feature Encoding: Categorical features are mapped to numerical values using custom mappings for each unique category within the dataset. This ensures compatibility with machine learning algorithms that require numerical input.

Time Series Aggregation: Time series statistics (e.g., mean, standard deviation) from the actigraphy data are computed and merged into the main dataset to create additional features for model training.

DATA MODELING AND VISUALIZATION

OBJECTIVE: PREDICT **PROBLEMATIC INTERNET USE (PIU)** SEVERITY USING WEARABLE-DERIVED ACTIVITY DATA.

EVALUATION METRIC: **ROOT MEAN SQUARED ERROR (RMSE)** AND **QUADRATIC WEIGHTED KAPPA (QWK)** FOR MODEL ACCURACY AND AGREEMENT.

GOAL: TO ASSESS THE PREDICTIVE PERFORMANCE OF VARIOUS MACHINE LEARNING MODELS AND VISUALIZE KEY TRENDS.

MACHINE LEARNING MODELS USED

LightGBM: Efficiently handles large datasets, optimizing for speed and accuracy while reducing overfitting.

XGBoost: Focuses on boosting accuracy through decision trees and regularization to mitigate overfitting.

CatBoost: Handles categorical data efficiently and performs well with minimal tuning.

We chose LightGBM, XGBoost, CatBoost due to their proven ability to handle complex datasets efficiently, boost predictive accuracy, manage overfitting, and work well with structured tabular data, making them ideal for our PIU severity prediction task.

HYPERPARAMETER TUNING

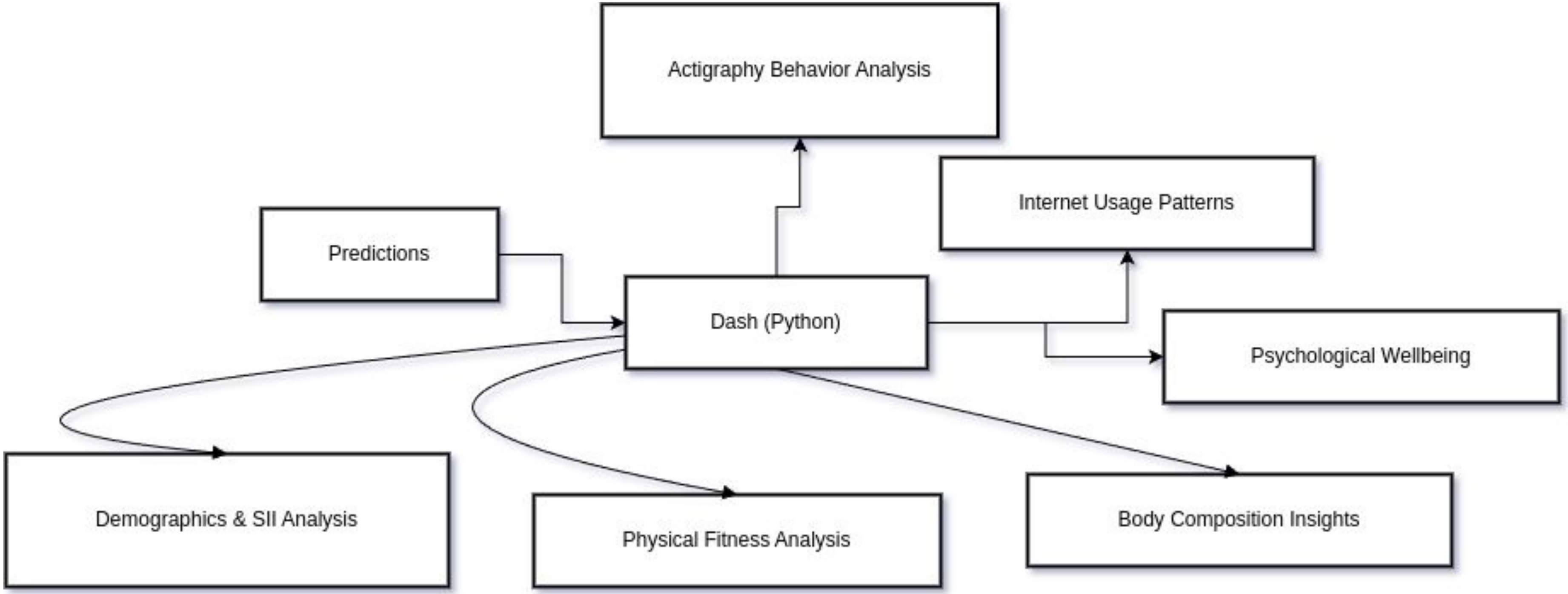
MODELS	Hyperparameter	Tuning Approach
LightGBM	Number of Leaves, Learning Rate, Feature Fraction	Random Search
XGBoost	Learning Rate, Max Depth, Number of Estimators	Grid Search
Catboost	Learning Rate, Depth, L2 Leaf Regularization	Grid Search

RESULTS AND ANALYSIS

MODELS	BEST PREDICTIONS
LightGBM	0.9225
XGBoost	1.2548
Catboost	1.3187

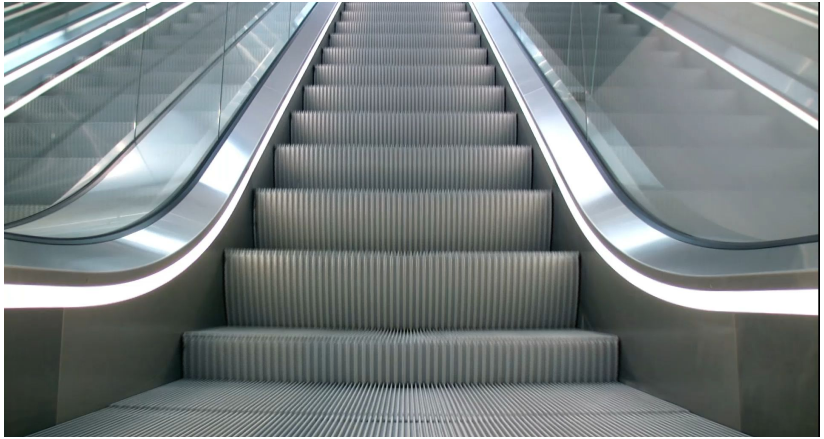
BEST METRICS	VALUES
Mean Squared Error (MSE)	0.3302
R-squared	0.3593
Quadratic Weighted Kappa (QWK)	0.5189

DASHBOARD OVERVIEW



CONCLUSION

We've built strong predictive models using **XGBoost**, **LightGBM**, and **CatBoost** to gauge **Problematic Internet Use (PIU)** severity. After fine-tuning the hyperparameters, **Catboost** really stood out in terms of accuracy. Moving forward, we can make these models even better by adding more features and exploring real-time predictions for proactive interventions.



PROJECT LIMITATION

1. **Data Imbalance**: Some PIU categories had fewer samples, which may have impacted model fairness and generalization.

2. **Feature Constraints**: The absence of behavioral and psychological factors in the dataset may have limited the predictive power of the models.

FUTURE RESEARCH

Incorporating Additional Features: Future models could benefit from including more behavioral and environmental factors, such as **mental health indicators** or **social media usage**, to improve prediction accuracy.

Real-Time Model Deployment: Exploring the potential for **real-time monitoring** and **dynamic prediction models** to track PIU in real-time and offer timely interventions.

REFERENCES

- **Problematic Internet Use Among US Youth: A Systematic Review**
- Source: *PMC*
- Link
- **Relationship between Internet Addiction and Mental Health in Adolescents**
- Source: *PMC*
- Link
- **Problematic Internet Use in Children and Adolescents: Associations with Psychiatric Disorders and Impairment**
- Source: *BMC Psychiatry*
- Link