

## Reporte de la Primera Etapa

# Child Mind Institute — Problematic Internet Use

<https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use/overview>

**Integrantes:** Julian Tabares Restrepo, Daniel Alejandro Garcia Mendez, Willian Alberto Reina Garcia, Andres Velasquez Malvey.

## 1. Análisis del Problema

### 1.1 Identificación Clara del Problema

El uso problemático de Internet (UPI) en jóvenes es un fenómeno creciente que conlleva consecuencias negativas para la salud mental como la aparición o agravamiento de trastornos mentales como la depresión o la ansiedad. Además el UPI, está asociado con alteraciones en los hábitos físicos, el deterioro en las relaciones sociales/familiares y dificultades en la comunicación interpersonal.

Este problema se caracteriza por una dependencia excesiva al uso de internet, la pérdida de control sobre el tiempo de conexión y una disminución en el rendimiento de las actividades diarias como son el estudio, deporte, descanso y las interacciones sociales. En la era digital actual, estos efectos se ven intensificados por el diseño adictivo de las plataformas digitales que utilizan mecanismos de recompensa, notificaciones constantes y algoritmos personalizados al usuario, que fomentan el uso excesivo del internet; como también la misma presión social que impulsa a los jóvenes a permanecer en línea de forma constante.

Frente a esta problemática, se vuelve esencial implementar estrategias de detección temprana del UPI, especialmente en el análisis de registros de actividad física infantil. Esta predicción permitirá intervenir de manera oportuna, reducir los efectos adversos a corto/largo plazo y fomentar hábitos digitales más equilibrados y saludables en los jóvenes.

### 1.2 Objetivos del Proyecto

El objetivo principal de este proyecto es desarrollar un modelo predictivo que pueda identificar el riesgo de uso problemático de Internet (UPI) en niños y adolescentes, utilizando datos de actividad física y otros factores relevantes descritos en el diccionario de datos. Los objetivos específicos incluyen:

**Objetivo General:** Desarrollar un modelo de clasificación con precisión mayor al 70% para la detección temprana del UPI en los jóvenes a registrados, que pueda ser implementado como una herramienta de apoyo en entornos clínicos y educativos.

#### **Objetivos Específicos:**

- Preprocesar y limpiar los datos de actividad física y datos tabulares asociados para garantizar su calidad y formato adecuado para el modelado.
- Realizar un análisis exploratorio de datos (EDA) para identificar patrones, correlaciones y características clave relacionadas con el UPI.
- Evaluar y comparar el rendimiento de diferentes modelos de clasificación (incluyendo Random Forest, K-Means, y modelos de Deep Learning/Bayesianos) para determinar el más adecuado para predecir el riesgo de UPI.
- Optimizar los modelos seleccionados para mejorar la precisión y la capacidad de generalización.

## Reporte de la Primera Etapa

- Generar un informe detallado de las soluciones propuestas, metodologías empleadas y resultados obtenidos.

### 1.3 Árbol de Problemas

Problema Principal: Uso problemático de internet (UPI) en adolescentes, caracterizado por dependencia, pérdida de control y afectación en el desarrollo cognitivo y emocional.

#### Causas (Factores que originan el UPI):

- Causas Individuales:
  - Baja autoestima y ansiedad social.
  - Falta de habilidades de regulación emocional.
  - Predisposición genética a conductas adictivas.
  - Temperamentos afectivos (ciclotímico, ansioso, irritable).
  - Impulsividad.
- Causas Familiares:
  - Supervisión parental insuficiente.
  - Conflictos familiares que llevan a refugiarse en internet.
  - Falta de comunicación.
- Causas Sociales y Tecnológicas:
  - Presión social por permanecer conectado (redes sociales, juegos online).
  - Diseño adictivo de plataformas digitales (recompensas variables, notificaciones constantes).
  - Disponibilidad y accesibilidad del internet.
  - Contenidos específicos (videojuegos, pornografía, redes sociales).

#### Consecuencias (Efectos del UPI):

- Efectos en la salud mental:
  - Depresión, ansiedad, trastornos del sueño.
  - Trastornos de la conducta alimentaria.
  - Aumento del estrés.
- Efectos académicos:
  - Bajo rendimiento escolar.
  - Falta de concentración.
  - Disminución del interés en actividades educativas.
- Efectos sociales:
  - Aislamiento, deterioro de relaciones interpersonales.
  - Reducción de actividades extracurriculares.
  - Conflictos familiares.
- Efectos físicos:
  - Sedentarismo, problemas de visión, hábitos de sueño irregulares.
- Problemas de desarrollo:
  - Afectación en el desarrollo cognitivo y emocional.
  - Dificultades en la regulación emocional y la toma de decisiones.

# Reporte de la Primera Etapa

## 2. Estado del Arte

### 2.1 Revisiones Bibliográficas

A continuación, se presentan cinco revisiones bibliográficas que exploran el uso problemático de Internet, sus factores asociados y el uso de técnicas de Machine Learning para su estudio y prevención:

1. **Jović, J., & Ćorac, A. (2024).** *Using machine learning algorithms and techniques for defining the impact of affective temperament types, content search and activities on the internet on the development of problematic internet use in adolescents' population. Frontiers in Public Health.*

Realizaron un estudio orientado a la identificación de factores psicológicos y conductuales asociados al desarrollo del uso problemático de Internet (UPI) en población adolescente, mediante la implementación de técnicas de aprendizaje automático. En su análisis, se incluyeron variables tales como los tipos de temperamento afectivo, las actividades digitales más frecuentes y los patrones de búsqueda en línea.

Los hallazgos obtenidos permitieron concluir que el temperamento ciclotímico, el tiempo promedio de conexión diaria y el deseo persistente de mantenerse en línea constituyen predictores significativos del UPI, este estudio evidencia que el uso de algoritmos de machine learning permite superar las limitaciones inherentes a los enfoques tradicionales, como los cuestionarios autoreportados, al ofrecer modelos predictivos más precisos que pueden contribuir al diseño de estrategias preventivas tempranas y eficaces.

2. **Karacan, E., et al. (2020).** *Problematic Use of the Internet - Using Machine Learning in a Prevention Programme. ResearchGate.*

Desarrollaron un sistema de clasificación basado en aprendizaje automático con el objetivo de prevenir el UPI, su propuesta consiste en un modelo que categoriza a los usuarios en distintos niveles de riesgo (normal, límite, crítico y severo) utilizando un cuestionario breve que permite recopilar datos de forma no invasiva. El modelo fue entrenado para establecer la relación entre el uso intensivo de Internet y trastornos como la depresión y el trastorno por déficit de atención con hiperactividad (TDAH). Los resultados confirman que, si bien las tecnologías digitales han contribuido a la expansión de comportamientos adictivos, también representan una herramienta valiosa para su identificación temprana, tratamiento y prevención cuando se integran adecuadamente en sistemas de salud y educación. Este estudio destaca el potencial de la inteligencia artificial como aliada en la lucha contra el UPI.

3. **Kuss, D. J., & Griffiths, M. D. (2015).** *Internet addiction and problematic Internet use: A systematic review of clinical research. PMC.*

Analizaron evidencia empírica de carácter clínico relacionada con la adicción a Internet y el UPI, aunque no incorporan directamente modelos computacionales ni técnicas de inteligencia artificial, su trabajo constituye un marco teórico robusto para comprender los determinantes del UPI.

La revisión contempla aspectos sociodemográficos, psicológicos y psiquiátricos, incluyendo factores como la baja autoestima, el aislamiento social, los trastornos del estado de ánimo y la impulsividad. Este corpus teórico proporciona una base fundamental

## Reporte de la Primera Etapa

para la selección de variables relevantes en modelos predictivos y para la interpretación clínica de los resultados obtenidos mediante enfoques automatizados.

4. **Carbonell, X., et al. (2024).** A study of impulsivity as a predictor of problematic internet use in university students with disabilities. *Frontiers in Psychiatry*.

Este estudio investiga la relación entre la impulsividad y el UPI, especialmente en poblaciones vulnerables como estudiantes universitarios con discapacidades, utiliza un modelo predictivo que incluye variables psicológicas y de funcionamiento vital.

Los hallazgos muestran que la impulsividad es un predictor significativo del UPI, junto con rasgos de personalidad y problemas de funcionamiento social. Esto sugiere que las características individuales pueden ser cruciales para los modelos de aprendizaje automático que buscan predecir el riesgo de UPI.

5. **Pontes, H. M., et al. (2023).** Problematic Internet Use among Adults: A Cross-Cultural Study in 15 Countries. *MDPI*.

Este estudio transcultural examina los patrones de uso de actividades en línea, la frecuencia del UPI y los factores de riesgo en adultos de 15 países, aunque se enfoca en adultos, las variables predictivas identificadas (como el networking social problemático y el juego, la falta de perseverancia y la depresión) son relevantes para comprender el UPI en un contexto más amplio y pueden informar el diseño de modelos predictivos en jóvenes, resalta las diferencias en la prevalencia del UPI entre regiones, lo que subraya la necesidad de enfoques culturalmente sensibles en la detección.

### 2.2 Análisis de Soluciones Existentes

En el foro de discusión de la competencia de Kaggle “Child Mind Institute - Problematic Internet Use”, se destacaron dos enfoques metodológicos aplicados al modelado del uso problemático de Internet.

**Solución 1** por modelo Bayesiano jerárquico abordó la predicción de la variable objetivo (PCIAT\_Total, posteriormente transformada a SII) mediante la imputación de valores faltantes utilizando la mediana y la selección de un conjunto reducido de características clave, este enfoque se caracteriza por su capacidad para incorporar conocimientos previos y ofrecer medidas de incertidumbre en las predicciones, siendo especialmente útil en contextos con datos incompletos y estructuras complejas. No obstante, su implementación puede resultar computacionalmente exigente y la interpretación de los resultados más desafiante en comparación con modelos más simples.

**Solución 2** por modelo Deep Autoencoders, redes neuronales profundas entrenadas para aprender representaciones comprimidas de los datos con el fin de optimizar la métrica Quadratic Weighted Kappa (QWK). Este método demostró ser altamente eficaz para el aprendizaje de características en contextos de alta dimensionalidad como los datos biométricos derivados de acelerómetros, permitiendo capturar patrones no lineales y relaciones latentes de forma eficiente. Sin embargo, su principal limitación radica en la necesidad de grandes volúmenes de datos para un entrenamiento adecuado, el alto costo computacional y la escasa interpretabilidad inherente a los modelos de redes neuronales profundas.

# Reporte de la Primera Etapa

## 3. Planeación

### 3.1 Tareas

La planeación propuesta se basa en las tareas clave para el desarrollo del proyecto.

#### Fase 1: Preparación y Análisis de Datos

- Tarea 1.1: Exploración Inicial de Datos: Revisión de la estructura, tipos de variables y calidad general de los datos.
- Tarea 1.2: Análisis Descriptivo: Cálculo de estadísticas básicas y generación de visualizaciones para entender distribución y comportamiento.
- Tarea 1.3: Análisis de Correlación: Identificación de relaciones entre variables independientes y con la variable objetivo (SII).

#### Fase 2: Procesamiento de Datos

- Tarea 2.1: Manejo de Datos Faltantes: Aplicación de métodos de imputación o eliminación según la naturaleza del dato.
- Tarea 2.2: Codificación y Escalamiento: Transformación de variables categóricas y ajuste de escala en variables numéricas.
- Tarea 2.3: Ingeniería de Características (si aplica): Creación de nuevas variables relevantes a partir de las existentes.

#### Fase 3: Selección y Configuración de Modelos

- Tarea 3.1: División de Datos: Separación del dataset en subconjuntos para entrenamiento, validación y prueba.
- Tarea 3.2: Selección y Justificación de Modelos Candidatos: Elección de modelos adecuados justificando su aplicabilidad.
- Tarea 3.3: Configuración Inicial de Modelos y Métricas de Evaluación: Establecimiento de parámetros base y métricas de desempeño.

#### Fase 4: Entrenamiento, Optimización y Evaluación

- Tarea 4.1: Entrenamiento de Modelos: Aplicación de algoritmos seleccionados al conjunto de entrenamiento.
- Tarea 4.2: Optimización de Hiperparámetros: Ajuste fino de parámetros
- Tarea 4.3: Evaluación Final del Modelo: Valoración del rendimiento en el conjunto de prueba mediante métricas y visualizaciones.
- Tarea 4.4: Comparación y Selección del Mejor Modelo: Elección del modelo con mejor balance entre precisión y robustez.

# Reporte de la Primera Etapa

## 3.2 Cronograma

(Tiempo estimado: 11 Semanas):

