KidSec Analytics: reportes del comportamiento digital y riesgo en el dispositivo

Sebastián Correa Fuentes

Carlos Samuel Jiménez Jiménez

Juan Pablo Chabur Rodríguez

Facultad de Ingeniería, Universidad EAN

Proyecto de Grado, Grupo 1

Director(a): Lina María Chacón Rivera

Bogotá 2026

Agradecimiento

Los autores expresan su agradecimiento a todas las personas que hicieron posible el desarrollo de este proyecto de investigación. En especial, se reconoce la colaboración y el aporte académico de Angie Dayanna Acosta Rodríguez, con quien se desarrolló la fase inicial del proceso investigativo durante el curso de seminario de investigación. Su participación fue fundamental en la construcción de las bases conceptuales y metodológicas que permitieron orientar el presente trabajo de grado.

De igual manera, se agradece a los padres de familia, niños y adolescentes que participaron de forma voluntaria en la aplicación de las encuestas y en la recolección de la información, ya que su colaboración permitió obtener los datos necesarios para el análisis y la comprensión de la problemática abordada.

Finalmente, se reconoce al acompañamiento de la docente y asesores académicos que, a través de sus orientaciones y recomendaciones, contribuyeron al fortalecimiento del desarrollo de este proyecto.

Resumen

El acceso temprano de los niños y adolescentes a dispositivos móviles ha incrementado significativamente la exposición a riesgos en entornos digitales, tales como malware, robo de información, ciberacoso, interacciones de alto riesgo en video juegos y redes sociales, así como la exposición a contenidos sensibles y maliciosos. En Colombia esta problemática se ve agravada por la falta de capacidad de los menores para identificar amenazas en línea y por las restricciones de los sistemas actuales de control parental, los cuales suelen ser reactivos, invasivos y poco eficaces al no ofrecer un monitoreo en tiempo real de comportamientos anómalos dentro de las aplicaciones.

Ante este escenario, el proyecto propone el análisis y diseño de un aplicativo de seguridad basada en Machine Learning orientada a la identificación automática de patrones de comportamiento digital sospechoso. La metodología implementada corresponde a un enfoque cuantitativo de tipo descriptivo – correlacional apoyado en una revisión bajo la metodología PRISMA y en el desarrollo de un prototipo funcional. Dicho prototipo integra un modelo de Machine Learning entrenado con datos simulados de comportamiento digital, permitiendo clasificar patrones de uso como seguros o riesgosos y generar alertas preventivas.

El sistema se implementa bajo una arquitectura cliente – servidor, utilizando backend en Python para el procesamiento de datos y la inferencia del modelo, operando en segundo plano de forma continua sin interferir directamente en la experiencia del usuario, se espera que el enfoque propuesto permita una detección proactiva de comportamientos digitales de riesgo, fortaleciendo la supervisión adulta y contribuyendo a la creación de entornos digitales más seguros para los menores, bajo principios de ética, privacidad y acompañamiento humano.

Palabras clave: Machine Learning, seguridad digital infantil, detección de comportamientos maliciosos, dispositivos móviles, Analísis de riesgos, supervisión parental.

Abstract

Early access to mobile devices for children and adolescents has significantly increased their exposure to risks in digital environments, such as malware, data theft, cyberbullying, high-risk interactions in video games and social media, as well as exposure to sensitive and malicious content. In Colombia, this problem is exacerbated by children's lack of ability to identify online threats and by the limitations of current parental control systems, which tend to be reactive, invasive, and ineffective because they do not offer real-time monitoring of anomalous behavior within applications.

Given this scenario, the project proposes the analysis and design of a machine learning-based security application aimed at automatically identifying patterns of suspicious digital behavior. The implemented methodology corresponds to a quantitative, descriptive-correlational approach supported by a review using the PRISMA methodology and the development of a functional prototype. This prototype integrates a machine learning model trained with simulated digital behavior data, allowing the classification of usage patterns as safe or risky and the generation of preventive alerts.

The system will be implemented using a client-server architecture, with a Python backend for data processing and model inference. Operating continuously in the background without directly interfering with the user experience, the proposed approach is expected to enable proactive detection of risky digital behaviors, strengthening adult supervision and contributing to the creation of safer digital environments for minors, based on principles of ethics, privacy, and human guidance.

Keywords: Machine Learning, child digital safety, detection of malicious behavior, mobile devices, risk analysis, parental supervision

Tabla de Contenido

Tablas

Figuras

Introducción

En Colombia, en numerosas ocasiones utilizar dispositivos móviles inteligentes ha sobrepasado el progreso de soluciones apropiadas en la seguridad digital. Según una investigación de la Universidad Nacional Abierta y a Distancia (UNAD, 2021), hubo diversas tácticas de robo entre 2018 y 2020 que afectaron a individuos de diferentes edades, incluyendo a los menores, lo cual causo perdidas de información, daños económicos y efectos colaterales. Esta investigación revelo que una porción significativa de la población no tiene conocimiento sobre medidas necesarias de protección, como el uso de contraseñas seguras y/o robustas, así como validar enlaces sospechosos a los que ingresan o la implementación de protocolos de seguridad. De esta manera aumenta la exposición a peligros en línea.

Esta situación resulta especialmente relevante cuando se analiza el uso temprano de dispositivos móviles por parte de los menores en Colombia, dicha situación se vuelve sumamente importante. Según la Comisión de Regulación de Comunicaciones, citada en Caracol Radio (2024), el 35% de los niños entre 6 y 9 años utilizan un teléfono inteligente de propiedad de sus padres, asimismo, el 61% posee un dispositivo propio, cantidad que crece hasta el 81% en los adolescentes de entre 14 y 17 años. También, el 40% de los niños tienen perfiles activos en redes sociales y este dato va en aumento al 77% entre los jóvenes más grandes. Estos datos muestran una participación constante en plataformas digitales abiertas, donde menores interactúan con otros usuarios, algoritmos y contenidos delicados, lo que aumenta el área de exposición a las amenazas digitales. Además, el 14% de los padres solamente dicen conocer herramientas para la alfabetización digital o de control parental, lo cual restringe la supervisión efectiva del uso de la tecnología (Gómez, 2025).

Investigaciones en la materia han logrado constatar que los riesgos se evidencian con el contenido, el contacto y el comportamiento dentro de plataformas en línea. Este hecho incluye el acceso a contenido inapropiado, el ciberacoso, la interacción con personas desconocidas y

la propagación de software infectado (Livingstone, 2013). Gath y Swit (2024) observaron en este sentido que a medida que los niños tienen más dispositivos personales, la probabilidad de sufrir daños en línea puede incrementar, indicando que el riesgo aumenta a un 21% por cada dispositivo adicional que tienen. Esto indica que tener acceso a dispositivos propios y libertad digital temprana aumenta la vulnerabilidad.

Desde este punto, Olguin y Arana-Llanes (2024) señalan desde la perspectiva técnica, que las aplicaciones móviles se pueden volver blancos de ataque si se utilizan permisos inseguros, si se realizan descargas a través de fuentes no confiables u oficiales, o si se aprovechan las debilidades del sistema operativo. Estas condiciones hacen más fácil la instalación de malware y la captura de información financiera o personal, afectando en mayor grado a los usuarios con menor criterio digital y capacidad intelectual reducida.

A pesar de que en Colombia y otros países se han motivado acciones legales para la protección digital de los niños, como lo son sistemas para verificar la edad y restricciones para poder acceder a ciertos contenidos, estas acciones no siempre ayudan a identificar conductas delictivas en tiempo real dentro de los dispositivos de los menores. Esto pone en evidencia la necesidad de acciones más eficientes que puedan detectar patrones anómalos de uso directamente en el entorno digital.

Dentro de este contexto, existen varios estudios acerca del uso de técnicas de inteligencia artificial como método de mitigación. Zaccagnino et al., (2021) desarrollaron un modelo de Machine Learning que reconoce adultos y menores mediante el análisis de patrones táctiles en dispositivos móviles. El estudio, que examino más de 9000 gestos táctiles de 147 participantes, tuvo una exactitud cercana al 88% y un AUC de 0.92. Esto demuestra que los patrones de uso son capaces de ser empleados como indicadores de comportamiento para calcular la edad del usuario y activar mecanismos automáticos de protección. De forma complementaria Mirabet (2024) diseño un sistema para detectar malware en aplicaciones de Android usando modelos supervisados, donde demuestra que el aprendizaje automático facilita

la automatización para la detección de peligros con gran efectividad en el sistema operativo más usado en el mundo.

A pesar de lo que se ha avanzado, todavía falta camino por recorrer en la protección de menores que viven constantemente usando el dispositivo móvil en Colombia y toda la responsabilidad no puede caer en el adulto. Para ello es necesario incorporar sistemas inteligentes que estén orientados únicamente a la protección de menores que usan todos los días un dispositivo móvil sin saber a qué riesgos estén expuestos. Estas herramientas deben ser capaces de identificar los riesgos digitales por sí mismos y no depender solamente de la supervisión del adulto.

En este contexto, la finalidad y objetivo principal de esta investigación es realizar un aplicativo basado en Machine Learning que ayude a la identificación y prevención automática de riesgos digitales en los dispositivos móviles que son usados por los menores en Colombia. Para lograrlo se van a estudiar las amenazas que más afectan a los menores hoy en día y cómo funcionan dichas vulnerabilidades en apps cotidianas, de igual manera se revisaran que modelos de aprendizaje autónomo sirven mejor para prevenir y frenar los riesgos, y al final, desarrollar un aplicativo funcional que aplique los mecanismos para mantener las medidas de seguridad y protección al día.

Así, la pregunta de investigación se define como: ¿Hasta qué punto un modelo de Machine Learning autónomo, implementado en dispositivos móviles, hace posible la identificación de forma temprana riesgos digitales en menores en Colombia y disminuir su exposición frente a la supervisión parental tradicional?

Marco Teórico

El análisis del desarrollo infantil en entornos digitales requiere una aproximación multidimensional que integre perspectivas cognitivas, socioemocionales y ecológicas. Estas características evolutivas los hacen especialmente vulnerables ante contenidos digitales engañosos o manipuladores, ya que carecen de la capacidad para evaluar críticamente la intencionalidad de los mensajes recibidos.

Las teorías de los efectos mediáticos complementan este análisis al explicar cómo los entornos digitales influyen en el aprendizaje, la percepción y la construcción de significados en los niños. La exposición prolongada a contenidos móviles no solo afecta el tiempo de pantalla, sino que moldea percepciones, actitudes y comportamientos de los menores en línea (Peñuela, 2023). Esta influencia se potencia por el diseño persuasivo de muchas aplicaciones infantiles, que utilizan mecanismos de recompensa intermitente y gamificación para prolongar el tiempo de uso, generando patrones de dependencia tecnológica.

La exposición a contenidos inapropiados constituye una de las principales preocupaciones en el ámbito de la seguridad digital infantil. Livingstone et al., (2019) identifican que los menores son particularmente vulnerables debido a su dificultad para comprender las implicaciones sobre privacidad, la falta de control en el manejo de datos personales y la exposición a algoritmos que no priorizan su bienestar. Los contenidos problemáticos incluyen material violento, sexual, discriminatorio o que promueve conductas de riesgo como trastornos alimentarios o autolesiones.

La vulneración de la privacidad infantil mediante la recolección indebida de datos personales representa una categoría de riesgo en crecimiento. Su et al., (2023) evidencian que el 4.47% de las aplicaciones clasificadas como "Family apps" solicitan permisos de ubicación, a pesar de que las políticas de Google Play prohíben explícitamente la recolección de datos de

ubicación para menores. Esta discrepancia entre normativa y práctica expone a los niños a riesgos de seguimiento, perfilamiento y potencial explotación comercial de sus datos.

Lopes et al., (2023) documentan una brecha significativa entre lo que las políticas de protección de datos exigen y lo que muchas aplicaciones infantiles realmente implementan, generando riesgos de daño físico, psicológico o de privacidad para los niños. Ali et al. (2020) complementan este análisis demostrando que incluso las soluciones de control parental presentan vulnerabilidades significativas en materia de seguridad y privacidad, lo que paradójicamente puede incrementar la exposición de los menores a riesgos digitales.

En Colombia, estudios de la UNAD (2021) documentan que entre 2018 y 2020 se presentaron diversas técnicas de ataque que afectaron a personas de todas las edades, incluyendo menores, con consecuencias que van desde la pérdida de datos sensibles hasta daños económicos y emocionales. La investigación revela que gran parte de la población desconoce medidas básicas de seguridad como el uso de contraseñas robustas, la verificación de enlaces o el empleo de protocolos de protección, lo que incrementa significativamente la exposición al riesgo.

Una de las líneas más consolidadas de aplicación de machine learning a la seguridad infantil es la detección de malware y amenazas ocultas en entornos Android. Mirabet (2024) desarrolló una herramienta para detección de malware en aplicaciones Android mediante técnicas de machine learning, fundamentada en el reconocimiento de que el sistema operativo Android representa el 85% del mercado de dispositivos móviles y es el más atacado por ciberdelincuentes. Una línea innovadora de investigación se centra en la identificación automática de menores mediante el análisis de sus patrones de interacción con dispositivos móviles. Zaccagnino et al., (2021) desarrollaron un enfoque de tecno-regulación basado en machine learning para distinguir entre menores y adultos mediante el análisis de gestos táctiles, con el objetivo de activar salvaguardas automáticas para la protección infantil en línea.

El diseño experimental aplico distintos métodos, con aproximaciones single-view y multi-view donde se empleó una muestra de 147 participantes y más de 9,000 gestos táctiles, esto permitió alcanzar un AUC (área bajo la curva ROC) de 0.92 y una precisión del 88%. El estudio afirma que es posible identificar la edad del usuario mediante gestos táctiles y poder aplicar medidas automáticas de protección, donde se puede reducir significativamente la necesidad de una supervisión constante por parte de los tutores principales.

Estudios complementarios de Nguyen et al., (2019), Tolosana et al., (2022) y Mane et al., (2023) han explorado ciertas técnicas de detección que se basa en biometría conductual, interacciones táctiles y reconocimiento de patrones de uso, donde aportaron herramientas para adaptar la experiencia digital a la edad del usuario y necesidades específicas del infante. Estos sistemas permiten el inicio contextual de medidas de protección, ajustando dinámicamente el nivel de restricción según el perfil del usuario que se detecta.

Al desarrollar e implementar herramientas para cuidar a los menores, deben basarse en principios éticos y morales, sin vulnerar la privacidad del infante 24/7, así logrando encontrar un equilibrio entre protección y privacidad en ámbitos de seguridad digital, así mismo, lograr otorgar su propio espacio a medida que estos crecen.

El tratamiento de datos personales está regulado tanto en el ámbito internacional como nacional. En Colombia, la base de todo es la Ley 1581 de 2012 donde se establecen las normas generales de protección de datos personales de los menores, actualmente han implementado nuevas reglas que se enfocan especialmente en cuidar de su información cuando se conectan a una red.

El concepto de safety by design, articulado por Faraz et al., (2022), afirman que las aplicaciones digitales infantiles deben ser seguras desde antes de lanzarla al mercado, en lugar de añadirlas posteriormente como parches correctivos cuando la aplicación entro en fallas. En la práctica, esto significa revisar desde el día uno como la aplicación afecta la privacidad, así pidiendo solo los datos estrictamente necesarios para el correcto funcionamiento, no

guardando datos de alta sensibilidad y explicado detalladamente cómo funcionan los algoritmos que toman decisiones por ellos.

Urbano (2024) afirma que desarrollar un marco de estrategias éticas basadas en inteligencia artificial puede mitigar la violencia digital contra los menores, enfatizando la necesidad de balancear la protección con respecto a derechos fundamentales. El autor identifica un choque en querer que los menores estén seguros y respetar cosas como lo es su privacidad, así mismo, opina que los menores deben aprender a defenderse y manejarse solos internet sin depender de las ayudas externas. Aportando un equilibrio entre protegerlos y dejarlos crecer.

Entre los principios éticos claves que se identifican son:

- Que, los papas como los menores sepan que están siendo vigilados y para qué.
- Proporcionalidad, limitando la vigilancia a lo mínimo necesario para garantizar protección y vigilar más allá vulnerando la privacidad de los menores
- Consentimiento informado, donde se explique al menor lo que se va a realizar en un idioma que pueda percibir fácilmente según su edad, para que logre estar de acuerdo
- Que, si un algoritmo o una Inteligencia Artificial toma una decisión, siempre haya un padre que pueda revisar que todo lo que está pasando detrás este correctamente
- Si el sistema genera una alerta, el menor está en su derecho de saber que paso y porque paso.

Organismos internacionales como UNICEF (2020) han gestionado normativas para la protección de derechos digitales de menores. Estos marcos reconocen que los niños están expuestos no solo riesgos a individuales, sino también amenazas estructurales relacionadas con desigualdades socioeconómicas que limitan la correcta intervención adulta y el acceso a herramientas de protección.

La Convención sobre los Derechos del Niño, en su Observación No. 25 sobre derechos de los niños con relación a entornos digitales, establece que los Estados deben garantizar la protección de los niños en cuanto a la violencia, la explotación y el abuso en entornos digitales, así mismo, que se respete su derecho a la participación, la privacidad y el acceso a la información. Estas normativas internacionales establecen la necesidad de soluciones que integren protección con empoderamiento digital.

En la revisión teórica se evidencia la convergencia de múltiples disciplinas para la comprensión y abordaje de la seguridad digital infantil. Desde la psicología del desarrollo se entiende la vulnerabilidad cognitiva y socioemocional en los menores; desde las ciencias de la computación se desarrollan herramientas de detección y protección; desde las ciencias sociales se estudian los contextos familiares y educativos que condicionan el comportamiento en entornos digitales; y desde el derecho y la ética se establecen normativas que equilibran protección con derechos fundamentales.

La efectividad de sistemas de protección basados en IA depende críticamente de su capacidad para operar en el equilibrio entre seguridad y privacidad, entre protección y autonomía, entre vigilancia y confianza. Los hallazgos revisados sugieren que este equilibrio es alcanzable mediante diseño cuidadoso, transparencia en operación, y mantenimiento de la supervisión humana en decisiones críticas.

Estado del Arte

Diseño metodológico

Con el objetivo de que el estudio del estado del arte sea sencillo y de fácil replicación, se organizó la búsqueda de información aplicando el método PRISMA. Ya que este facilita el estudio del uso del Aprendizaje Automático (Machine Learning) para proteger a los niños cuando hacen uso de un dispositivo móvil y así construir un estado del arte que sea firme y claro.

Estrategia de Búsqueda

La investigación se llevó a cabo usando las herramientas proporcionadas por la Biblioteca de la Universidad EAN como lo son las bases de datos que se indagaron.

- Google Scholar
- Scopus

Donde se logró conseguir artículos, revistas, investigación en donde se trabajó con Machine Learning, que se relacionan con el tema de investigación.

De igual manera, se aplicó una ecuación de búsqueda para identificar y poder mejorar las instrucciones para encontrar lo más relevante para la investigación. La ecuación de búsqueda para scopus, **TITLE-ABS-KEY (("machine learning" OR "artificial intelligence") AND (child* OR adolescen*) AND (cyberbullying OR grooming OR "problematic smartphone use" OR "internet addiction") AND (prediction OR detection OR classification)) AND (LIMIT-TO (EXACTKEYWORD , "Machine-learning") OR LIMIT-TO (EXACTKEYWORD , "Cyberbullying") OR LIMIT-TO (EXACTKEYWORD , "Cyberbullying Detection") OR LIMIT-TO (EXACTKEYWORD , "Grooming Detection")) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (LANGUAGE , "English") OR LIMIT-TO (LANGUAGE , "Spanish"))**. Su implementación tuvo filtros en artículos de revistas indexadas, publicaciones revisadas por pares y la posibilidad de acceder al texto completo, facilitando

realizar con éxito la investigación. Además, se aplicó otra ecuación de búsqueda para Google Scholar, **(“artificial intelligence”) AND menores AND (“riesgos digitales”) AND ciberseguridad.**

Criterios de Inclusión y Exclusión

Criterios de inclusión

- Investigaciones publicadas en artículos.
- Publicaciones realizadas entre los años 2020 y 2025
- Estudios empíricos o aplicados.
- Investigaciones donde integraron Machine Learning o IA.
- Relación directa con la seguridad de los niños y la identificación de amenazas en dispositivos móviles inteligentes.

Criterios de exclusión

- Investigaciones que se enfocan únicamente en el adulto.
- Investigaciones sin validación académica.
- Investigaciones sin información numérica que permitan entender el daño que puede causar una actividad sospechosa en el dispositivo de los menores.

Filtrado y Selección de Estudios

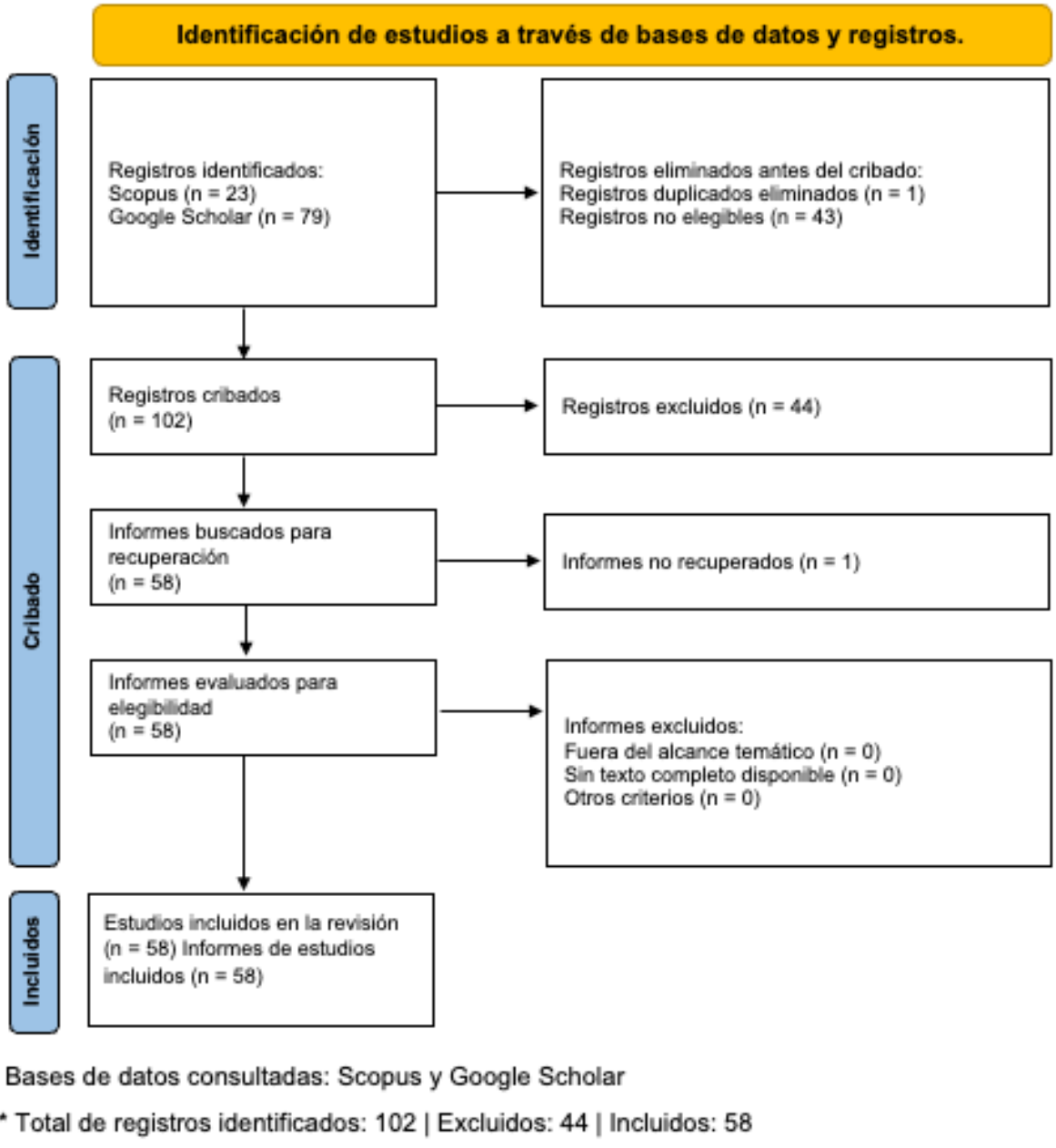
La búsqueda en un principio detecto un conjunto amplio de referencias (102). Tras evaluar por título y resumen, se eliminaron (44) documentos por no cumplir con los criterios de inclusión. Los 52 trabajos restantes fueron examinados. De ellos, se seleccionaron los 58 que cumplían con todos los criterios metodológicos, habían publicados en revistas o conferencias indexadas y estaban directamente relacionados con el tema.

El diagrama de selección se organizó en un diagrama PRISMA, que mide cada fase del cribado. Este diagrama se presenta como evidencia visual de la rigurosidad del proceso y facilita que otros investigadores la reproduzcan.

Diagrama PRISMA

Figura 1

Diagrama de flujo de proceso de identificación, cribado y selección de estudios.



Nota. El proceso sistemático de identificación de estudios mediante bases de datos y registros es representado en el diagrama, seguido por las fases de cribado, evaluación de elegibilidad e

inclusión final. Se presentan las cifras de los registros identificados (n = 102), los informes se analizaron y se tomaron como parte del análisis comparativo final (n = 58), excluyendo (n = 44) informes. Elaboración propia.

Tendencias Metodológicas y resultados recurrentes

El examen de las investigaciones elegidas muestra varias tendencias metodológicas que son claras para el proyecto.

- **Uso de Machine Learning y aprendizaje automático:** Todos los trabajos escogidos usan técnicas de aprendizaje automático con distintos objetivos: detección de amenazas en sistemas digitales para niños, análisis de interacción móvil para identificar usuarios infantiles y creación automatizada de herramientas de protección. Por ejemplo, Faraz et al., (2022) y Faraz et al., (2024) utilizan la arquitectura de inteligencia artificial para desarrollar modelos de protección infantil en ambientes de videojuegos. Tolosana et al., (2022) plantea un modelo más grande para identificar la edad del usuario cuando se usa un dispositivo móvil a través de la biometría conductual.
- **Enfoque en privacidad y seguridad de aplicaciones móviles:** El estudio de Sun et al., (2023) resalta practicas relacionadas con la privacidad de las aplicaciones que son empleadas por menores, lo que revela una inclinación que asciende no únicamente a examinar las amenazas a las que están expuestos, sino que también como las aplicaciones mismas pueden vulnerar la información personal de los menores.
- **Integración de IA en herramientas practicas:** los trabajos que fueron revisados no se limitan únicamente a las descripciones teóricas, sino que también muchas de ellas presentan propuestas como lo son, (chatbots con Inteligencia Artificial enfocados en la seguridad para video juegos, sistemas de detección para verificar si un adulto o un menor usa el dispositivo móvil a través de gestos táctiles) esto representa grande

avances que van orientados en la aplicación práctica de dispositivos y plataformas reales.

Estas tendencias indican que la literatura actual cambia enfoques tecnológicos (modelos ML, algoritmos de clasificación) por aplicaciones dirigidas hacia contextos cotidianos para niños, que van desde los videojuegos hasta las aplicaciones móviles.

Vacíos de Conocimiento Identificados

A pesar de los aportes de los estudios seleccionados, también existen vacíos claros en la literatura.

- **Limitada integración de múltiples variables:** Aunque hay investigaciones sobre como identificar la detección de edad y otros sobre seguridad de aplicaciones, pocas investigaciones combinan la detección de usuario infantil, análisis de los comportamientos sospechosos y poder prevenir las amenazas digitales en un solo modelo. Esto limita la capacidad para encontrar soluciones que traen estos aspectos al mismo tiempo.
- **Contextos regionales poco explorados:** la mayoría de los estudios se centran en contextos tecnológicos desarrollados en (EE. UU, Europa), donde hay un gran vacío en términos de aplicaciones y amenazas en regiones como lo es Latinoamérica.
- **Escasez de validaciones extensivas:** varios trabajos presentan propuestas, pero pocas cuentas con evaluaciones a gran escala.

Estos vacíos señalan oportunidades de investigación donde aportes novedosos pueden tener impacto significativo.

Relación del Estado del Arte con el Proyecto Actual

La revisión sistemática ha permitido conocer que, si bien se han logrado progresos en aplicar la IA para proteger a los niños en los entornos digitales, pero todavía no hay un modelo

integrado que combine detección de edad, análisis de amenazas y mitigación en un solo sistema que se pueda aplicar específicamente a dispositivos móviles fuera del entorno de los juegos y en aplicaciones fragmentadas. Para ello, el enfoque del proyecto es poder desarrollar un sistema de protección proactivo y contextualizado que aplica técnicas de Machine Learning en los dispositivos móviles usados por lo niños, representa una innovación significativa en comparación con las investigaciones revisadas. La propuesta tiene como objetivo por cerrar la brecha detectada a través de un enfoque más integral y contextualizado, incorporando factores técnicos, conductuales y ambientales que aún no han funcionado de manera exhaustiva en trabajos anteriores

Tabla 1

Estado del arte sobre protección infantil digital e inteligencia artificial.

Tema y subtema	Teoría / Modelo / Concepto	Descripción o idea central	Autor y año	Fuente APA
Ciberbullying / Detección de intención	NLP + FastText con análisis de intención y puntaje bully-victim	Detecta cyber harassment e intención del texto en redes sociales usando FastText y análisis de orden de palabras, reportando mayor exactitud que métodos previos.	Abarna et al. (2022)	Abarna, S., Sheeba, J. I., Jayasrilakshmi, S., & Devaneyan, S. P. (2022). Identification of cyber harassment and intention of target users on social media platforms. *Engineering Applications of Artificial Intelligence*. https://doi.org/10.1016/j.engappai.2022.105283
Ciberbullying / Detección con ML y DL	Modelo híbrido BERT + rasgos léxico-lingüísticos	La combinación de transformers con características lingüísticas manuales alcanza 98.6% de accuracy en detección de cyberbullying, superando modelos individuales.	Abdullah et al. (2025)	Abdullah, Latif, I., Hafeez, N., Ullah, F., Sidorov, G., Felipe-Riverón, E., & Gelbukh, A. (2025). Cyberbullying detection on social media using machine learning techniques. *Computación y Sistemas*. https://doi.org/10.13053/cys-29-3-5481
Herramientas de control parental	Riesgos de control	Algunas soluciones de control parental generan nuevas vulnerabilidades.	Ali et al. (2020)	Ali, S., Elgharabawy, M., Duchaussoy, Q., Mannan, M., & Youssef, A. (2020, Diciembre 7). Betrayed by the Guardian: Security and Privacy Risks of Parental Control Solutions. In Annual Computer Security Applications
Ciberbullying / Detección en redes sociales	MLP con selección de características (BOW, TF-IDF, n-grams)	Modelo de deep learning aplicado a tuits en turco con 93.2% de accuracy, demostrando la utilidad de la selección de características en contextos lingüísticos específicos.	Aliyeva y Yağanoğlu (2025)	Aliyeva, Ç. O., & Yağanoğlu, M. (2025). Deep learning approach to detect cyberbullying on Twitter. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-19869-3

Ciberbullying / Detección multilingüe	Transformers híbridos con feature fusion y ensemble voting	Fine-tuning de CAMeLBERT+AraGPT2 y AraBERT+XLM-R con fusión de características supera modelos individuales y enfoques tradicionales en árabe.	Alsuwaylimi y Alenezi (2025)	Alsuwaylimi, A. A., & Alenezi, Z. S. (2025). Leveraging transformers for detection of Arabic cyberbullying on social media: Hybrid Arabic transformers. *Computers, Materials & Continua*. https://doi.org/10.32604/cmc.2025.061674
Sistemas IDS y ML	Algoritmos de aprendizaje automático	Evaluación de algoritmos ML aplicados a detección de intrusos.	Álvarez, Cardenoso (s. f.)	Alvarez, D. G, Cardenoso, D. V. (s. f.). Sistemas de detección de intrusos basados en técnicas de Machine Learning. [Trabajo de fin de grado, Universidad vallolid].
Legislación digital infantil	Responsabilidad compartida	Se planea enfoque legal donde fabricantes y proveedores también asumen responsabilidad.	Argelich Comelles (2025)	Argelich Comelles, C. (2025, mayo 29). La proyectada Ley Orgánica para la protección de las personas menores de edad en los entornos digitales: del control parental defectivo a las obligaciones a fabricantes [Artículo de investigación]. SSRN.
Ciberbullying / Detección multimodal	CNN para detección de armas en imágenes vinculadas a cyberbullying	Framework visual que amplía la noción de cyberbullying más allá del texto, detectando amenazas con armas en imágenes con menores con 86% de accuracy.	Barbosa-Santillán et al. (2025)	Barbosa-Santillán, L. I., Guzman-Velazquez, B. P., Orozco-Aguilera, M. T., & Flores-Pulido, L. (2025). Unraveling cyberbullying dynamics: A computational framework empowered by artificial intelligence. *Information*. https://doi.org/10.3390/info16020080
Ciberbullying / Desbalance de clases en modelos	PLMs con cost-sensitive learning y focal loss	Aborda el desbalance de clases en datasets de cyberbullying usando aumento de datos y aprendizaje sensible al costo, logrando recalls entre 78% y 96%.	Benaissa et al. (2025)	Benaissa, A. R., Harbaoui, A., & Ben Ghezala, H. H. (2025). Class imbalance-sensitive approach based on pretrained language models for the detection of cyberbullying in English and Arabic datasets. *Behaviour & Information Technology*. https://doi.org/10.1080/0144929X.2024.2313142

Victimización digital en adolescentes / Predicción de riesgo	Random Forest con oversampling y regresión logística multivariable	Modelo predictivo de victimización por cyberbullying en países en desarrollo, identificando 12 predictores significativos con 82% de accuracy y AUROC de 0.83.	Braga y Tyrrell (2025)	Braga, P. R. V., & Tyrrell, K. R. (2025). Predicting cyberbullying victimisation in emerging markets and developing countries using the Global School-Based Health Survey. *Machine Learning with Applications*. https://doi.org/10.1016/j.mlwa.2025.100646
Estrategias de IA en psicología infantil	Exposición temprana a la tecnología	Mas de un tercio de los niños en Colombia entre los 6 y 9 años ya usan un teléfono móvil, aumentando su exposición a riesgos digitales desde edades tempranas.	Caracol Radio (2025)	Caracol Radio. (2025, mayo 23). El 35% de niños entre 6 y 9 años en Colombia ya manejan un celular: Estudio CRC. Caracol Radio.
Riesgos en infancia temprana	Aplicaciones de IA en entornos educativos y salud mental	Se plantea el uso de herramientas basadas en inteligencia artificial para fortalecer procesos de acompañamiento psicológico infantil en contextos educativos.	Castro Chaparro & Benítez Vizcaíno, (2025)	Castro Chaparro, I. C., & Benítez Vizcaíno, P. I. (2025, mayo 28). Estrategias con IA para la Psicología Infantil en Cergenial. Universidad de Santander.
Ciberbullying / Intervención en tiempo real	Prototipo de detección con Perspective API y umbrales configurables	Herramienta de chat "Serenity" que alerta a emisor y receptor ante toxicidad, con umbral ajustable por el usuario, demostrando viabilidad práctica sin comprometer privacidad.	Chan et al. (2025)	Chan, J., Kishore, S., & Yang, X. (2025). A real-time cyberbully checker. *Software Impacts*. https://doi.org/10.1016/j.simpa.2024.100710
Ciberbullying / Detección con LLMs	Evaluación comparativa de	Comparación de 20 LLMs y 24 modelos ML/NLP para detectar cyberbullying; Claude 3.0 y	Cirillo et al. (2025)	Cirillo, S., Desiato, D., Polese, G., Solimando, G., Sugumaran, V., & Sundaramurthy, S. (2025). Exploring the ability of emerging large language models to detect cyberbullying in social posts through prompt-based

	LLMs con prompt-based classification	Mistral destacaron en desempeño y calidad de explicaciones según expertos.		classification approaches. *Information Processing & Management*. https://doi.org/10.1016/j.ipm.2024.104043
Ciberbullying / Generalización cross-domain	BERT con evaluación cross-domain en Facebook en turco	BERT logró macro-F1 de 0.928 y demostró robustez en distintos dominios, incluyendo perfilamiento de usuarios agresivos.	Coban, Ozel e Inan (2023)	Coban, O., Ozel, S. A., & Inan, A. (2023). Detection and cross-domain evaluation of cyberbullying in Facebook activity contents for Turkish. *ACM Transactions on Asian and Low-Resource Language Information Processing*. https://doi.org/10.1145/3580393
Ciberbullying / Optimización de modelos ensemble	GloVe + ELSTM-AB con optimización TSGSO	Modelo ensemble que combina embeddings GloVe con LSTM y AdaBoost, optimizado mediante Glowworm Swarm para mejorar detección en Twitter.	Daniel et al. (2023)	Daniel, R., Murthy, T. S., Kumari, C. D. V. P., Lydia, E. L., Ishak, M. K., et al. (2023). Ensemble learning with tournament selected glowworm swarm optimization algorithm for cyberbullying detection on social media. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3326948
Ciberbullying / Construcción de datasets	Dataset multidimensional con agresión, repetición, peerness e intención de daño	Propone un dataset más representativo del fenómeno al incorporar cuatro dimensiones clave ausentes en datasets tradicionales, con línea base en modelos ML.	Ejaz, Razi y Choudhury (2024)	Ejaz, N., Razi, F., & Choudhury, S. (2024). Towards comprehensive cyberbullying detection: A dataset incorporating aggressive texts, repetition, peerness, and intent to harm. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2023.108123
Protección digital con Inteligencia Artificial	Safety by desing	Integración de salvaguardas desde el diseño de plataformas digitales.	Faraz et al. (2022)	Faraz, A., Mounsef, J., Raza, A., & Willis, S. (2022). Child safety and protection in the online gaming ecosystem [Artículo]. IEEE Access, PP(99), 1–…
Seguridad en juegos online	Chatbot IA protectbot	Desarrollo de chatbot IA para reforzar seguridad infantil en videojuegos.	Faraz, et al. (2024)	Faraz, A., Ahsan, F., Mounsef, J., Karamitsos, I., & Kanavos, A. (2024). Enhancing Child Safety in Online Gaming: The Development and Application of Protectbot, an AIPowered Chatbot Framework. Information, 15(4), 233. https://doi.org/10.3390/info15040233

Riesgos digitales en menores / Apoyo a poblaciones vulnerables	Compañero virtual con IA para reconocimiento de cyberbullying	Herramienta web con T5-small y MobileBERT que ayuda a adolescentes con autismo a identificar y responder al cyberbullying mediante escenarios virtuales.	Ferrer, Ali y Hughes (2024)	Ferrer, R., Ali, K., & Hughes, C. (2024). Using AI-based virtual companions to assist adolescents with autism in recognizing and addressing cyberbullying. *Sensors*. https://doi.org/10.3390/s24123875
Seguridad móvil	Phishing y ML	Modelo supervisado para detectar sitios web fraudulentos en dispositivos móviles inteligentes.	Gómez Varela (2019)	Gómez Varela, C. A. (2019, diciembre). Detección de sitios web falsos en dispositivos móviles mediante algoritmos supervisados de aprendizaje automático. [Tesis, Universidad CENFOTEC] Repositorio Institucional CENFOTEC
Capacitación parental	Guía multimedial	Se destaca la importancia de formar a padres en riesgos digitales.	Guevara Tibaduiza (2023)	Guevara Tibaduiza, G. M. (2023). Ciberseguridad: guía práctica para la protección de niños entre los 4 y 12 años de los riesgos en internet. [Trabajo de grado - Especialización, Universidad Católica de Colombia]. RIUCaC.
Seguridad digital en móviles	Factores de riesgo de daño digital	Una proporción relevante de niños de 8 años ha experimentado daño digital, asociado al uso individual de dispositivos y supervisión limitada.	Hinkley et al. (2024)	Hinkley, T., Fisher, J., Iyer, P., & Berrigan, L. (2024). Digital media in early childhood: risk factors for online harm and psychosocial correlates. Frontiers in Developmental Psychology, 3, 1390276.
Riesgos en línea	Videojuegos online	Se identifican experiencias adversas en videojuegos multijugador.	Jerge (2024)	Jerge, E. (2024, septiembre 5). Adverse Experiences in Online Gaming [Informe]. ClickSafe Intelligence.
Riesgos en Roblox	Seguridad en Roblox	Se analiza riesgos de explotación y seguridad digital en Roblox.	Jerge (2024)	Jerge, E. (2024, octubre 30). Building Blocks: Investigating Child Exploitation and Safety Features of Minecraft [Informe].

Riesgos en videojuegos	Seguridad en Minecraft	Se examinan riesgos de contacto inapropiado en Minecraft.	Jerge (2025)	Jerge, E. (2025, julio 14). Beyond the Virtual Playground [Informe]. ClickSafe Intelligence.
Cumplimiento legal en las aplicaciones	Evaluación GDPR	Se detectan incumplimientos en protección de datos en aplicaciones enfocadas a infantes.	Lopes et al. (2023)	Lopes, R., Ta, V. T., & Korkontzelos, I. (2023, mayo). On the conformance of Android applications with children's data protection regulations and safeguarding guidelines. 10.48550/arXiv.2305.08492
Biometría conductual	Modelos de ML táctiles	Comparación de algoritmos de ML para clasificar usuarios infantiles.	Mane et al. (2023)	Mane, S., Mandhan, K., Bapna, M., & Terwadkar, A. (2023). Child Detection by Utilizing Touchscreen Behavioral Biometrics on Mobile. In G. Ranganathan, G. A. Papakostas & Á. Rocha (Eds.), Inventive Communication and Computational Technologies (pp. 227–244). Springer. 10.1007/978-981- 99-5166-6_16
Seguridad digital en móviles	Machine Learning aplicado a malware	Se desarrolla un modelo de aprendizaje automático orientado a la detección de malware en dispositivos Android.	Mirabet Herranz (2024)	Mirabet Herranz, J. (2024). Aplicación de técnicas de machine learning para la detección de malware en dispositivos Android. [Trabajo fin de grado, Universitat Politècnica de València]. Repositorio institucional UPV
Detección de edad	Interacción táctil con dispositivo móvil	Identificación automática de menores mediante comportamientos táctiles en el dispositivo móvil.	Nguyen, Roy & Memon (2019)	Nguyen, T., Roy, A., & Memon, N. (2019, julio). Kid on The Phone! Toward Automatic Detection of Children on Mobile Devices. Computers & Security, Volume 84, Pages 334-348, ISSN 0167-4048. https://doi.org/10.1016/j.cose.2019.04.001
Riesgos digitales en menores / Herramientas de control parental	Plataforma web de detección de cyberbullying validada con actores educativos	SocialBullyAlert integra IA y redes neuronales para detectar contenido ofensivo en menores, validada por docentes, psicólogos y padres de familia.	Nina-Gutiérrez et al. (2024)	Nina-Gutiérrez, E. A., Pacheco-Alanya, J. E., & Morales-Arevalo, J. C. (2024). SocialBullyAlert: A web application for cyberbullying detection on minors' social media. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2024.0150776

Ciberseguridad en aplicaciones móviles	Amenazas digitales	Se describen las principales modalidades de ataques dirigidos a los teléfonos inteligentes mediante aplicaciones móviles.	Olguín Romero & Arana-Llanes (2024)	Olguín Romero, A., & Arana-Llanes, J. Y. (2024). Ataques a teléfonos celulares mediante el uso de aplicaciones móviles: una revisión narrativa. Tecnociencia Chihuahua, sin paginación
Apoyo a crianza digital	Recomendador IA parental	Sistema móvil basado en ML que recomienda estrategias de crianza digital.	Özaydın Aydoğdu (2023)	Özaydın Aydoğdu, Ö. (2023, agosto 11). Designing, Developing and Examining the Effectiveness of a Machine Learning–Based Mobile Recommendation System for Parents' Digital Parenting Skills. Child & Family Social Work. https://doi.org/10.1111/cfs.13218
Sistemas IDS y ML	Minería de datos IDS	Propuesta de integración de ML y minería de datos para mejorar la detección de intrusiones.	Pellufó, Capobianco, & Echaiz (2014)	Pellufó, I. C., Capobianco, M., Echaiz, J. (2014, mayo). Machine Learning aplicado en sistemas de detección de intrusos. XVI Workshop de Investigadores en Ciencias de la Computación (WICC). [Trabajo de fin de grado, Universidad Nacional de la Plata]. Repositorio institucional de la UNPL.
Ciberseguridad infantil	Web 3.0 y Metaverso	Se estudian riesgos emergentes en nuevas plataformas.	Peñuela Jiménez (2023)	Peñuela Jiménez, L. C. (2023). Ciberseguridad: el peligro al acecho de los niños, niñas y adolescentes. Informática y Derecho: Revista Iberoamericana de Derecho Informático (segunda época), 15–28.
Prevención ciberacoso	Juegos serios con IA	Juegos apoyados en IA permiten estudiar y prevenir el ciberacoso.	Pérez et al (2023)	Pérez, J., Castro, M., Awad, E., & López, G. (2023). Virtual Harassment, Real Understanding: Using a Serious Game and Bayesian Networks to Study Cyberbullying.
Vulnerabilidad infantil ante la tecnología	Perspectiva familiar en el uso digital	La dinámica familiar incide directamente en la vulnerabilidad infantil cuando no existe mediación activa en el uso de la tecnología.	Ptaszek & Pyżalski (2023)	Ptaszek, G., & Pyżalski, J. (2023). Children’s vulnerability to digital technology within the family: A scoping review. Societies, 13(1), 11.

Protección infantil digital	Deepfake y regulación digital	Se analiza el deppdake como amenaza a la privacidad de menores y se propone regulación.	Puetate et al. (2024)	Puetate Paucar, J. M., Toro Hernández, G. A., & Coka Flores, D. F. (2024). Deepfake: un desafío latente para la seguridad de los menores ecuatorianos en el entorno digital. Dilemas contemporáneos: Educación, Política y Valores.
Tendencias ML en ciberseguridad	Análisis bibliométrico PRISMA	Estudio PRISMA que identifica tendencias actuales en ML aplicándolo en ciberseguridad.	Ramírez et al (2023)	Ramírez, D. M., Garcés-Giraldo, L. F., Doria-Orozco, T., Franco-Castaño, S., Valencia-Arias, A., Rodríguez-Correa, P. A., & Espinoza Román, J. (2023). Tendencias investigativas en el uso de Machine Learning en la ciberseguridad. Revista Ibérica de Sistemas e Tecnologias de Informação, 60–72.
Ciberbullying / Optimización de hiperparámetros	SVM con Cuckoo Search y TF-IDF	Optimización de SVM mediante Cuckoo Search para detección de cyberbullying en árabe, mejorando la accuracy de 85.8% a 87.1%.	Saadi y Dhannoon (2023a)	Saadi, M. Q., & Dhannoon, B. N. (2023). Arabic cyberbullying detection using support vector machine with cuckoo search. *Iraqi Journal of Science*. https://doi.org/10.24996/ijs.2023.64.10.37
Ciberbullying / Comparación de modelos boosting	Random Forest vs. XGBoost con Cuckoo Search en árabe	XGBoost (0.861) y RF (0.849) comparados en detección de cyberbullying en árabe; tras optimización, RF alcanzó 0.867.	Saadi y Dhannoon (2023b)	Saadi, M. Q., & Dhannoon, B. N. (2023). Comparing random forest vs. extreme gradient boosting using cuckoo search optimizer for detecting Arabic cyberbullying. *Iraqi Journal of Science*. https://doi.org/10.24996/ijs.2023.64.9.40
Ciberbullying / Comparación de algoritmos ML	Clasificadores clásicos con CountVectorizer sobre FormSpring	Comparación de 9 algoritmos ML para detección binaria de cyberbullying; AdaBoost obtuvo la mejor accuracy con 86.52%.	Sahana, Anil Kumar y Darem (2023)	Sahana, V., Anil Kumar, K. M., & Darem, A. A. (2023). A comparative analysis of machine learning techniques for cyberbullying detection on FormSpring in textual modality. *International Journal of Computer Network and Information Security*. https://doi.org/10.5815/ijcnis.2023.04.04
Inteligencia artificial en delitos digitales	IA en investigación digital	Se examina el uso de la IA en investigaciones relacionadas con grooming online en el ámbito judicial español.	Salat Paisal, (2025)	Salat Paisal, M. (2025). El uso de la inteligencia artificial en la investigación policial y judicial de delitos de online child sex grooming en España. IDP. Revista de Internet, Derecho y Política, (42), 1–12.

Riesgos digitales en menores y grooming online	Permite detectar en tiempo real con embeddings convesacionales	En este entorno Safecon emplea un modelo que combina lo siguiente, universal sentence encoder, pcs y k-means, para poder identificar patrones en conversaciones en tiempo ral, con un fin único de poder orientar acciones de moderación y protección.	Sallahuddin et al. (2025)	Sallahuddin, S., Hameed, M. F., Ismail, M., Ali, S., & Ahmed, A. (2025). SafeCon: AI-powered real-time cyber grooming detection system. *VFAST Transactions on Software Engineering*. https://doi.org/10.21015/vtse.v13i2.2118
Digital risks to minors and online grooming	Detecta contextualmente con transformers (BERT, RoBERTa)	Plantea combinar el análisis a nivel de mensajes con la determinación del contexto conversacional para detectar de una manera mucho más robusta el grooming que con texto aislado.	Street et al. (2025)	Street, J., Ihianle, I. K., Olajide, F., & Lotfi, A. (2025). Enhanced online grooming detection employing context determination and message-level analysis. *Intelligent Systems with Applications*. https://doi.org/10.1016/j.iswa.2025.200607
Ciberbullying	Revisión de técnicas para el uso de Machine Learning en ciberbullyng	Resumen del marco metodologico que predomina en el área: recolección, procesamiento, extracción de características y clasificación, y orientación de investigaciones futuras.	Sultan et al. (2023)	Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., et al. (2023). A review of machine learning techniques in cyberbullying detection. *Computers, Materials & Continua*. https://doi.org/10.32604/cmc.2023.033682
Privacidad en aplicaciones usadas por los menores	Brechas de privacidad en las aplicaciones	Muchas de las aplicaciones enfocadas a un público infantil toman datos sensibles sin transparencia y permisos adecuados.	Sun et al. (2023)	Sun, R., Xue, M., Tyson, G., Wang, S., Camtepe, S., & Nepal, S. (2023, abril 30). Not Seen, Not Heard in the Digital World! Measuring Privacy Practices in Children's Apps. Association for Computing Machinery. 10.1145/3543507.3583327
Interacción infantil en entornos moviles	Dataset ChildCI	Presenta un dataset longitudinal para detectar la	Tolosana et al. (2022)	Tolosana, R., Ruiz-Garcia, J. C., VeraRodriguez, R., Herreros-Rodriguez, J., Romero-Tapiador, S., Morales, A., & Fierrez, J. (2022). Child-Computer Interaction with Mobile Devices: Recent Works, New Dataset, and Age

		edad mediante interacción neuromotora.		Detection. IEEE Transactions on Emerging Topics in Computing. 10.1109/TETC.2022.3150836.
Ciberatacantes en Colombia	Riesgos y sus consecuencias	Se analizan ciberataques ocurridos en Colombia entre el 2028 y 2020 y sus efectos en la población.	UNAD (2021)	Universidad Nacional Abierta y a Distancia, UNAD (2021). Ciberataques, riesgos y consecuencias que han afectado a la población colombiana entre los años 2018 y 2020. [Proyecto de grado, Universidad Nacional Abierta y a Dustancia]. Repositorio Institucional UNAD.
Protección infantil con IA	Uso ético de IA	Se reflexiona sobre la aplicación ética de la IA para reducir la violencia digital infantil.	Urbano (2024)	Urbano, R. (2024). Strategies to protect children from violence in artificial intelligence. Rotura – Revista de Comunicação, Cultura e Artes, 40-49.
Protección infantil	Tecno regulación y salvaguardas inteligentes	Se propone el análisis de gestos táctiles mediante ML para distinguir entre menores y adultos y activar protecciones automáticas.	Zaccagnino, Capo, Guarino, Lettieri & Malandrino (2021)	Zaccagnino, R., Capo, C., Guarino, A., Lettieri, N., & Malandrino, D. (2021). Techno-regulation and intelligent safeguards: Analysis of touch gestures for online child protection. Multimedia Tools and Applications, 80(21–23), 32159–32186.

Nota. En la tabla se resumen estudios teóricos, empíricos y de revisión sobre protección infantil en contextos digitales enfocados a menores, el uso de Inteligencia Artificial aplicada en el área de ciberseguridad y detección de riesgos en dispositivos móviles. Estos estudios abarcan investigaciones académicas indexadas, trabajos de grado, revisiones narrativas, informes técnicos y propuestas normativas que están publicadas entre los años 2014 y 2025. Esta información presentada permite identificar técnicas metodológicas, enfoques tecnológicos como el Machine Learning y biometría conductual), así como la identificación de vacíos existentes en la integración de soluciones automatizadas orientadas en la protección integral de los menores en el uso de dispositivos móviles en Colombia. Elaboración propia.

Metodología

Enfoque y alcance de la investigación

La investigación se llevó a cabo desde un enfoque cuantitativo con alcance descriptivo–correlacional, en concordancia con la hipótesis del estudio: el cual es evaluar si los patrones de interacción y conducta digital permiten detectar y anticipar riesgos mediante un modelo de Machine Learning (Hernández-Sampieri & Mendoza, 2018). El diseño es de tipo no experimental, dado que las variables no se manipulan; se analizarán asociaciones y desempeño predictivo a partir de datos tanto como obtenidos y simulados.

El trabajo se estructurará en fases que permitan conectar las evidencias empíricas y el desarrollo del prototipado:

Revisión de la literatura y delimitación de los escenarios de riesgo (PRISMA). Se utilizará PRISMA para reconocer las amenazas más frecuentes, variables empleadas y estrategias de modelado en seguridad digital infantil. Este resultado será una matriz con categorías de riesgo y criterios operativos para su modelado posterior (Page et al., 2021; Livingstone, 2013; Romero & Arana-Llanes, 2024).

Estudio e inclusión de encuestas a padres y menores en Colombia. Se realizarán encuestas estructuradas previamente aplicadas en Seminario de Investigación a padres de familia, niños y adolescentes, orientadas a revisar los hábitos de uso, la supervisión parental, la alfabetización digital por los padres, la exposición percibida a los riesgos digitales y eventos donde hubo interacción de alto riesgo. Estas encuestas se utilizarán como una evidencia empírica para el análisis descriptivo-correlacional y como un elemento de ajuste para simulación de datos, como rangos de horas de uso plausibles o niveles de supervisión parental.

Modelado de variables y construcción del dataset simulado. A partir de la literatura y de las evidencias de las encuestas realizadas a padres y menores, se trabajará con un dataset

sintético el cual representa patrones de uso. La muestra analítica para el componente de Machine Learning está compuesta por 3.000 registros y 20 campos, donde se incluye: edad, género, horas_uso_diario, uso_nocturno, descargas_fuentes_externas, enlaces_sospechosos, enlaces_sospechosos_detectados, interacciones_desconocidos, numero_interacciones_desconocidos, permisos_excesivos_apps, tipo_app_principal, control_parental, control_parental_activado, supervision_parental_percibida, alfabetizacion_digital, impulsividad, autoestima, necesidad_aprobacion_social, tolerancia_frustracion y la variable objetivo nivel_riesgo. Se asume que la simulación puede introducir sesgos; por ello, los hallazgos se interpretan como evidencia de viabilidad técnica, y no como una estimación definitiva en población real.

Se entrena un modelo supervisado para clasificar patrones como “seguros” o “riesgosos”. Se aplica una partición del dataset y validación una k-fold. Junto con las métricas globales, se revisa la matriz de confusión y los errores de clasificación para controlar lecturas optimistas basadas solo en exactitud, particularmente por el efecto que producen la existencia de falsos positivos y falsos negativos en un contexto preventivo (Zaccagnino et al., 2021; Mirabet, 2024).

En el desarrollo del prototipo funcional e integración. Se implementa una arquitectura cliente–servidor, donde el backend en “Python” ejecuta el preprocesamiento e inferencia en segundo plano y donde logra generar alertas preventivas según accesos definidos. Donde se fijarán parámetros de operación para lograr minimizar el impacto en la experiencia del menor (Mirabet, 2024).

Instrumentos

Protocolo PRIMA y matriz de extracción para derivar escenarios y variables (page et al., 2021).

Encuestas estructuradas a padres y menores en Colombia (cuestionarios).

Especificaciones del Dataset simulado: diccionario de datos, rangos, reglas de generación y criterio de etiquetado de nivel_riesgo.

Pipeline de preprocesamiento y modelo supervisado.

Prototipo cliente-servidor con backend en “Python” y mecanismo de registro de alertamientos

Validación de instrumentos

Las variables (Riesgo, exposición, supervisión y conducta), también escenarios y cuestionarios se fundamentan en la literatura que se seleccionó. La consistencia interna se medira con el alfa de Cronbach, y para la confianza en el modelo se verificará mediante validación cruzada (Zaccagnino et al., 2021).

Tabla 2

Distribución de la muestra según tipo e instrumento aplicado.

Tipo de participante	Instrumento aplicado	Frecuencia (n)	Porcentaje
Estudiantes	Encuesta estructurada	122	62.2%
Padres de familia	Encuesta estructurada	74	37.8%
Docentes	Entrevista semiestructurada	3	
Total		196	100%

Nota. la tabla evidencia el total de la muestra de investigación, donde está compuesta por 196 encuestas y 3 entrevistas. Donde participaron 122 estudiantes que representa el 62.2% del total y 74 padres de familia representando el 37.8% del total de las encuestas. Estos instrumentos permiten conocer y estructuras como se maneja y se conocen los riesgos en los

dispositivos de los menores para complementar la interpretación de los resultados estadísticos. Elaboración propia.

Técnicas para el análisis de la información

El análisis de hábitos de uso, supervisión y exposición combinara estadística descriptiva y una correlación spearman. Para el modelo de Machine Learning se evaluará su desempeño mediante métricas tradicionales como lo son:

- accuracy
- precisión
- F1
- ROC-AUC
- La matriz de confusión (falsos positivos y negativo)
- Procesamiento de datos mediante Python con pandas y TensorFlow

Resultados

Los instrumentos para recolección de la información fueron diseñados tomando los diferentes factores involucrados en el contexto del estudio, Se aplicaron encuestas a padres de familia, niños y docentes. Esta segmentación permitió obtener perspectivas sobre la problemática analizada, así garantizando una visión más integral y objetiva de la realidad. Las encuestas se seleccionaron como instrumento principal debido a su capacidad para recopilar datos de manera estructurada, estandarizada y cuantificable. El análisis de esos resultados se organizará en secciones, según cada instrumento de medición, con el fin del facilitar su análisis, comparación e interpretación.

Los resultados obtenidos permiten responder al objetivo de esta investigación, orientado a identificar los riesgos digitales en dispositivos móviles usados por menores mediante técnicas de aprendizaje automático aplicadas tanto a texto como a variables de perfil. A partir del procesamiento de las encuestas, del dataset estructurado de perfiles y de un léxico especializado de riesgos, se establecieron dos conjuntos de entrenamiento complementarios, un componente textual y un componente de perfil.

El conjunto textual quedó conformado por 542 registros, mientras que el conjunto de perfiles estuvo generado por 3000 observaciones y 19 variables relacionadas con hábitos de uso, supervisión y factores psicosociales de los menores.

En el componente textual se observó una gran cantidad de expresiones asociadas con situaciones de riesgo digital. Del total de 542 registros, 519 correspondieron a casos clasificados como riesgo y 23 a casos seguros, lo que representa 95,76 % y 4,24 %, respectivamente. Asimismo, la mayor proporción de ejemplos se centró en la categoría survey_risk, con 296 registros (54,61 %), seguida de un lenguaje abusivo con 130 casos (23,99 %), grooming con 42 (7,75 %), ciberbullying con 29 (5,35 %) y riesgo personal con 22 (4,06 %). Estos resultados muestran que las respuestas abiertas recolectadas en las encuestas

aportaron una variedad importante de expresiones o manifestaciones de riesgo y enriquecieron el componente textual del sistema para su detección.

De manera complementaria, se identificó que la base léxica de riesgo está compuesta por 229 patrones. De ellos, 136 correspondieron a lenguaje abusivo, 42 a grooming, 29 a ciberbullying y 22 a riesgo personal. Dentro del total, 63 patrones fueron definidos como alertas inmediatas. Este hallazgo evidencia que la propuesta no depende únicamente del aprendizaje estadístico, sino que también necesita reglas explícitas construidas a partir de un conocimiento para detectar agresiones verbales, solicitudes indebidas de información y expresiones asociadas con amenazas.

En cuanto al desempeño del clasificador textual, las métricas reportadas fueron favorables. El modelo alcanzó una exactitud de 94,85 %, una precisión de 95,56 %, un recall de 99,23 % y un F1 de 97,36 %. Estos valores evidencian una alta capacidad para reconocer mensajes problemáticos dentro del entorno de prueba. En la validación funcional, por ejemplo, expresiones como “mándame una foto ahora” fueron clasificadas como grooming con alerta inmediata y frases ofensivas como “eres una puta” fueron clasificadas como lenguaje abusivo. En contraste, una expresión neutra como “vamos a jugar esta tarde” no activó alerta porque su puntuación quedó por debajo del umbral establecido, aunque sí registró una probabilidad de riesgo intermedia, lo que muestra que el sistema no opera únicamente con una clasificación binaria, sino con una lógica de decisión que se basa en umbrales.

En el componente de perfil también se identificaron hallazgos relevantes, aunque con menor rendimiento predictivo. El dataset de 3000 registros se distribuyó en 1716 perfiles de bajo riesgo (57,2 %) y 1284 de alto riesgo (42,8 %). Además, el promedio general de edad fue de 12,45 años y el promedio de uso diario del dispositivo fue de 6,24 horas. En este caso, el clasificador alcanzó una exactitud de 79,47 %, una precisión de 74,93 %, un recall de 78,19 % y un F1 de 76,52 %. Aunque estos indicadores fueron inferiores a los obtenidos por el

componente textual, los resultados muestran que el análisis de hábitos de uso, exposición y supervisión también permite anticipar escenarios de vulnerabilidad digital.

Como síntesis general de los datos procesados, en la Tabla 3 se presenta la composición de las fuentes utilizadas para el entrenamiento del sistema.

Tabla 3

Composición de los datos procesados para el entrenamiento.

Conjunto	Tamaño final	Distribución principal
Datos textuales	542	519 riesgo y 23 seguros
Datos de perfil	3000	1284 alto riesgo y 1716 bajo riesgo
Encuesta a niños	113 respuestas	Insumo para fragmentos seguros y de riesgo
Encuesta a acudientes	74 respuestas	Insumo para fragmentos seguros y de riesgo
Léxico de riesgo	229 patrones	63 de alerta inmediata

Nota. A partir de los archivos del proyecto reportados en el proceso de entrenamiento. Elaboración propia

Descripción de la Información Recolectada

Durante el proceso de la investigación se recolectaron datos mediante (encuestas, entrevistas, instrumentos aplicados), que se aplicaron a (199 participantes), pertenecientes a un grupo especifico de padres, menores y docentes. La información que se obtuvo permitió analizar las variables relacionadas con el objeto de estudio, enfocándose principalmente en, (la hora de uso del dispositivo, edad que adquirían de un dispositivo, nivel de conocimiento de los padres, a su vez 3 entrevistas a docentes lo que permitió conocer la percepción que se tiene sobre la IA aplicada a la seguridad infantil. Este conjunto de datos cuantitativos y cualitativos sirvió como una base para realizar el cruce de las variables y así estructurar el diagnóstico inicial del entorno digital de los menores en Colombia.

Para la representación de los textos se utilizó TF-IDF con unigramas y bigramas, limitando el espacio de características a 2500 variables y posteriormente se entrenó una

regresión logística balanceada. La partición de entrenamiento y prueba se realizó de forma estratificada en proporción 75 % y 25 %, con el fin de mantener la distribución de clases en ambos conjuntos. Esta metodología permitió obtener un modelo textual que permite reconocer patrones lingüísticos asociados con grooming, lenguaje abusivo, ciberbullying y riesgo personal.

Procesamiento de Datos

Una vez recolectados los datos de las encuestas se organizaron y limpiaron datos innecesarios que impedían el correcto análisis de estos. En el caso de los datos cuantitativos, se realizaron medidas estadísticas básicas como en tablas y representados en graficas para facilitar su entendimiento.

En el caso de los datos cualitativos, se realizó un proceso de categorización, identificando patrones comunes y agrupando las respuestas según el tema. Esto permitió establecer tendencias y relaciones significativas entre las variables analizadas.

En el modelo de perfil, el procesamiento se realizó sobre 19 variables numéricas y categóricas relacionadas con edad, tiempo de uso, uso nocturno, descargas externas, interacciones con desconocidos, enlaces sospechosos, control parental, autoestima, impulsividad, necesidad de aprobación social y alfabetización digital. Los valores faltantes se imputaron mediante la mediana y posteriormente, las variables fueron estandarizadas antes del entrenamiento de una regresión logística balanceada. Este procedimiento permitió comparar perfiles de bajo y alto riesgo bajo un esquema estadístico homogéneo y facilitar la identificación de patrones diferenciales entre ambos grupos.

Las principales diferencias observadas entre los perfiles de bajo y alto riesgo se resumen en la Tabla 4.

Tabla 4

Principales diferencias entre perfiles de bajo y alto riesgo.

Variable	Bajo riesgo	Alto riesgo
Horas de uso diario	5,49	7,24
Uso nocturno	30,9 %	53,2 %
Interacciones con desconocidos	3,46	5,36
Enlaces sospechosos detectados	3,13	6,12
Control parental activado	43,0 %	26,9 %

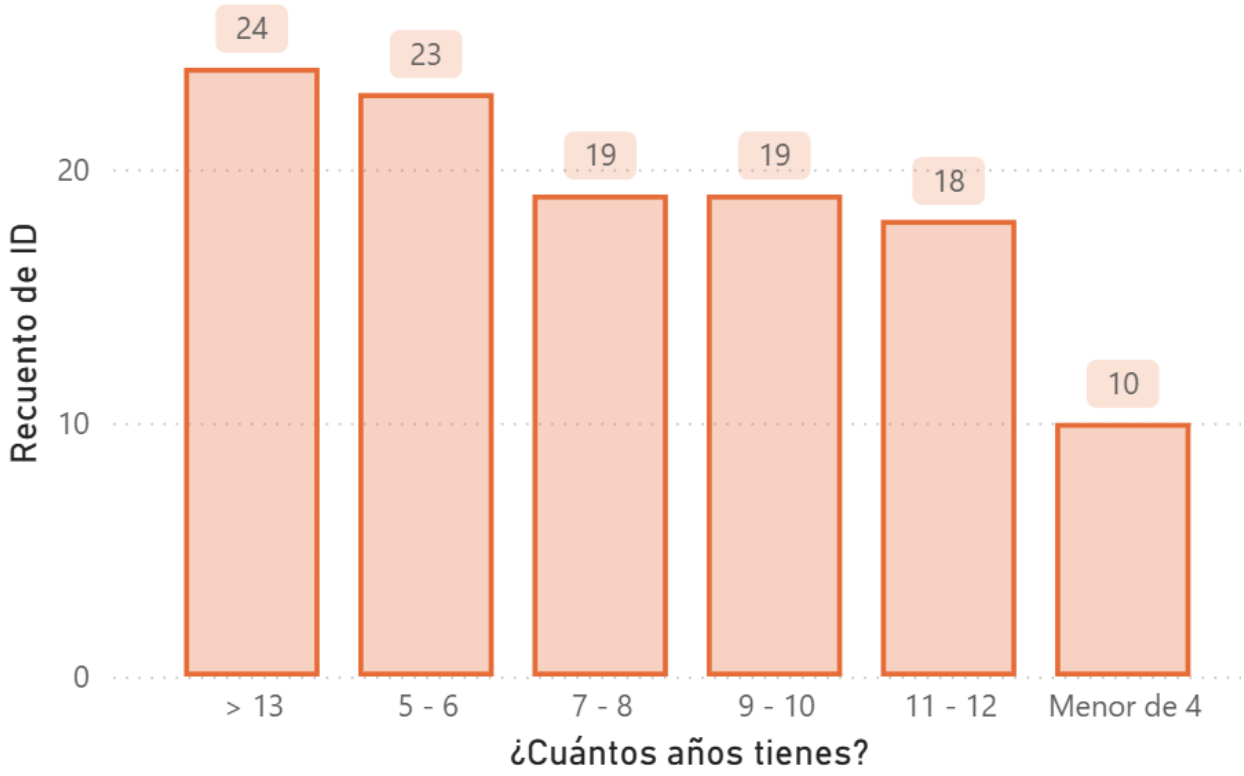
Nota. Elaboración propia.

Encuesta a los Niños

Esta encuesta es necesaria, ya que permite conocer el nivel de conocimiento que tienen los menores frente a los riesgos que existen al usar un dispositivo móvil. Dicho esto, se realizará un análisis detallado de cada una de las preguntas, haciendo una comparación de cada una de las gráficas. El proceso de recolección de datos no alcanzo el total esperado de 122 encuestas, sin embargo, se obtuvo una participación de 113 niños, lo que constituye un número suficiente para el análisis en cuestión.

Figura 2

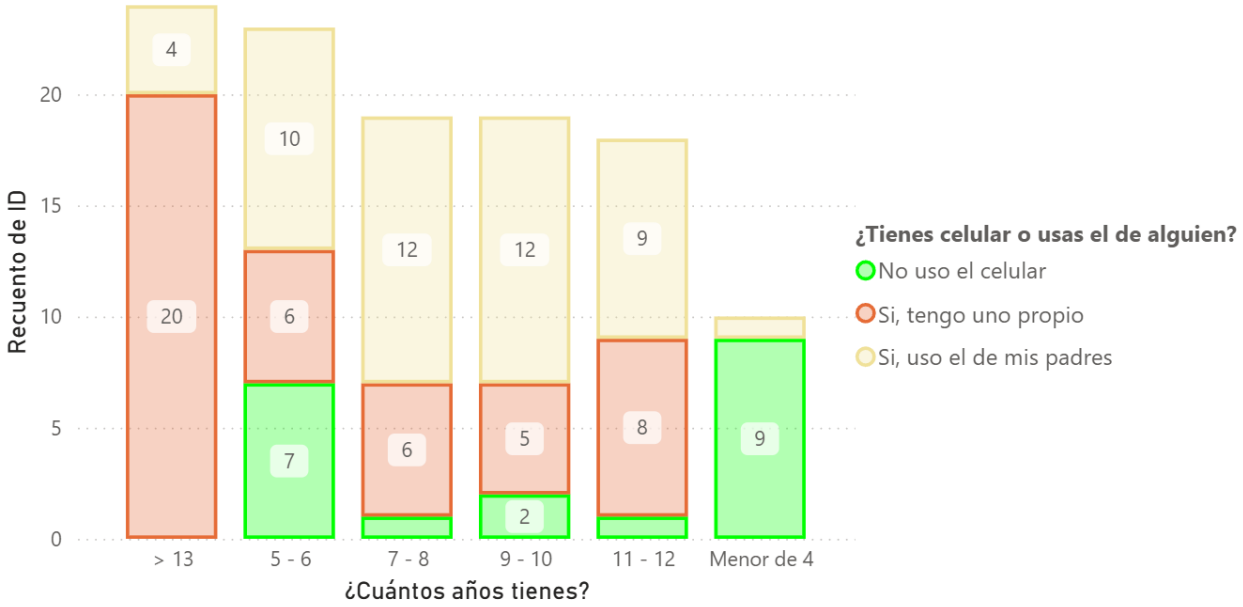
Edades de los Niños



Nota. Al revisar las edades, se ve que el grupo está bastante equilibrado entre los 5 y 12 años, rango que constituye la investigación con un promedio de 19 personas por categoría. Sin embargo, el pico de la participación lo tienen los más grandes (n > 13), mientras que los más pequeños, son por mucho, el grupo menos numeroso, dando como resultado que entre los 113 encuestados los más grandes usan un dispositivo móvil. Elaboración propia

Figura 3

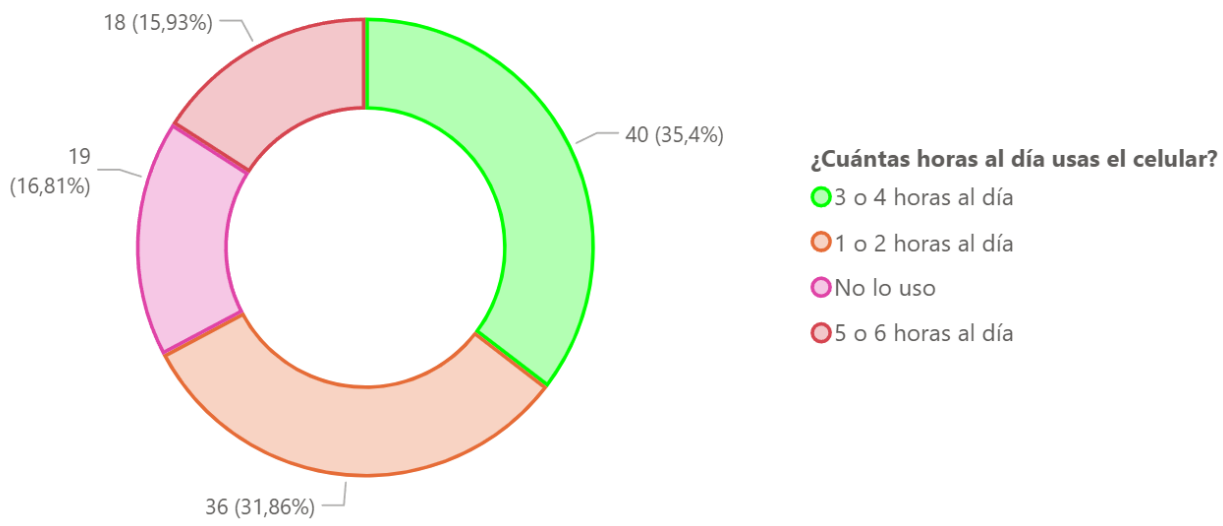
Acceso a dispositivos móviles según el rango de edad



Nota. Lo más claro aquí es la brecha generacional en la propiedad del celular. Mientras que en el grupo de mayores de 13 años casi todos tienen dispositivos propios, en los niños de 5 años a 12 años predomina el uso del teléfono de los padres. Un dato que salta a la vista es que, en el grupo de los menores de 4 años, casi nadie usa el celular por no decir que ninguno, manteniéndose al margen de esta tendencia. A partir de los 11 – 12 años se marca la brecha que va en aumento notable de quienes ya tienen celular propio, entre los 7 y 10 años, la gran mayoría depende del teléfono de los padres. Elaboración propia.

Figura 4

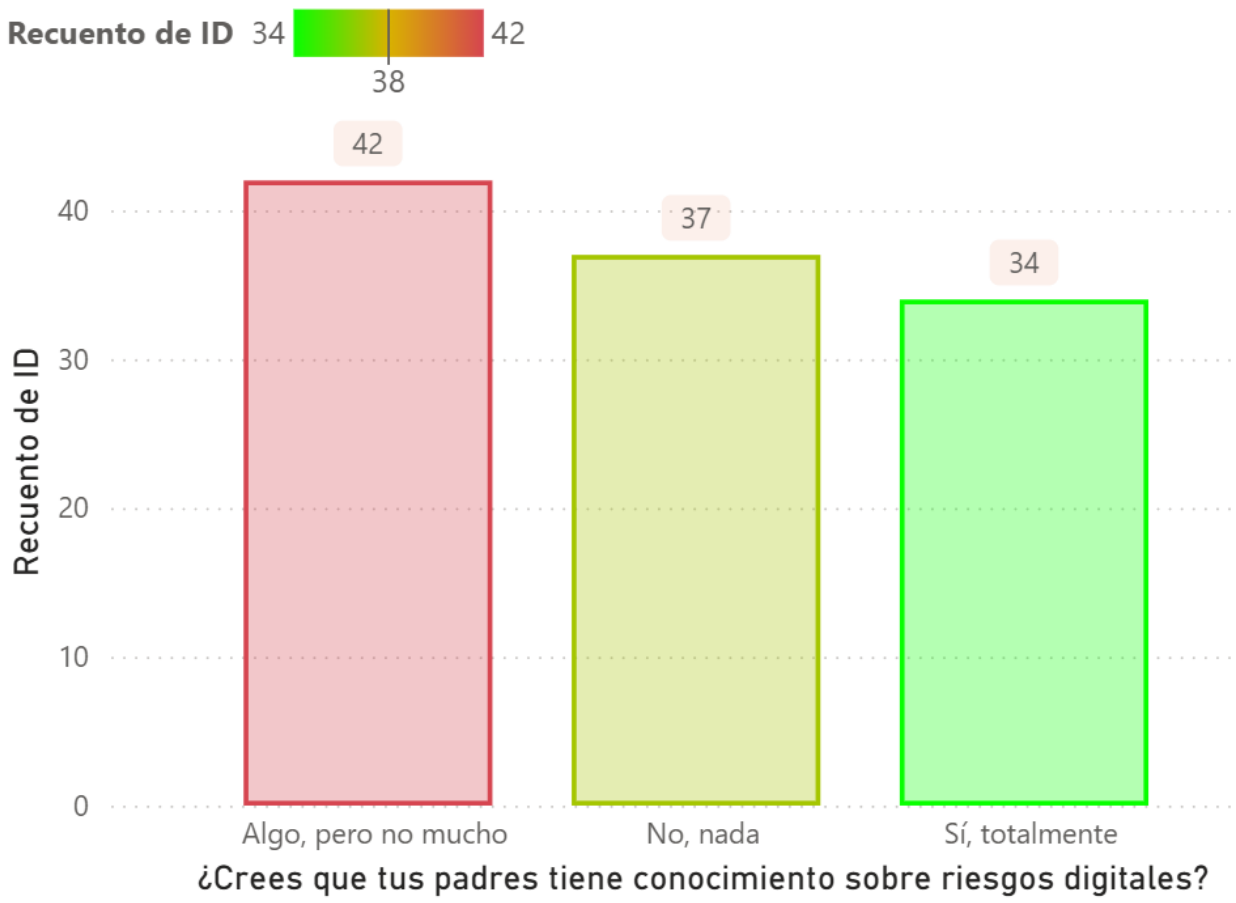
Tiempo promedio de uso diario del celular



Nota. Lo que más resalta es que la gran mayoría de los niños (más del 67%) pasa entre 1 y 4 horas al día pegados en el celular. Es el estándar de consumo. Por otro lado, solo un grupo muy pequeño (cerca del 16%) admite que pasan más tiempo entre 5 a 6 horas, y curiosamente, casi la misma cantidad de personas asegura que directamente no lo usa. Elaboración propia.

Figura 5

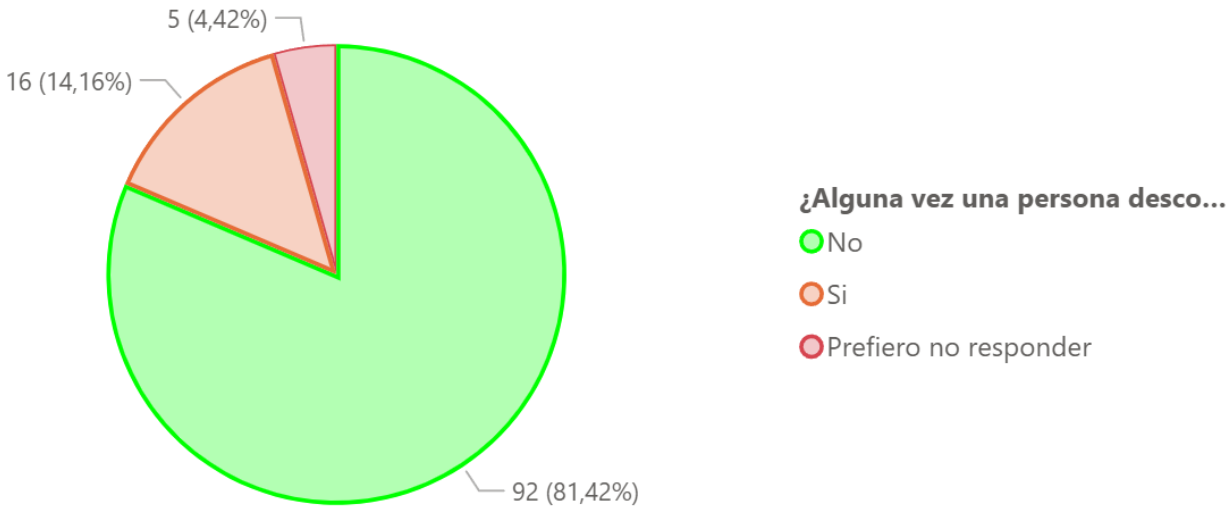
Percepción sobre el conocimiento de los padres en riesgos digitales



Nota. La opinión está bastante dividida, pero lo que más resalta es que la mayoría siente que sus padres tienen una idea superficial sobre riesgos en internet. El grupo que respondió “algo, pero no mucho” es el más numeroso (42 personas). Lo más preocupante es que el segundo grupo más grande (37 personas) considera que sus padres no saben “nada” del tema, superando a los que confían totalmente en el conocimiento de sus padres. En resumen, hay una sensación general de que falta preparación en cada para enfrentar los temas como la seguridad en los dispositivos móviles que usan a diario los menores. Elaboración propia.

Figura 6

Interacciones con personas desconocidas



Nota. La grafica muestra los resultados sobre la experiencia con personas desconocidas que los niños encuestados han tenido, se logra observar que la gran mayoría de los menores corresponde a (91 personas, que respondieron que no han tenido contacto con personas desconocidas mientras hacen uso del celular. Por otro lado (16 personas) afirmo haber tenido dicha experiencia, mientras que el dato más pequeño de la encuesta (5 personas) prefirieron no revelar esa información. Elaboración propia.

En la Tabla 5 se presentan las explicaciones escritas por los niños que dijeron haber estado en contacto con personas desconocidas mientras usaban el dispositivo móvil. Los reportes están relacionados sobre todo con requerimientos de datos personales, solicitudes de fotos y mensajes donde los amenazan en las redes sociales y videojuegos, lo que permite evidenciar la existencia de peligros vinculados a la interacción digital sin una supervisión apropiada.

Tabla 5

Descripción textual de las situaciones reportadas por contacto con desconocidos.

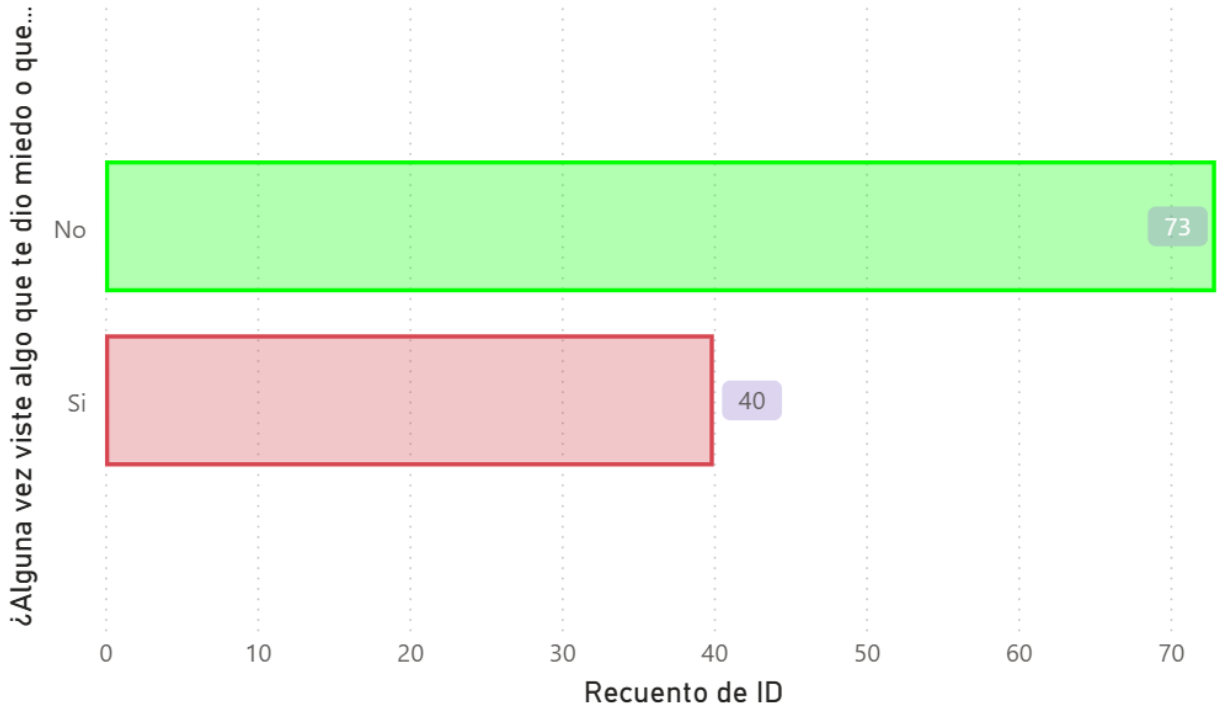
ID	Descripción textual reportada
S1	Por juegos, cuando entro a partidas una vez me dijeron donde vives
S2	Una persona una vez me pidió mi dirección en un juego
S3	una persona por un juego me pidió mi celular y donde vivía me dio mucho miedo
S4	En Facebook una persona me dijo que me daba dinero por mandar fotos mías
S5	En un juego una persona me insulto y me trato muy feo
S6	Me escribió un numero por wasap y me dio mucho miedo ya que no conocía esa persona
S7	Cuando le di a un anuncio en internet luego una persona que escribió por watsap, me dio mucho miedo y le dije a mi papa y el borro el mensaje
S8	En un juego me pidieron que mandara fotos a un número y me daban dinero
S9	Una persona por un juego me escribió y me dio miedo
S10	No me escribe una señora X, y pues no le contesté
S11	Una vez alguien me mando mensaje por una app y me pidió que le mandara fotos de mí, donde se viera mi cuerpo, pero no le mande nada ya que me dio mucho miedo ya que no sabía quién era esa persona
S12	Una persona me llamo para pedirme datos de mis papas
S13	Una vez una persona por una app me escribió que tenía mis contraseñas y que tenía que mandarle dinero para que no hiciera nada malo
S14	Una persona en un juego me pedía fotos de mi sin nada

S15 Un man quería que le mandara fotos mías sin nada

Nota. En la “Figura 6” podemos ver cuántos de los menores interactuaron con personas desconocidas, la tabla recopila los que marcaron “Si” y detallaron la experiencia que tuvieron en ese momento, donde la mayoría de ellos sintió demasiado temor cuando hablaron con estas personas, ya sea por juegos o redes sociales, uno de ellos describió Grooming digital cuando un desconocido lo contacto para manipularlo y decirle que tenía sus contraseñas, esto hace evidente que es necesario un aplicativo que pueda contrarrestar estas malas acciones. Elaboración propia.

Figura 7

Percepción de situaciones de temor o inquietud



Nota. Este grafico permite detallar las respuestas a la pregunta “¿Alguna vez viste algo que te dio miedo o que no entendiste?”, de un total de 113 menores encuestados, las respuestas que dominan son negativa, con (73 personas) donde indican que no han tenido esa experiencia.

Por el otro lado, (40 personas) respondieron de manera afirmativa, donde afirman haber presenciado situaciones que en algún momento cuando usaban el teléfono les causo temor. Elaboración propia.

La Tabla 6 muestra las circunstancias que los niños mencionaron al confirmar que sintieron miedo o confusión mientras hacían uso de un dispositivo móvil. Se nota que la mayor parte de los casos tienen que ver con mensajes raros, enlaces sospechosos, publicidad engañosa y contacto con contenido para los adultos. Esto hace posible detectar patrones de riesgo digital que repiten en contextos comunes de navegación.

Tabla 6

Descripción textual de situaciones que generaron miedo en línea.

ID	Descripción textual reportada
M1	No me acuerdo
M2	Videos de terror
M3	me pidieron mi número de celular
M4	mensaje donde me ganaba un celular
M5	mensajes extraños por correo
M6	me pedían mi dirección en el juego una persona que no conocía
M7	en Google me salió anuncios de que mi teléfono tenía virus
M8	me llego un mensaje donde me pedían mi teléfono
M9	mensajes en Google cuando buscaba para hacer mi tarea
M10	cosas en internet
M11	mensajes cuando intenté instalar algo desconocido en mi celular
M12	mensajes extraños
M13	Muerte de personas, uso de drogas
M14	cosas raras en Facebook

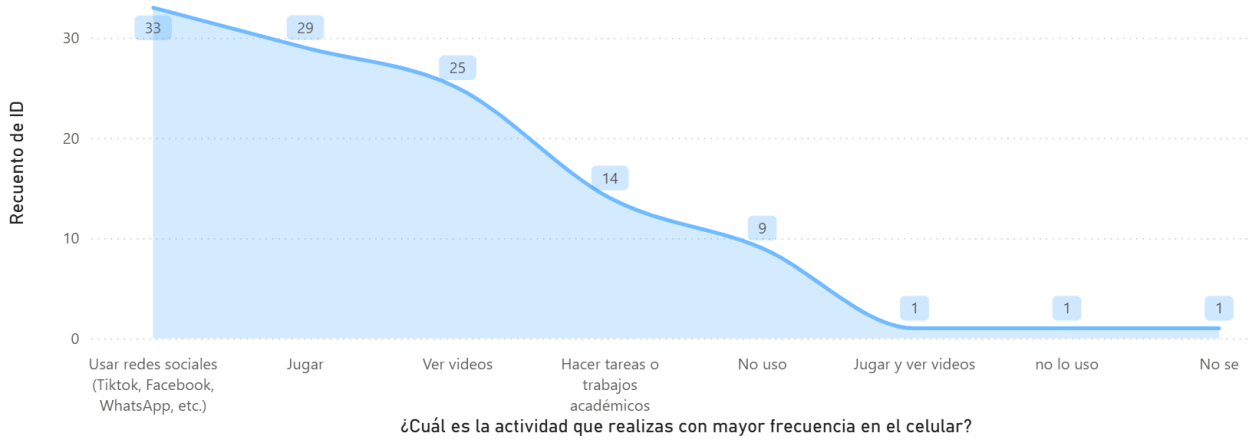
M15	me hice amigos de personas en un juego y me dijeron donde vivía
M16	Una vez vi un video con cosas feas y gritos, me asusté y le dije a mi mamá para que lo quitara
M17	Que me roben mis monedas del juego
M18	Me salieron cosas paranormales que no me dejaron dormir
M19	me salió una cosa rara en videos de redes
M20	me apareció un mensaje y le di sí sin leer
M21	mensajes raros en Instagram
M22	llamadas donde no contestan y cuelgan
M23	mensajes raros en aplicaciones
M24	videos que me asustaron en TikTok o así
M25	me robaron mis personajes
M26	me mandaron un link y cuando le di me puso algo raro en el celular
M27	anuncios raros y di like y me apareció que mi teléfono tenía virus
M28	me hablaron por el juego para pedirme fotos y no sabía qué hacer o decir
M29	una vez me llegó un mensaje de texto y entré y era para mandar cosas mías o de mi mamá
M30	Vi una imagen que se movía y me dio susto
M31	videos y mensajes raros en Facebook de alguien que no conocía
M32	me enviaron videos para gente grande y no sabía qué hacer
M33	en internet salen anuncios raros, en las aplicaciones no mucho
M34	anuncios diciendo que uno se ganó un celular, le di y se abrieron pestañas con cosas raras para gente grande
M35	se abren pestañas en el navegador cuando uno hace click en cualquier lado
M36	en un juego un muchacho me pedía mi Facebook y luego fotos a cambio de diamantes
M37	me llegó mensaje extraño al WhatsApp con un link, pero no lo abrí y mi mamá lo bloqueó

M38	en YouTube al dar en algo me mandó a otra página y no sabía qué hacer
M39	en una aplicación me mandaron un link que decía PDF, pero abajo decía HTML y lo borré porque investigué que era para robar información
M40	en un juego me dijeron algo que no entendí, pero no sé si era algo malo o bueno

Nota. La tabla permite evidenciar que los menores que participaron en la encuesta y identificaron algo que les dio miedo marcando con un “Si”, describen lo que les paso en entornos digitales, donde la mayoría fue la aparición de mensajes extraños, enlaces sospechosos, anuncios engañosos, solicitudes de datos personales y exposición al contenido para adultos. También se logra identificar manipulación en juegos y redirecciones a paginas desconocidas. En general, la participación de riesgo se centra en eventos inesperados que generan miedo, temor y desconfianza cuando se hace uso del dispositivo móvil. Elaboración propia.

Figura 8

Principales usos y actividades frecuentes en los dispositivos móviles

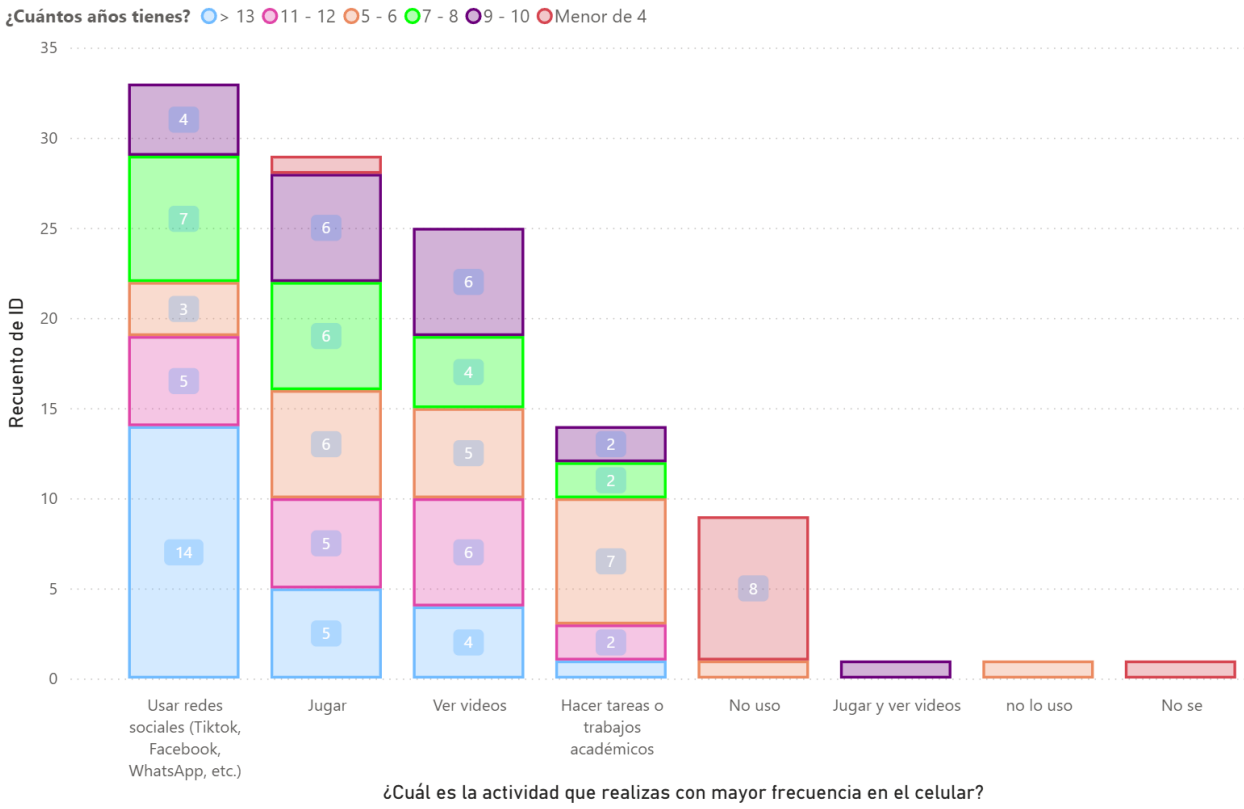


Nota. El grafico permite observar una clara tendencia hacia el entretenimiento y la socialización. El uso de redes sociales lidera con (33 personas), seguido de cerca por el juego (29 personas) esto da como resultado que dichas actividades son las que más frecuentan los

menores, y es donde más atención se debe prestar, el consumo de videos (25 personas). Se logra notan un descenso en actividades productivas como la escuela (14 personas) y finalmente una minaría reporta no usar el celular o no tener claro que es lo que hacen cuando lo usan. Elaboración propia.

Figura 9

Rango de edad y uso de actividades

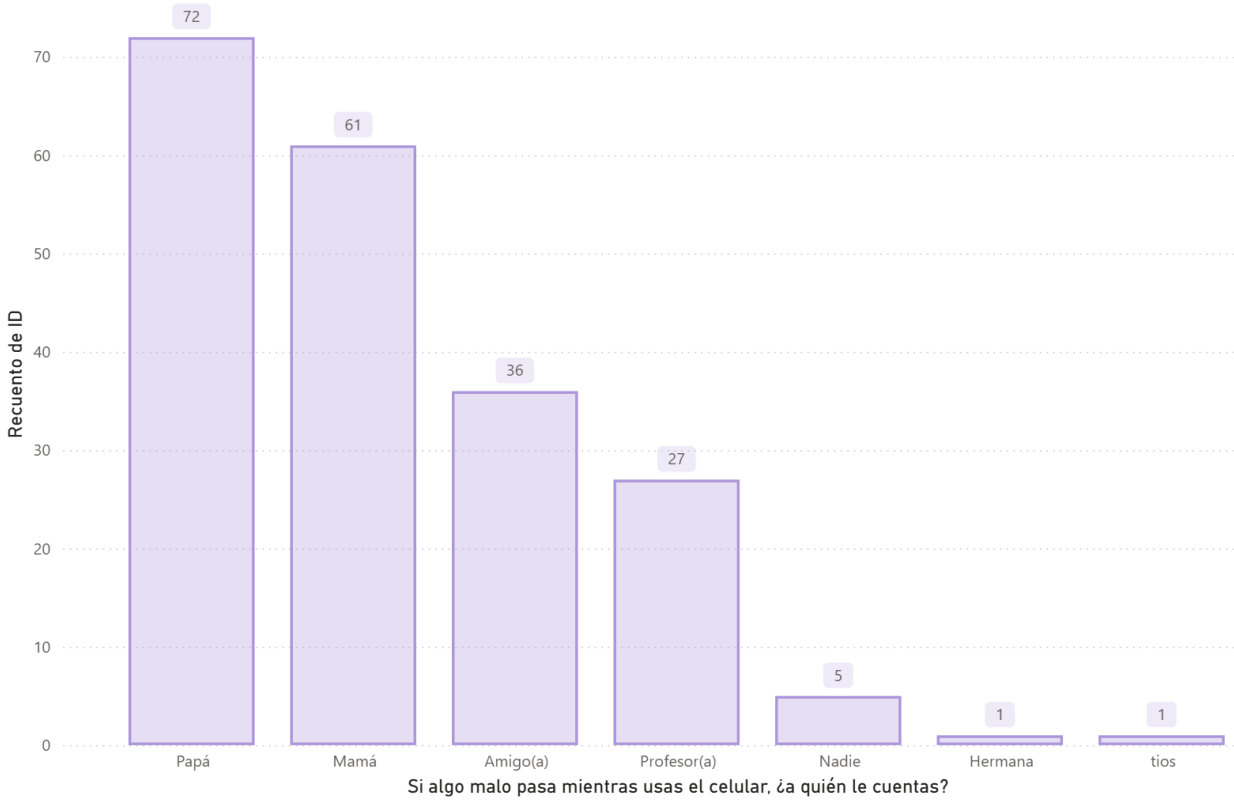


Nota. Al observar el grafico es parecido al “Grafico 8”, que permite identificar las actividades que realizan los menores, a diferencia este cuenta con la edad que tienen los menores y las actividades que realizan al estar en contacto con un dispositivo móvil, se logra observar que la edad no restringe que los menores realicen cualquier tipo de actividad, en el apartado de redes sociales se identifican que (3 personas) entre los 5 - 13 años usan el dispositivo móvil para ingresar a las redes sociales, lo que hace necesario un aplicativo que los pueda proteger ya

que anteriormente observamos que los menores cuentan con poco criterio para identificar amenazas. Elaboración propia.

Figura 10

Confianza cuando algo malo paso al usar el dispositivo móvil



Nota. Al observar el grafico, se analiza que cuando algo malo sucede al usar el dispositivo móvil a quien acude el menor, lo que resultando que la mayor confianza la tienen el papa y la mama como parte del núcleo principal de menor. Por otro lado, se logra identificar que los “Amigos” tienen una participación significativa en la encuesta con 36 votaciones, lo que causa curiosidad ya que dichas personas pueden tener la misma edad que el menor encuestado y resultar en empeorar o no saber manejar el tema, superando a los profesores, y finalmente el resto con una votación mínima. Elaboración propia.

La Tabla 7 muestra las categorías de riesgo digital que fueron identificadas a partir de las respuestas abiertas que ingresaron los menores. Los peligros más escritos describen el

contacto con personas desconocidas, el robo de información, malware y los contenidos para los adultos. Esto permite demostrar que los niños identifican de cierta manera una que otea amenaza digital, sin embargo, no siempre entienden con exactitud sus consecuencias.

Tabla 7

Percepción de riesgos digitales identificados por niños en el uso de dispositivos móviles.

ID	Categoría de riesgo	Descripción de los riesgos mencionados
R1	Contacto con desconocidos	Mensajes de personas extrañas en aplicaciones de redes sociales o videojuegos, que piden datos personales.
R2	Robo de información y hackeo	Robo de las contraseñas, fotos y tarjetas, hackeo de cuentas, suplantaciones de identidad, filtrado de datos personales.
R3	Virus y software malicioso	Instalación de virus, enlaces de procedencia sospechosa, anuncios engañosos, el dispositivo se vuelve lento por malware.
R4	Espionaje y acceso no autorizado	Activación de la cámara si permisos previos, acceso al historial de navegación o las fotos.
R5	Extorción y amenazas	Llamadas donde amenazan a las personas para extorción, temor a que conozcan la dirección del hogar.
R6	Contenido inapropiado	Videos para adultos, páginas que no están hechas para menores.
R7	Ciberacoso	Bullyng en juegos y redes sociales, mensajes ofensivos o que figan cosas feas.
R8	Perdida de cuentas o recursos digitales	Robo de monedas en juegos, perdida de cuentas, eliminación accidental de la información y progreso.
R9	Exposición de información personal	Solicitud de fotos o videos personales, pedir número de teléfono y datos personales o de familiares.
R10	Riesgos emocionales	Miedo, sustos, experiencias feas, ansiedad ante situación en línea.
R11	Falta de control parental	Ausencia de supervisión cuando se usa el dispositivo móvil.
R12	Desinformación	Información falsa en internet o contenidos engañosos

R13	Uso excesivo del dispositivo	Permanecer mucho tiempo jugando y no poder parar.
R14	Riegos de violencia	Miedo a agresiones físicas por la exposición de los datos personales.
R15	Sin percepción del riesgo	Respuestas como “Nada”, “No sé”, “No tengo”.

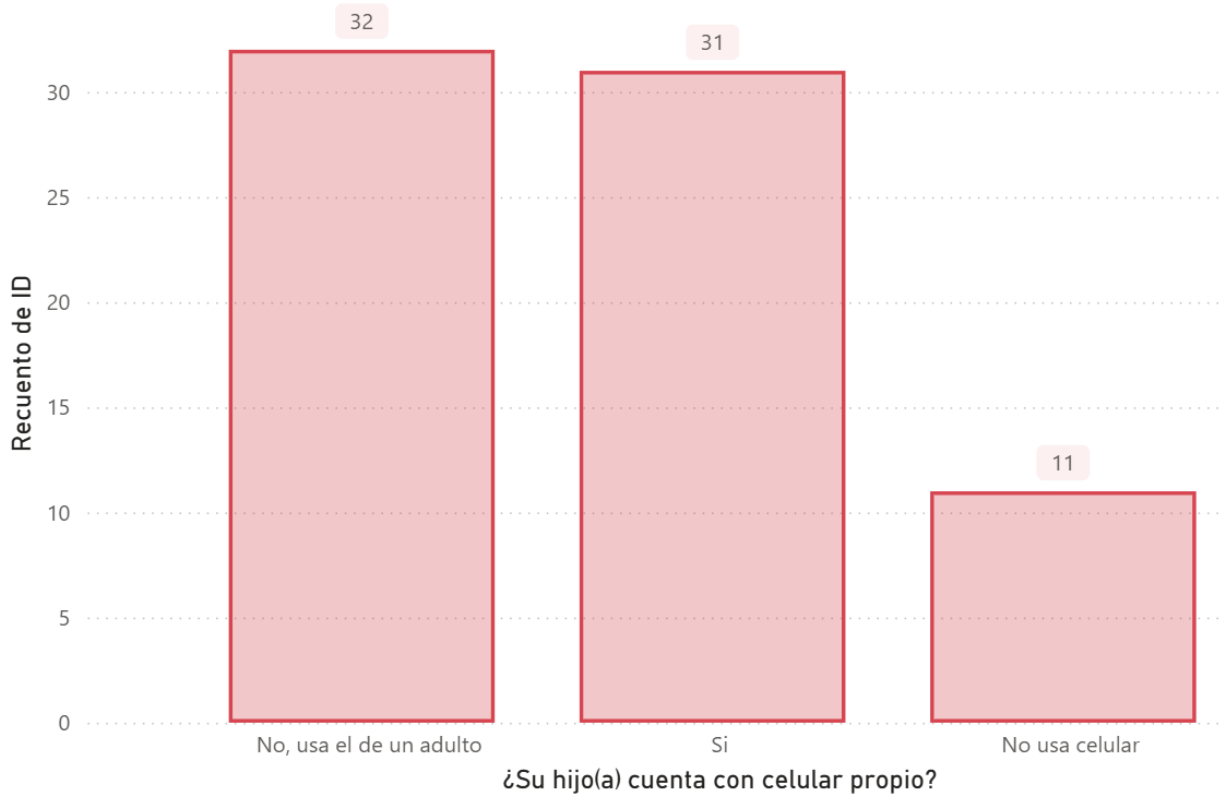
Nota. La tabla permite mostrar respuestas abiertas contestadas en la encuesta hecha por los menores, las categorías se construyeron mediante proceso de codificación cualitativa, donde se agruparon las expresiones similares bajo dimensiones conceptuales comunes para facilitar el análisis e interpretación. Se logra identificar que los menores no solo le tienen miedo a lo que pueda pasar en línea, sino también cuanto tiempo permanecen conectado y no pueden parar. Elaboración propia.

Encuestas Padres de familia

La encuesta a los padres permite conocer la percepción que tienen frente a los riesgos, y la forma en que la pueden mitigar y prevenir, se realizaron preguntas sobre el uso y tiempo que le dan sus hijos a los dispositivos móviles, los resultados evidencian que la mayoría de los menores hacen uso y tiene un teléfono (41,89%) lo usan entre 4 o más horas al día, lo que representa el 22.97% de total del 31.08% de los que se encuentran dentro del rango de uso, esto evidencia que el uso de una herramienta automática mediante el uso del Machine Learning, puede ayudar a que los padres puedan tener un mejor control con lo hacen y con quien hablan.

Figura 11

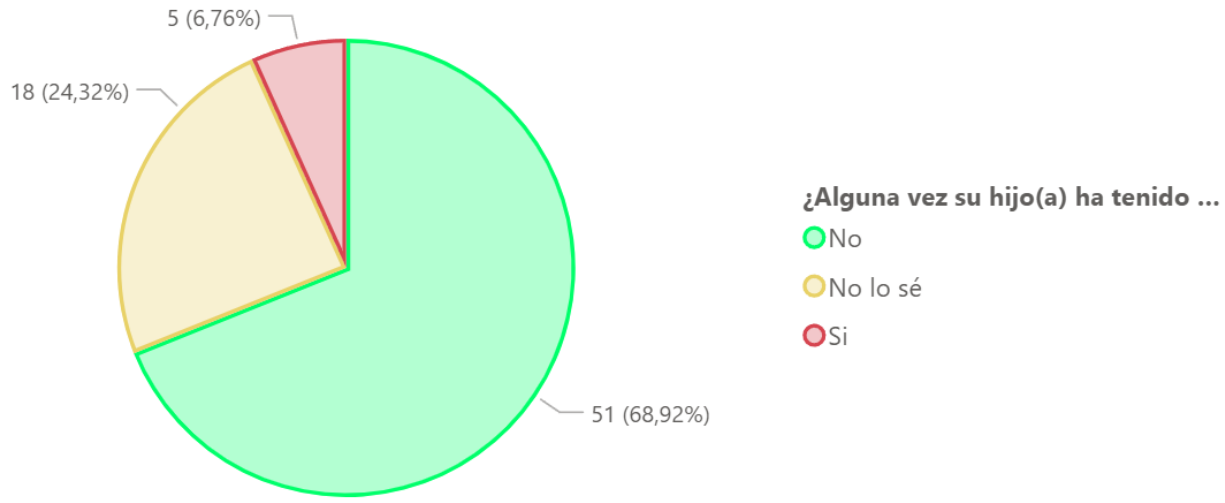
Uso del teléfono según padres de familia



Nota. En el grafico los datos permiten revelar que la mayoría de los padres (32) permiten a sus hijos utilizar dispositivos de adultos, mientras que (31) asegura que su hijo ya tiene uno propio. Esto sugiere que el acceso a la tecnología se comparte mayormente a través de los dispositivos de los familiares. Elaboración propia.

Figura 12

Experiencias previas reportadas por padres sobre sus hijos(as)



Nota. El grafico de pastel permite observar una brecha de información significativa, ya que casi una cuarta parte de los padres (24,3%) desconoce si sus hijos han tenido algún riesgo digital, aunque el (68.9%) reporta que sus hijos no han tenido la experiencia negativa de algún riesgo mientras se encuentran usando el dispositivo móvil, solo (5) padres reportan que sus hijos si han tenido algún tipo de riesgo digital y se muestra en la “Tabla 6”. Elaboración propia.

En la Tabla 8 se describen textualmente los casos de riesgo digital que los padres informaron en la encuesta sobre las experiencias que han tenido sus hijos en ambientes digitales. Se describe que los eventos mencionados abarcan situaciones de ciberacoso, solicitud de contenido explicito e íntimo y comunicación con individuos desconocidos, lo que demuestra que la presencia de peligros reales en el ámbito digital infantil.

Tabla 8

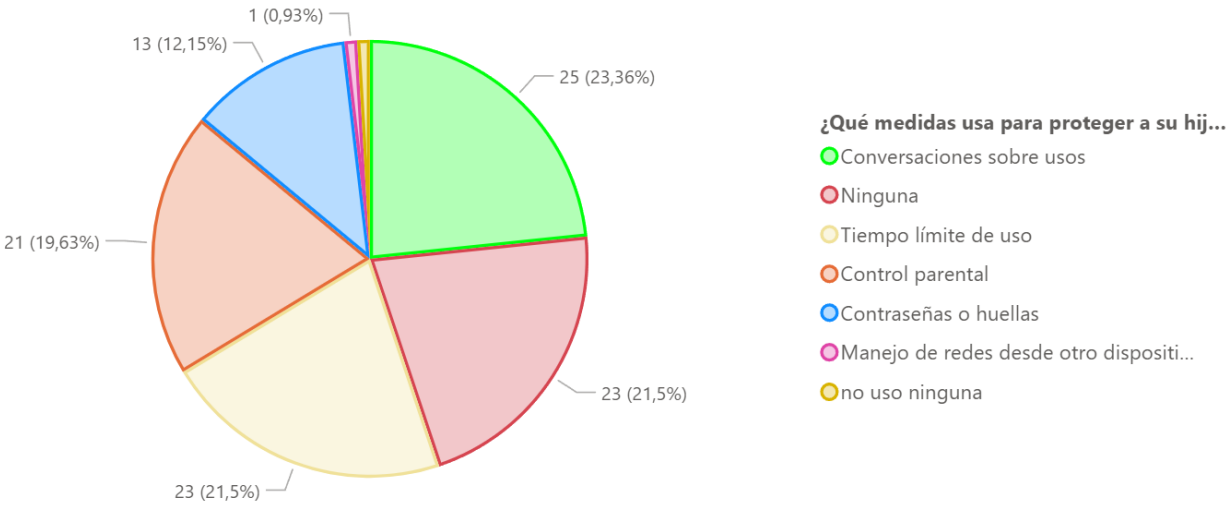
Situaciones de problemas digitales reportadas por padres de familia.

ID	Descripción textual reportada
P1	Una ocasión recibió un mensaje raro de una persona, claro que yo intervine y pude bloquear ese número que era de otro país.
P2	Una persona en un juego lo trató mal y le hacía bullying, diciéndole que lo iba a matar y cosas como esas.
P3	Una persona por Facebook le pidió fotos íntimas y no sabía hasta que ella me lo contó.
P4	Una persona extraña le escribió por TikTok para que se hicieran amigos; no me di cuenta hasta que empezó a pedir fotos y videos. Acudimos a la policía, pero esa persona ya había borrado la cuenta.
P5	Le hackearon el Facebook.

Nota. La tabla permite identificar los testimonios de los padres que evidencias riesgos críticos como el ciberacoso, solicitudes de contenido intimo (grooming) y contacto con extraños en redes sociales y videojuegos. Mediante los testimonios se logra observar que, en varios de los casos, el incidente ocurrió por el contacto con personas desconocidas, lo que fuerza la necesidad de un aplicativo que permita detectar comportamientos anómalos en mensajes. Elaboración propia.

Figura 13

Medidas de seguridad y protección usadas por los padres



Nota. El grafico es claro, la mayor estrategia usada por los padres de familia es el dialogo sobre los riesgos al usar el dispositivo móvil, donde se habla de malware, hablar con personas desconocidas y lo que puede pasar con (23,3%) superando ligeramente a medidas restrictivas como control parental y límite de tiempo de eso en el dispositivo, un (12,1%) usa contraseñas o huella, pero no es suficiente ante riesgos de instalación de malware y contacto con desconocidos ya que las contraseñas no previenen de ninguna manera estos riesgos. Elaboración propia.

La Tabla 9 demuestra que los problemas más importantes que los padres han indicado en relación con las medidas de protección que implementan para el uso de dispositivos móviles por sus hijos. Se nota que las restricciones más comunes se evidencian con la falta de controles parentales, la filtración de contenido inapropiado y inestabilidad mental de los niños, lo que demuestra que se necesitan herramientas complementarias efectivas para mitigar los riesgos a los que los niños están expuestos.

Tabla 9

Fallas y dificultades compartidas por los padres de familia.

ID	Dificultad percibida por los padres
D1	Solo bloquea las apps no el contenido
D2	Su privacidad
D3	Ninguna
D4	rebeldía de mi hijo
D5	Ninguna
D6	Las plataformas para niños aún tienen fugas y se llega a faltar contenido no adecuado para los niños
D7	Ninguna
D8	Que el niño se estresa o se enoja al no darle el teléfono, no tienen el teléfono y se frustra se enoja, contesta fea y demás actitudes grotescas
D9	En las plataformas para niños aún suele haber contenido no apto para niños
D10	Insistencia a más tiempo usando el celular, distraerse mucho
D11	Ninguna dificultad
D12	que no siempre son seguras del todo
D13	nunca eh probado alguna medida de protección

D14	Nunca le eh permitido el teléfono a mi hijo ya que no quiero que este pegado a ese aparato y también me preocupa su seguridad, ya que cualquier persona le puede escribir
D15	nunca eh probado medidas con mi hija, pienso que es bueno darle libertad por si se equivoca pueda aprender de eso que le pueda suceder, no me gusta la vigilancia ni quitarle la privacidad a mi hija
D16	muchas veces los niños son muy inteligentes y desbloquean el control parental con facilidad
D17	que muchas veces no funciona bien ya que no bloquea el teléfono, cuando no esta en la misma red
D18	no me gusta vigilar a mi hijo con aplicaciones ya que confio en el
D19	muchas veces no funcionan al 100% es necesario que alguien pueda intervenir en la seguridad de los niños, ya que los padres no estamos siempre en casa o con ellos
D20	nunca eh usado una de esas herramientas, ya que no le permito el celular a mi niña
D21	no uso ninguna, ya que no conozco medidas de protección
D22	Siempre le presto mi teléfono a mi hijo, pero solo 1 o a veces 2 horas en el dia para que juegue, no puede ingresar a nada más que los juegos, ya que las demás aplicaciones las tengo bloqueadas con faceID para que solo ingrese yo
D23	ninguna, la app funciona muy bien
D24	los niños no escuchan lo que puede pasar cuando no saben usar un teléfono correctamente, pero algunas veces le hablo sobre eso
D25	no
D26	no
D27	ninguna, ya que no le controlo el celular
D28	no conozco
D29	no uso aplicaciones, pero le hablo sobre que no hable con personas que no conozca
D30	no uso ninguna para monitorear lo que pasa en el teléfono de mis hijos
D31	la app solo funciona cuando está en la misma red
D32	no uso ninguna ya que no conozco esas apps y tampoco las sabría manejar
D33	no conozco aplicaciones para ver que hacen
D34	Ninguna
D35	No son muy seguras
D36	No son eficaces
D37	Hasta el momento ninguna
D38	No hay restricciones de los videos
D39	No tengo conocimiento alguno sobre estas aplicaciones para controlar el teléfono de los jóvenes
D40	No controlo lo que hace mi hija en el teléfono
D41	No tiene teléfono, pero si le presto el mío, pero estoy pendiente de que no se meta a otras aplicaciones que no sea el juego
D42	muy mal ya que siempre se enojan porque quieren andar pegados al teléfono
D43	Solo le presto el teléfono cuando debe hacer tareas, ya que no me gusta que ande siempre con el teléfono ya que no confió mucho ya que puede hablar con personas o algo parecido y me da miedo
D44	no es muy eficaz ya que los jóvenes no escuchan ni atienden a las indicaciones que les dan sus padres o profesores en las escuelas
D45	no se manjar esas aplicaciones
D46	no conozco

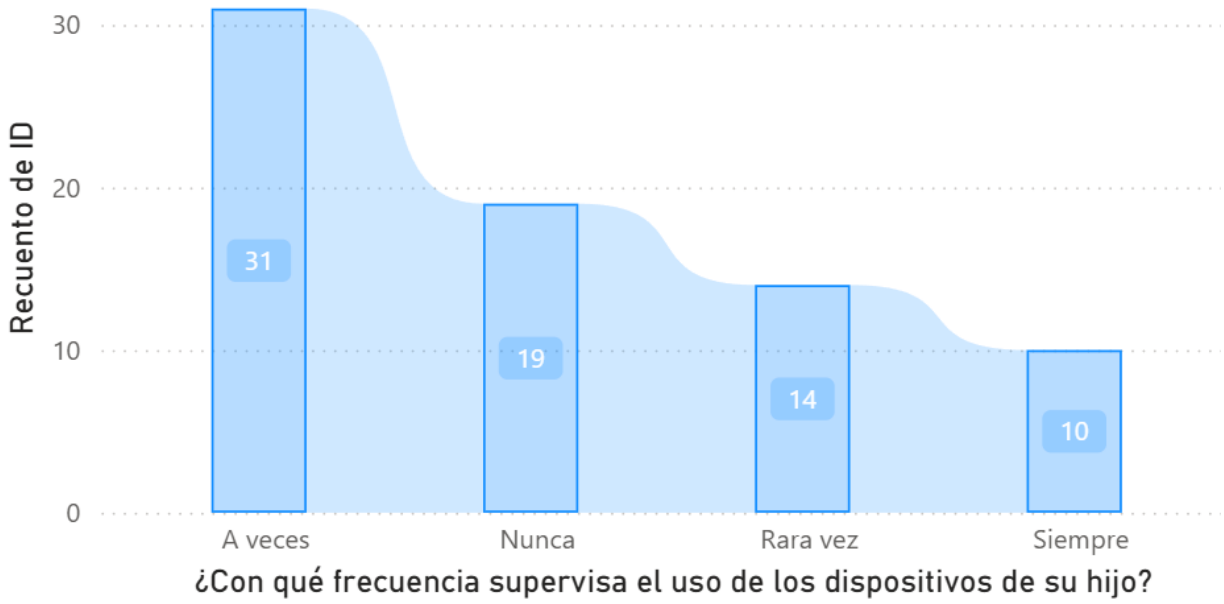
D47	No me gusta controlar lo que hacen mis hijos, ellos ya son bastante grandes para saber lo que hacen y a lo que conllevan sus actos
D48	no tengo algún conocimiento de estas aplicaciones y algunas son de pago por suscripción y en la casa no alcanza para algo así
D49	siempre le converso a mi hija sobre no usar el teléfono tan pequeño, siempre me hace berrinche de que quiere un iPhone, pero está muy niña y no se lo puedo dar ya que puede hacer cosas locas con esos aparatos
D50	siempre gritan o pelean cuando no se les presta el teléfono
D51	no me gusta controlar lo que hace mi hijo con su teléfono, confió en el
D52	muchas veces los niños no escuchan y se enojan porque no se les pasa el teléfono
D53	solo le presto el teléfono para que haga sus tareas del colegio
D54	le hablo sobre cosas que pueden pasar o cosas que vi que les pasaron a otras personas, pero los jóvenes son tercos y no escuchan
D55	siempre estoy a su lado, y verifico que solo vea YouTube
D56	no uso ninguna aplicación ya que no conozco, ni tampoco sé cómo protegerlo o decirle
D57	los jóvenes hacen lo que quieren y no hacen caso a lo que uno les dice, así vean lo que les pasa a otras personas siempre responde con que nunca les pasara a ellos
D58	el mismo compro su teléfono, no puedo tener control sobre el
D59	solo le presto el mío para hacer tareas
D60	siempre me hace pataleta porque le quito el teléfono
D61	Personas extrañas contactando a los chicos
D62	Ninguna porque ya son mayores
D63	No es completo
D64	Jaker
D65	Ninguna
D66	No aplico ninguna
D67	En ocasiones las restricciones no cumplen con solo el contenido para niños, se filtra información que es para adolescentes o mayores.
D68	Ninguna
D69	Que incumpla los horarios establecidos y que este en juegos de guerra porque ayudan a la maldad
D70	Este tipo de medidas no los protegen de otras personas.
D71	Comprensión del riesgo
D72	Ninguna
D73	A veces se saltan los tiempos permitidos.
D74	ninguna

Nota. Los testimonios que contaron los padres identifican cuatro barreras criticas para la mediación parental, en primer lugar limitaciones técnicas que hacen ineficientes los filtros de contenido y la dependencia de la red local, de igual manera el desconocimiento digital, donde muchos de los padres admiten no saber usar las herramientas de control que existen, por otro lado los conflictos conductuales, donde los padres reportan que los menores al no obtener el

teléfono realizan berrinches, rebeldía y actitudes grotescas lo que los hace y evidencia la adicción al teléfono, por ultimo hablan de los dilemas éticos, donde evidencia preocupación por la privacidad del menor como un motivo para no vigilar. Estos datos demuestran la poca alfabetización que tienen los padres con respecto a la seguridad digital infantil. Elaboración propia.

Figura 14

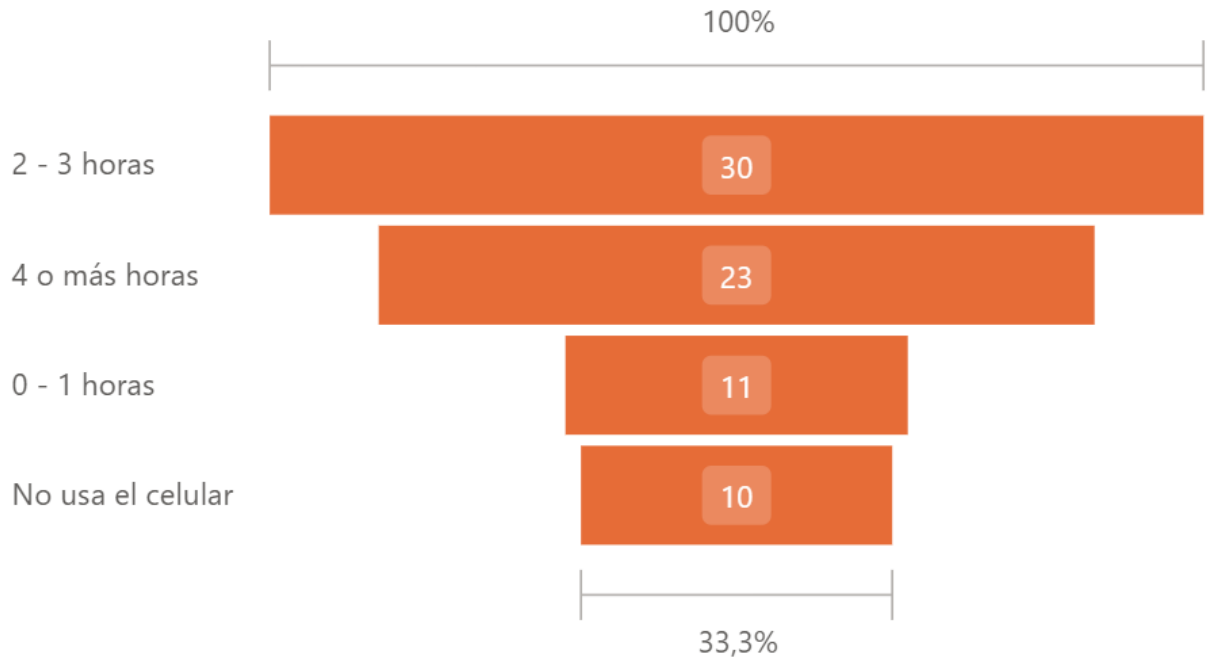
Frecuencia de supervisión parental por parte de los padres



Nota. La supervisión cuando un menor hace uso de un dispositivo es esencial, como se muestra en la gráfica, la mayoría de los padres (31 personas), solo supervisan a sus hijos a veces, llama la atención mucho más que (19 personas) nunca supervisan a sus hijos, esto les hace desconocer todo lo relacionado con los riesgos a los que pueden estar expuestos, duplicando las (10 personas) minoría que siempre supervisa lo que hacen sus hijos en los dispositivos móviles. Elaboración propia.

Figura 15

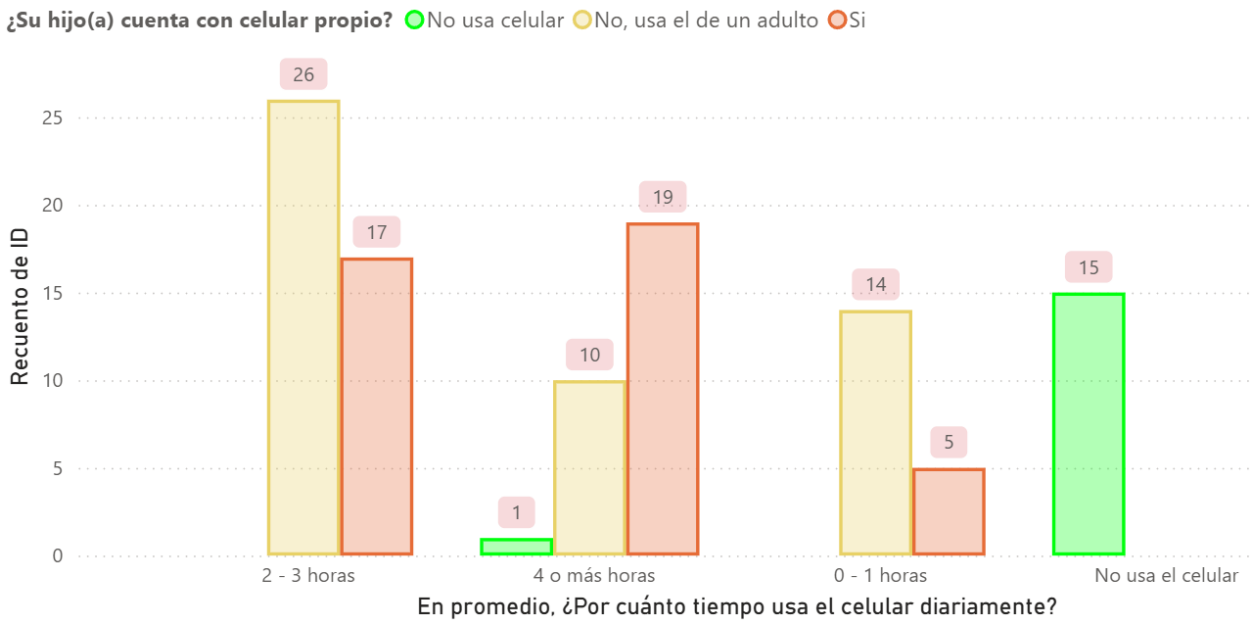
Tiempo de uso del teléfono por parte de los padres a sus hijos(as)



Nota. Como se evidencia en el grafico el habito de consumo se centra en niveles altos, 23 padres le permiten usar a sus hijos un dispositivo móvil de entre 4 o más horas diarias, superando en gran desmedida a los que no lo usan, solo (11 personas) le permiten usar a sus hijos el dispositivo móvil de 0 – 1 hora. Elaboración propia.

Figura 16

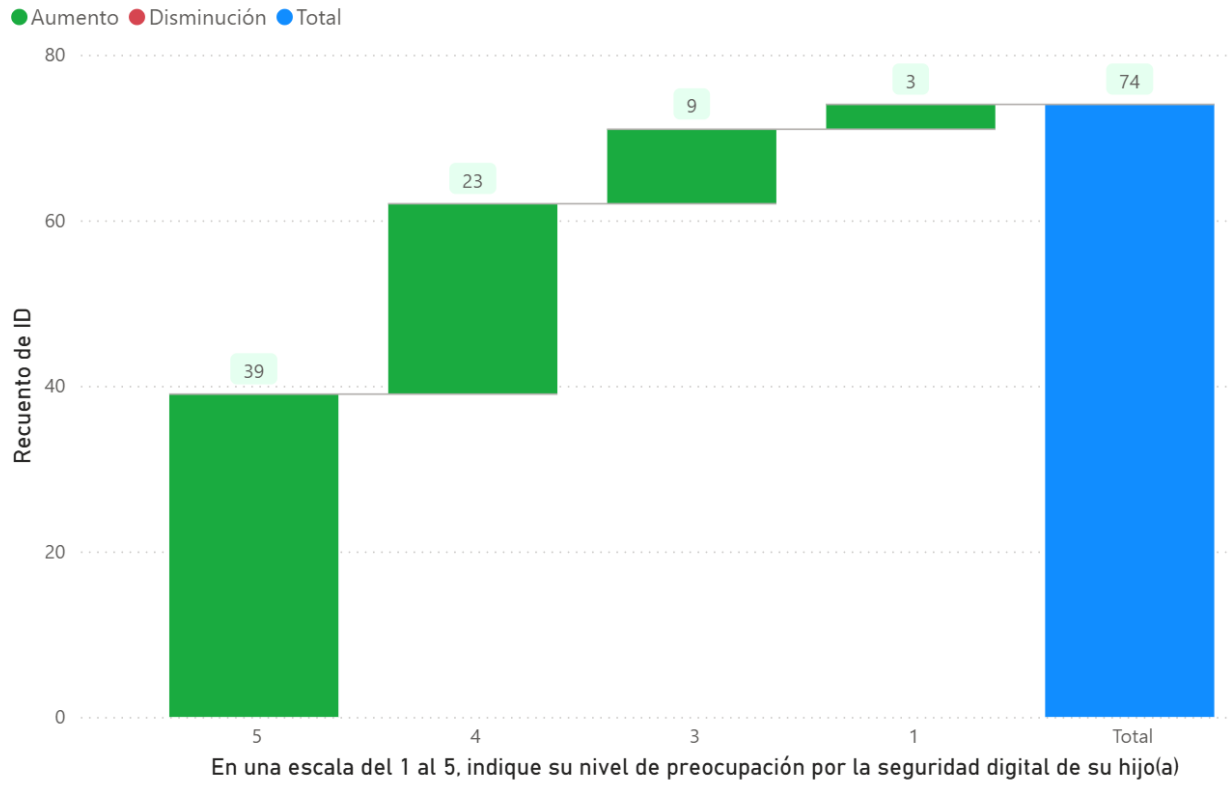
Tiempo de uso dispositivo móvil



Nota. En este grafico se puede observar la comparación entre tiempo en el teléfono y uso del teléfono, lo que se puede observar es que los que mayormente los menores usan el dispositivo de sus padres, detrás de los que usan el de los padres se encuentran los que ya tienen uno propio, y (16 personas) no les permiten el teléfono a sus hijos. Por este motivo, es esencial un aplicativo que permita que los jóvenes que más usan el teléfono puedan estar seguros en todo momento y no tener miedo a los riesgos que se presentan en el día a día en los dispositivos móviles. Elaboración propia.

Figura 17

Nivel de preocupación de los padres respecto a la seguridad digital en los menores



Nota. En el grafico se logra observar que existe un consenso de alerta entre los padres, donde 62 de las 74 personas encuestadas se sitúan en los niveles de preocupación de 4 a 5. Esto permite reflejar que una gran parte de los padres tiene una percepción alta respecto al entorno digital en el que interactúan sus hijos. Elaboración propia.

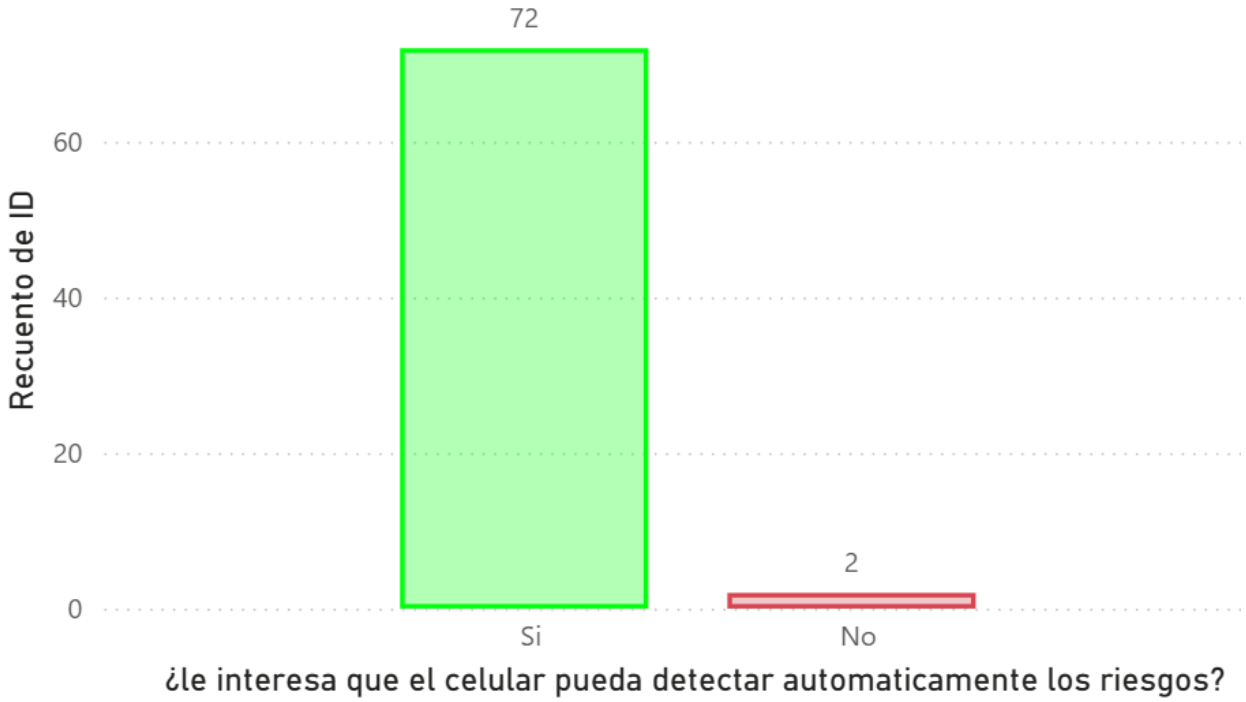
Figura 18

Actividades principales realizadas por los menores en los dispositivos móviles

Nota. El grafico se observa que los padres identifican que sus hijos mayormente se la pasan en juegos, videos y redes sociales, (31 personas) realizan tareas escolares ocupando el tercer lugar por debajo del entretenimiento puro, inclusive de las redes sociales, solo una gran minoría no lo usa o tal vez configuran sus teléfonos sus padres. Elaboración propia.

Figura 19

Interés parenteral para la implementación de sistemas de detección automática en dispositivos móviles



Nota. Se logra observar que a la mayoría de los padres (72 personas) les interesa que en los dispositivos se integren soluciones tecnológicas que automaticen la detección de riesgos para poder proteger de mejor manera a sus hijos, solo (2 personas) indican que no les interesa dicha solución, posiblemente porque es un aplicativo que debe estar en segundo plano y registrando todas las actividades que realiza su hijo y de cierta manera opacar la privacidad o confían en sistemas manuales, aunque no es obstáculo para los padres que confían en la tecnología como un aliado para cubrir la brechas de supervisión parental. Elaboración propia.

Resultados Encuestas a Docentes

1. **Riesgos y conciencia digital**

Los 3 docentes coinciden en que uno de los mayores peligros al que los menores están expuestos es el contacto con desconocidos, acceso a contenido para adultos y la manipulación por parte de terceros en video juegos y redes en línea. Se preocupan porque en Colombia falta conocimiento y que los adultos no están informados 100& como creen; incluso estos añaden que existen padres que exponen a sus hijos a un dispositivo móvil desde edades muy tempranas.

2. **Educación y acompañamiento**

Existe un debate entre estos sobre quien es el responsable principal de evitar que los menores estén expuestos a riesgos, unos afirman que los profesores deben de ser capacitados en ciberseguridad para orientar a los alumnos y que las instituciones de educación se involucren más con la protección digital de sus estudiantes, otros sostienen que el responsable es el padre de familia como núcleo principal y ser quienes dan acceso a los menores al uso del dispositivo móvil.

3. **Machine Learning**

La mayoría ve con muy buenos ojos el uso de inteligencia artificial, sin embargo, otros creen que su uso genera dependencia de esta tecnología y puede frenar el desarrollo si les va a facilitar la vida. Creen que para aprender primero debes haberte caído, sin embargo, el uso de inteligencia artificial los preocupa por la privacidad de sus hijos, los consentimientos, los falsos positivos y el riesgo de sobreproteger de más.

4. **Viabilidad del modelo propuesto**

Los 3 docentes consideran que es viable un modelo que genere alertas automáticas si sospecha que el menore está en peligro, consideran que debería realzarse para sistemas operativos de computadores y así poder instalarlos en los sistemas institucionales

permanentemente para dar protección a los niños y realizar reuniones informativas con los padres para dar a conocer su funcionamiento. Antes de instalarlo insisten en que se debe demostrar como previene los riesgos en redes sociales para ganar la confianza de los padres y los menores en los que se encuentran niños y adolescentes.

5. **Visión futura**

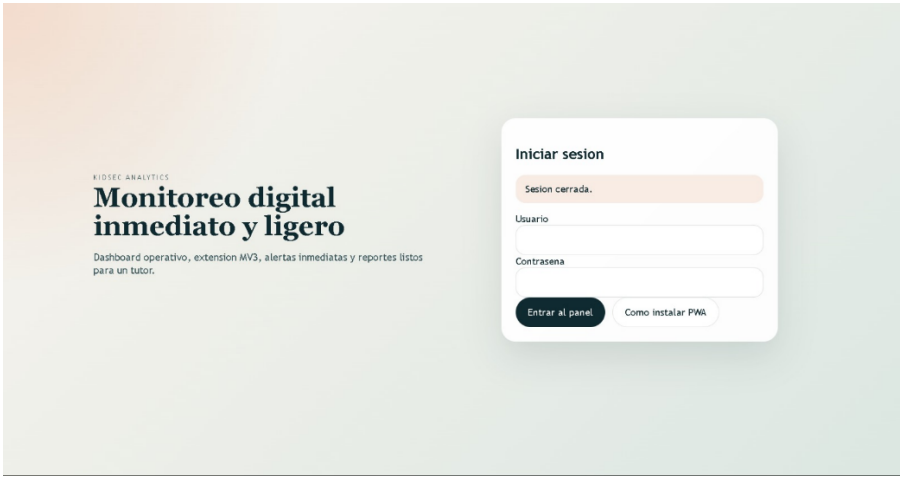
Los 3 docentes proponen estrategias lúdicas para informar a los menores sobre los riesgos a los que se exponen sin confundirlos, proponen juegos educativos, dispositivos físicos como pulseras inteligentes de alerta, y aparte de esto se preocupan de que las redes sociales prioricen el negocio sobre la seguridad y que Colombia pueda implementar políticas públicas para la protección digital e IA.

Resultados Construcción PWA

Dando solución a los objetivos planteados y a la problemática general, se desarrolló un aplicativo tipo PWA basado en Machine Learning y con una extensión MV3. Con un panel de inicio de sesión, un centro de monitoreo con generación de dashboard en tiempo real y un apartado para la instalación del PWA y cómo instalarlo junto con su extensión MV3 validada por Chrome.

Figura 20

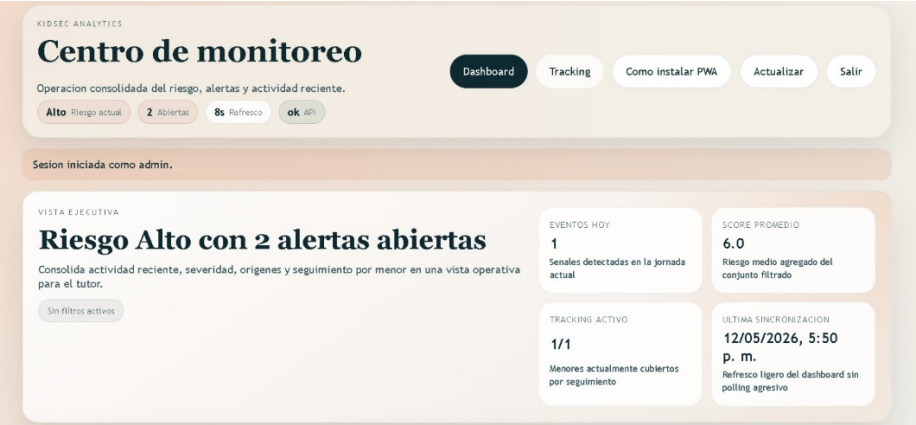
Panel Inicio de Sesión



Nota. Demostración práctica del inicio de sesión, donde el padre accede a la PWA. Elaboración propia.

Figura 21

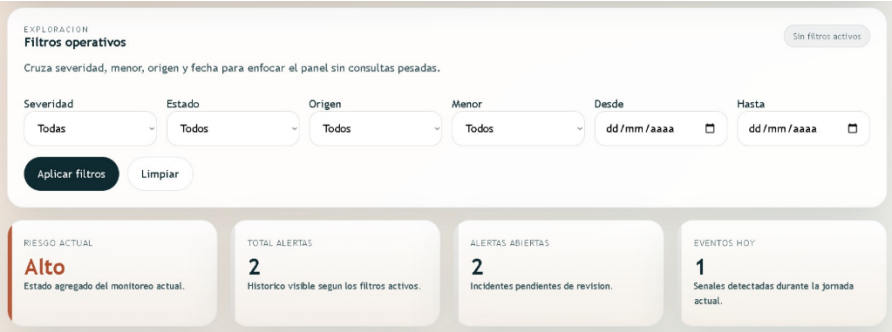
Centro de Monitoreo



Nota. Primer apartado con el que el padre interactúa dentro de la PWA. Elaboración propia.

Figura 22

Filtro de Búsqueda



Nota. Filtros operativos, donde el padre busque según la clasificación de riesgo dentro del PWA. Elaboración propia.

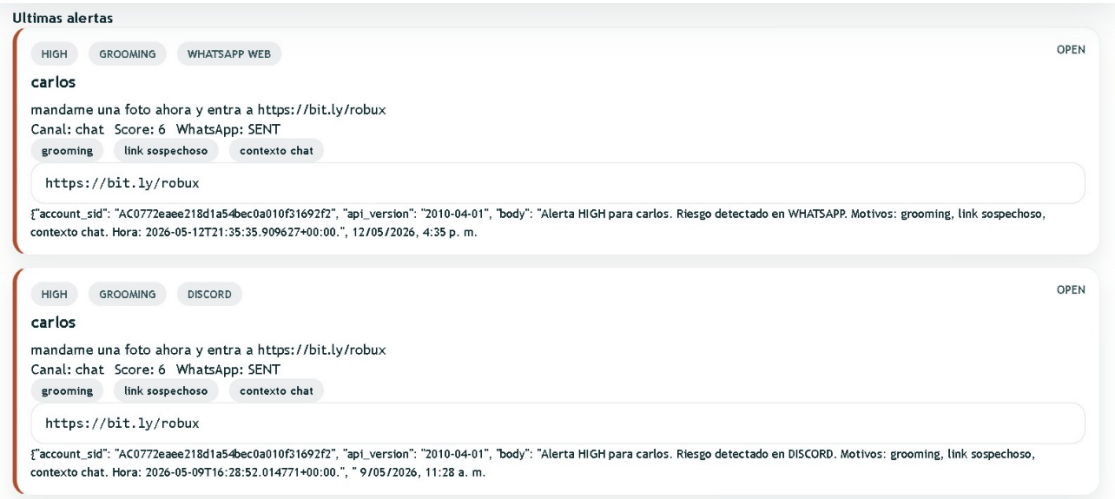
Figura 23

Dashboard General

Nota. Dashboard en los cuales el padre encuentra el riesgo actual, alertas y actividad reciente para monitorear las actividades dentro de cada aplicación. Elaboración propia.

Figura 24

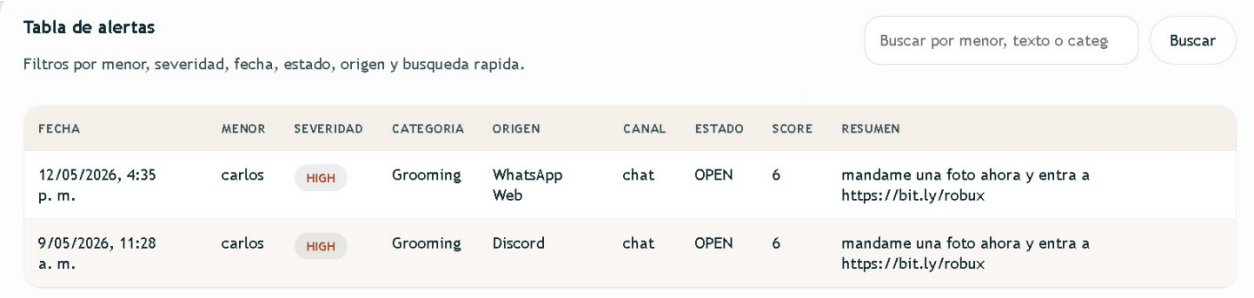
Alertas Recientes



Nota. El padre observa la gravedad de la alerta, el tipo de vulnerabilidad y la aplicación en la que se genera la alerta. Elaboración propia.

Figura 25

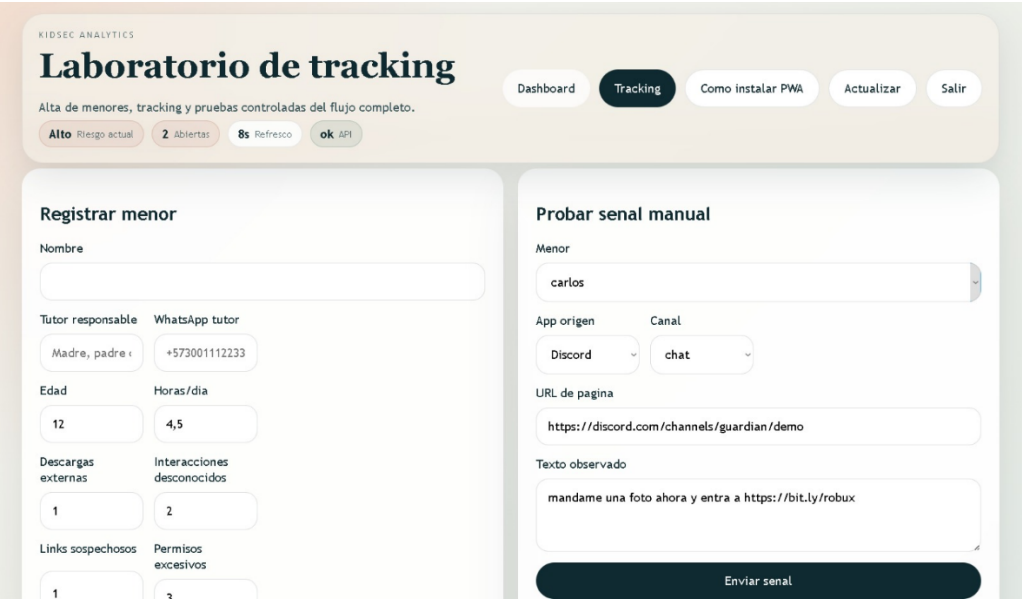
Especificación de Alerta



Nota. El padre puede observar la fecha, severidad, fecha, estado, origen y búsqueda rápida. Elaboración propia.

Figura 26

Laboratorio de Tracking



Nota. El padre registra la información necesaria del menor para el correcto monitoreo en tiempo real. Elaboración propia.

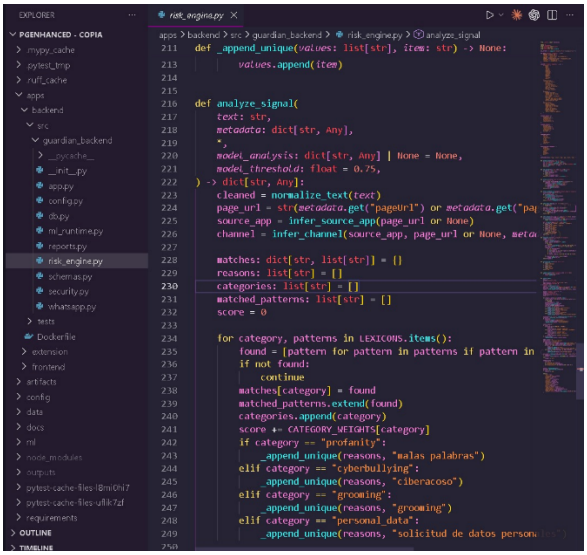
Figura 27

Estado Tracking y configuración de extensión

Nota. El padre configura la extensión encargada de generar la alerta en WhatsApp. Elaboración propia.

Figura 28

Variables de comportamiento digital

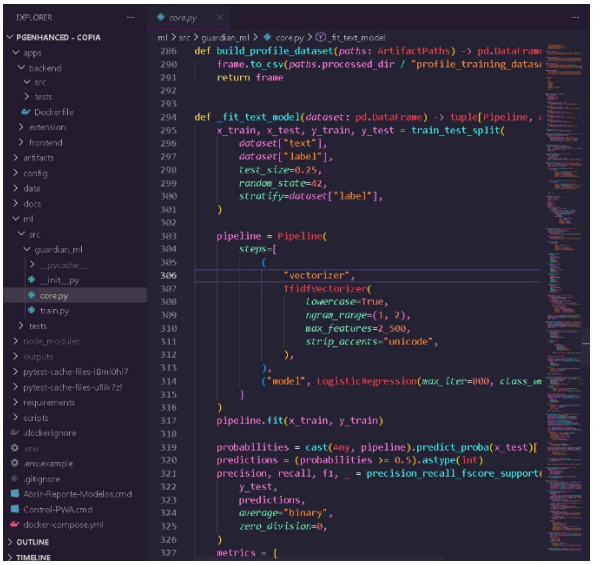


Nota. Para construir el perfil de riesgo de cada menor, definimos 19 variables que capturan distintas dimensiones de su vida digital. Algunas tienen que ver con el uso directo del

dispositivo: cuántas horas pasa conectado al día, si se conecta de noche cuando hay menos supervisión. Otras reflejan la exposición a situaciones potencialmente peligrosas: si descarga archivos de fuentes desconocidas, si interactúa con extraños o si accede a enlaces que no provienen de fuentes confiables. El tercer grupo surge de las encuestas que aplicamos directamente a niños y padres en Colombia, y recoge aspectos psicosociales como la impulsividad del menor, su nivel de autoestima y qué tanto entiende realmente sobre el entorno digital en el que se mueve. Juntas, estas 19 variables forman el vector que entra al modelo de perfil de riesgo. Elaboración propia.

Figura 29

Clasificador de texto (señales de riesgo en conversaciones)

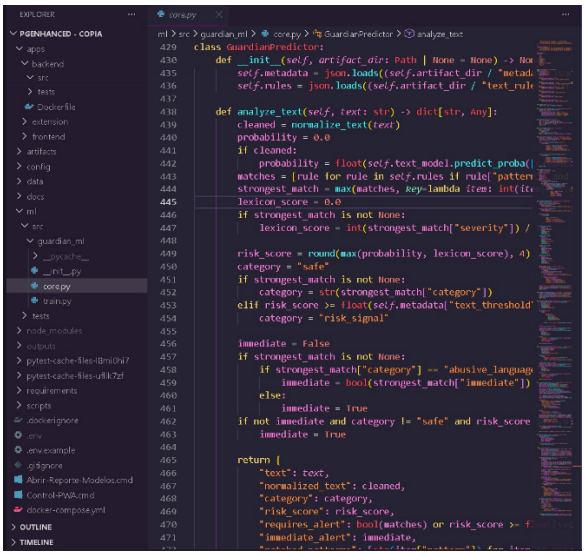


Nota. El primer modelo de machine learning que implementamos tiene una tarea concreta: leer un fragmento de texto y determinar si contiene señales de riesgo. Para representar el texto numéricamente usamos TF-IDF con bigramas, lo que nos permite capturar frases compuestas como "mándame una foto" en lugar de analizar solo palabras sueltas. El clasificador es una Regresión Logística con class_weight="balanced", una decisión que tomamos porque los mensajes problemáticos son mucho menos frecuentes que los seguros, y sin este ajuste el

modelo simplemente los ignoraría. Los datos de entrenamiento combinan las respuestas de 113 niños y 74 padres colombianos con más de 1.800 términos de abuso y 400 señales de riesgo digital que cubramos manualmente. Elaboración propia.

Figura 30

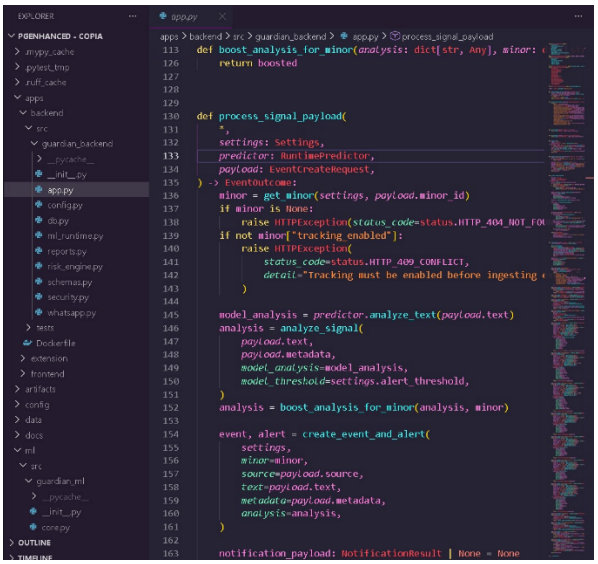
Clasificador de perfil conductual



Nota. Este segundo modelo trabaja sobre perfiles completos, no sobre textos. Lo entrenamos con 3.000 perfiles del dataset colombiano unificado. El pipeline maneja primero los valores faltantes (algo inevitable en encuestas reales) imputando por mediana, luego normaliza con StandardScaler y finalmente aplica Regresión Logística para asignar una etiqueta: riesgo bajo, medio o alto. Esa etiqueta acompaña al perfil del menor de forma permanente y le indica al sistema cuánto peso darle a cada señal que detecte más adelante. Elaboración propia.

Figura 31

Fusión entre ML y reglas léxicas



Nota. El riesgo final no lo decide un solo modelo. Optamos por una estrategia de fusión: tomamos el máximo entre la probabilidad que arroja el clasificador estadístico y la puntuación del patrón léxico más grave encontrado. La lógica detrás de esto es simple: cuando el modelo ML tiene dudas, los lexicones actúan como red de seguridad. Si alguien escribe una frase de grooming documentada o pide datos personales de forma explícita, ese patrón se detecta aunque el modelo probabilístico no esté seguro. Elaboración propia.

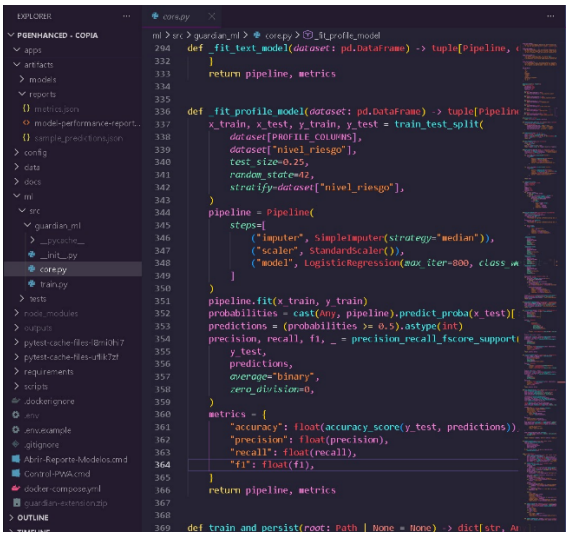
Figura 32

Motor de análisis de riesgo (backend)

Nota. Este es el núcleo de todo el sistema de detección. Analiza cada texto en cuatro capas sucesivas. Primero, busca coincidencias contra cuatro categorías de amenaza con pesos distintos: profanidad (peso 1), ciberacoso (peso 2), grooming y solicitud de datos personales (peso 3 cada uno). Segundo, detecta URLs sospechosas: acortadores, direcciones IP directas, dominios con extensiones de alto riesgo como .zip o .cam, o fragmentos típicos como "free-robux". Tercero, aplica un boost contextual según la plataforma de origen, sumando un punto adicional si el mensaje proviene de Discord, WhatsApp, Roblox o TikTok. Cuarto, integra el score del modelo ML cuando supera el umbral configurado. El resultado final es una puntuación de 0 a 8 que se traduce en cuatro niveles de severidad: LOW, MEDIUM, HIGH y CRITICAL. Elaboración propia.

Figura 33

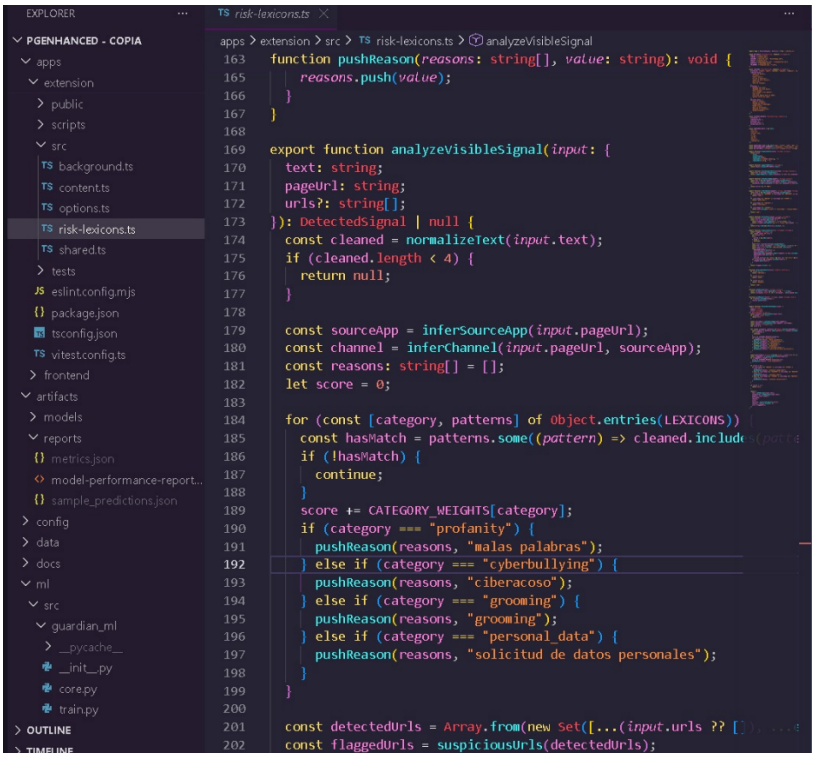
Flujo completo de procesamiento



Nota. Cada vez que la extensión captura un texto sospechoso, el backend sigue una secuencia definida: verifica que el menor tenga seguimiento activo, aplica el análisis ML, ejecuta el motor heurístico y, si el perfil del menor tiene etiqueta de riesgo alto, refuerza la severidad del resultado. Si se genera una alerta, el sistema persiste el evento y envía una notificación de WhatsApp al acudiente a través de la API de Twilio. Es el punto donde el aprendizaje automático, las reglas expertas y la notificación en tiempo real se unen en un solo flujo. Elaboración propia.

Figura 34

Detección en el dispositivo (extensión Chrome)



Nota. La última capa de detección ocurre directamente en el navegador del menor, sin depender del historial del dispositivo ni de acceso al servidor. La extensión, construida con Manifest v3, observa los cambios en el DOM de plataformas como Discord, Roblox, TikTok, WhatsApp Web y Telegram Web. Normaliza el texto visible eliminando tildes, espacios repetidos y mayúsculas, y aplica los mismos lexicones y pesos del backend para mantener consistencia. Cuando detecta algo, envía el fragmento al servidor para el análisis completo con ML y la generación de alerta si corresponde. Elaboración propia.

Análisis de Datos y Evaluación del Modelo Predictivo

Se llevo a cabo un estudio estadístico descriptivo y de correlación sobre variables en entornos digitale que se analizaron con el objetivo de entender los elementos que afectan la vulnerabilidad digital en los menores y poder crear un sistema de alerta temprana. Después, se entrenaron y analizaron 4 modelos de aprendizaje automatico “Machine Learning”, para prever el nivel de riesgo de los usuarios de forma automática.

A continuación, se presentan los resultados del diagnóstico de datos y el rendimiento comparativo de cada uno de los algoritmos implementados.

Después de la limpieza, el dataset queda con 3000 registros, 20 variables, sin nulos y sin duplicados exactos. Además, las variables binarias y categóricas quedaron tipadas de forma más coherente para análisis y modelado.

El dataset ya tiene un nivel de integridad suficiente para continuar. El único problema residual es que algunas columnas originalmente tuvieron una proporción alta de faltantes, por lo que cualquier conclusión sobre esas variables debe leerse con cautela. Sin embargo, la estrategia aplicada permite conservar la muestra completa y sostener un análisis reproducible.

Figura 35

Análisis descriptivo del conjunto de datos de estudio

	count	mean	std	min	25%	50%	75%	max
edad	3,000.00	12.45	2.91	8.00	10.00	12.00	15.00	17.00
horas_uso_diario	3,000.00	6.24	3.27	0.50	3.40	6.33	9.06	12.00
uso_nocturno	3,000.00	0.40	0.49	0.00	0.00	0.00	1.00	1.00
descargas_fuentes_externas	3,000.00	0.29	0.45	0.00	0.00	0.00	1.00	1.00
numero_interacciones_desconocidos	3,000.00	6.27	2.44	0.00	5.00	6.00	7.00	21.00
enlaces_sospechosos_detectados	3,000.00	5.74	4.83	0.00	3.00	4.00	6.00	20.00
permisos_excesivos_apps	3,000.00	5.49	3.28	0.00	4.00	5.00	6.00	15.00
tipo_app_principal	3,000.00	1.03	0.62	0.00	1.00	1.00	1.00	2.00
control_parental_activado	3,000.00	0.69	0.46	0.00	0.00	1.00	1.00	1.00
nivel_riesgo	3,000.00	0.43	0.49	0.00	0.00	0.00	1.00	1.00
sexo	3,000.00	0.48	0.50	0.00	0.00	0.00	1.00	1.00
interacciones_desconocidos	3,000.00	7.93	2.91	1.00	6.00	8.00	10.00	21.00
enlaces_sospechosos	3,000.00	3.01	1.69	0.00	2.00	3.00	4.00	12.00
control_parental	3,000.00	0.61	0.49	0.00	0.00	1.00	1.00	1.00
impulsividad	3,000.00	5.11	1.85	0.00	3.88	5.06	6.35	10.00
autoestima	3,000.00	5.88	1.96	0.08	4.43	5.94	7.27	10.00
necesidad_aprobacion_social	3,000.00	5.00	1.92	0.00	3.66	4.97	6.33	10.00
tolerancia_frustracion	3,000.00	5.82	1.89	0.61	4.55	5.81	7.08	10.00
supervision_parental_percibida	3,000.00	6.12	1.33	0.00	6.11	6.11	6.11	10.00
alfabetizacion_digital	3,000.00	6.00	1.29	0.00	6.01	6.01	6.01	10.00

Nota. En la siguiente tabla se presenta un resumen estadístico de las 3.000 observaciones que se registraron, esta permite observar la distribución, medidas y desviaciones estándar de las variables como lo son la edad, con una medida de 12.4 años), factores de comportamiento como las horas de uso que se le da al dispositivo y las interacciones con desconocidos, de igual manera como los rasgos psicométricos como lo son el autoestima y la impulsividad, las cuales sirven de base para el modelo predictivo propuesto. Elaboración propia.

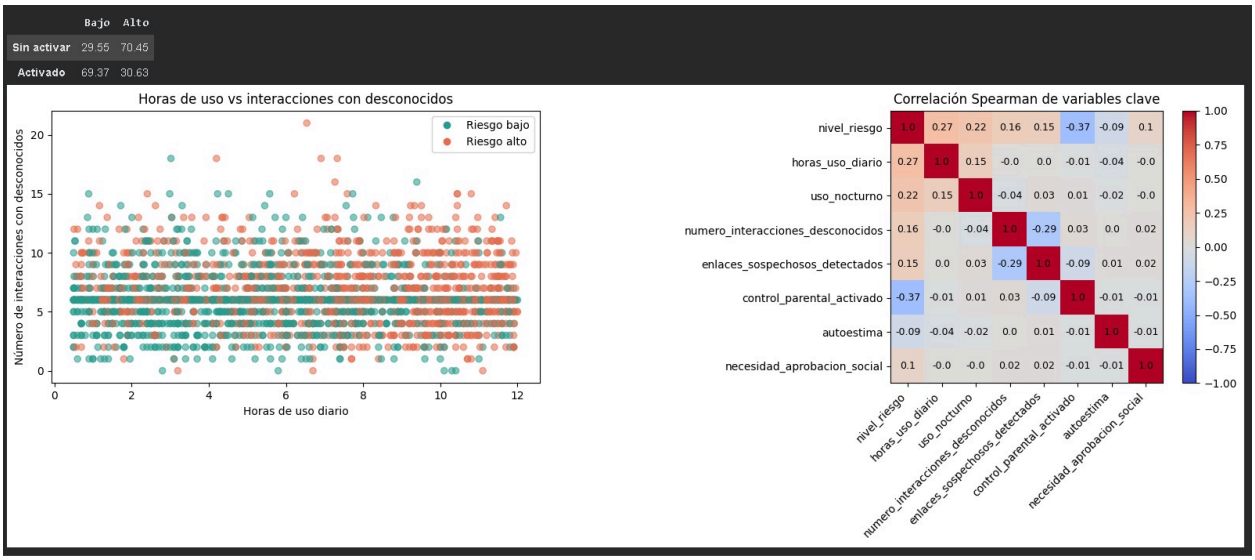
El análisis multivariado confirma que el riesgo digital surge por la interacción de varias dimensiones. En la matriz de Spearman del dataset limpio, las asociaciones positivas más visibles con nivel_riesgo corresponden a horas_uso_diario (0.266), uso_nocturno (0.225), numero_interacciones_desconocidos (0.157) y enlaces_sospechosos_detectados (0.150). En cambio, la relación negativa más fuerte es control_parental_activado (-0.371), lo que sugiere un efecto protector importante.

Este patrón también se observa en las proporciones: cuando el control parental no está activado, el porcentaje de riesgo alto llega a 70.45%; cuando sí está activado, baja a 30.63%. El gráfico de dispersión muestra además que los casos de mayor riesgo tienden a concentrarse en zonas de alta exposición temporal y mayor interacción con desconocidos, es decir, en un espacio donde se combinan conductas intensivas y menor protección contextual.

Se opto por usar scatterplot para observar una interacción conductual entre dos variables continuas y una matriz de correlación para sintetizar múltiples relaciones simultáneas de forma transparente.

Figura 36

Correlación de variables y distribución del riesgo digital



Nota. En la Figura se presenta un análisis visual compuesto por una relación entre las horas de uso diarias del dispositivo y las interacciones con usuarios desconocidos, junto con una matriz de correlación Spearman. Se destaca de manera inversa la relación entre el control parental activo y el nivel de riesgo (-0.37), lo que demuestra estadísticamente el impacto de la supervisión parental en entornos digitales. Elaboración propia.

En este problema las métricas más importantes no son solo accuracy, sino especialmente recall, precision, F1 y ROC-AUC. La razón es que un falso negativo implica no detectar a tiempo un caso de riesgo alto, lo cual es más sensible que equivocarse en algunos falsos positivos. Por eso el recall es clave; la precision ayuda a no saturar el sistema con alertas innecesarias; el F1 equilibra ambas; y el ROC-AUC evalúa la capacidad general de discriminación probabilística.

Los resultados muestran que SVM RBF alcanza el mejor equilibrio global con Accuracy = 0.8717, Recall = 0.8833, F1 = 0.8550 y ROC-AUC = 0.9464. La Red Neuronal (MLP) empata en accuracy y obtiene un ROC-AUC ligeramente superior (0.9476), pero queda por debajo en F1 (0.8499) y deja más falsos negativos. Random Forest y Regresión Logística ofrecen desempeños aceptables, pero inferiores en capacidad de detección del riesgo alto.

En términos operativos, el resultado favorece modelos que reduzcan falsos negativos, porque el sistema debe servir como herramienta preventiva. Por eso no basta con mirar la exactitud total.

Figura 37

Evaluación comparativa de modelos de clasificación

	Modelo	Accuracy	Precision	Recall	F1	ROC_AUC	TN	FP	FN	TP
0	SVM RBF	0.87	0.83	0.88	0.85	0.95	296	47	30	227
1	Red Neuronal (MLP)	0.87	0.85	0.85	0.85	0.95	305	38	39	218
2	Random Forest	0.84	0.82	0.81	0.81	0.93	297	46	49	208
3	Regresión Logística	0.83	0.79	0.83	0.81	0.91	285	58	43	214

Nota. Con el finde seleccionar el algoritmo óptimo para la detección de riesgos de forma temprana en la figura se puede observar que se evaluaron cuatro arquitecturas distintas donde se usaron métricas estándar de clasificación y junto su matriz de confusión (verdaderos positivos, falsos positivos, etc.). El modelo que ajusto a las necesidades del problema fue el primero “SMV RBF” aportando a la herramienta a detectar de mejor manera verdaderos falsos positivos. Elaboración propia.

Figura 38

Reporte de clasificación detallado modelo optimo (SMV RBF)

```
Mejor modelo según F1: SVM RBF
              precision    recall  f1-score   support

           0     0.9080    0.8630    0.8849       343
           1     0.8285    0.8833    0.8550       257

    accuracy                         0.8717       600
   macro avg     0.8682    0.8731    0.8699       600
weighted avg     0.8739    0.8717    0.8721       600
```

Nota. En la figura se observa el mejor modelo según la metrica F1-Score con un (0.85), aquí se desglosa el reporte detallado para el algoritmo SMV con kernel de función base radial (RBF).

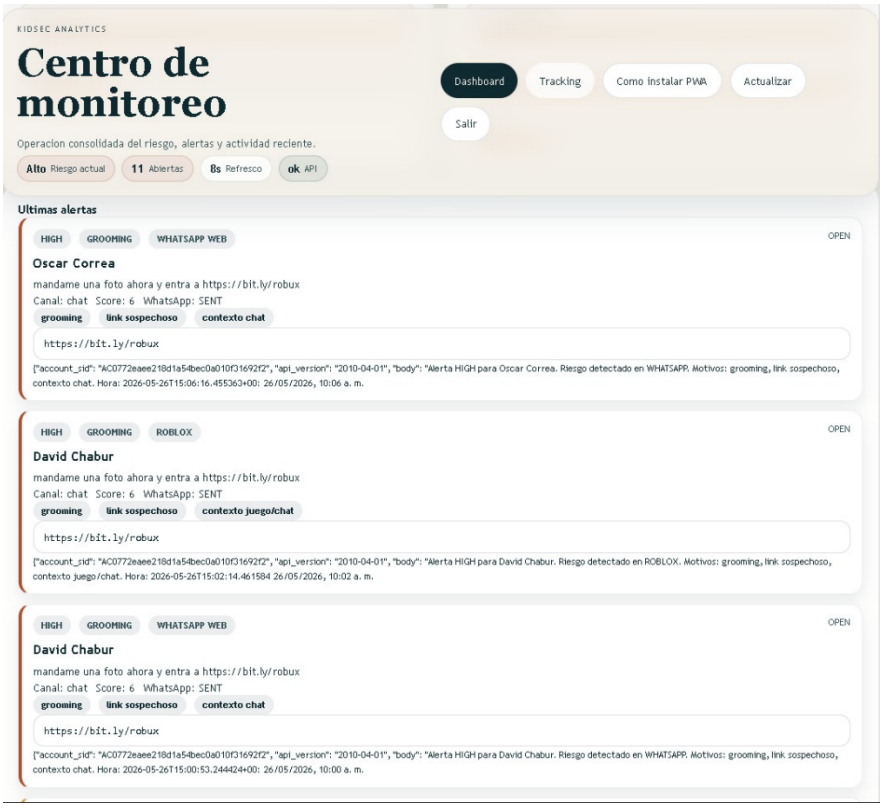
Este modelo demostró tener un excelente equilibrio en la detección de situaciones seguras.

Elaboración propia.

Resultados aplicación del PWA

Figura 39

Resultado 1

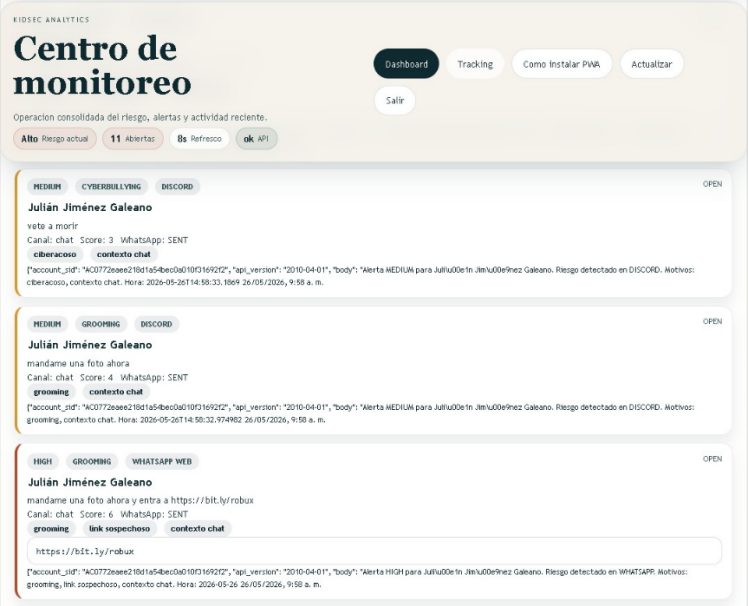


Nota. Se logra observar los resultados prácticos de 2 menores en un ambiente controlado.

Elaboración propia.

Figura 40

Resultado 2

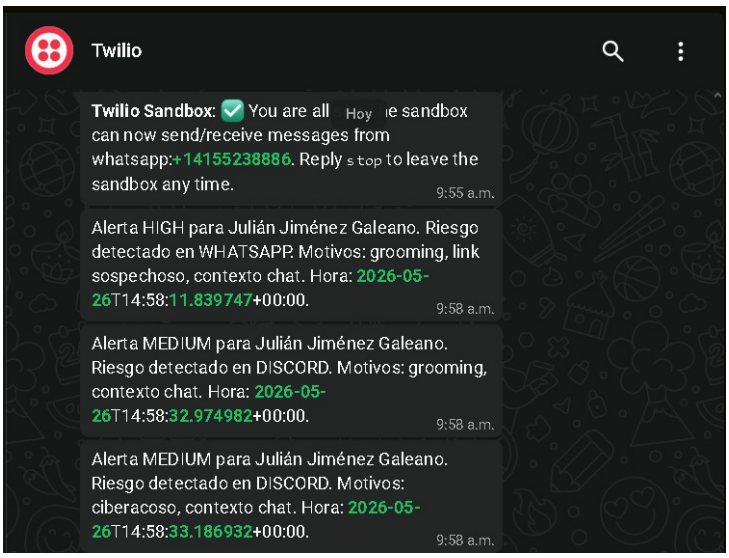


Nota. Se logra observar los resultados prácticos de 1 menor en un ambiente controlado.

Elaboración propia.

Figura 41

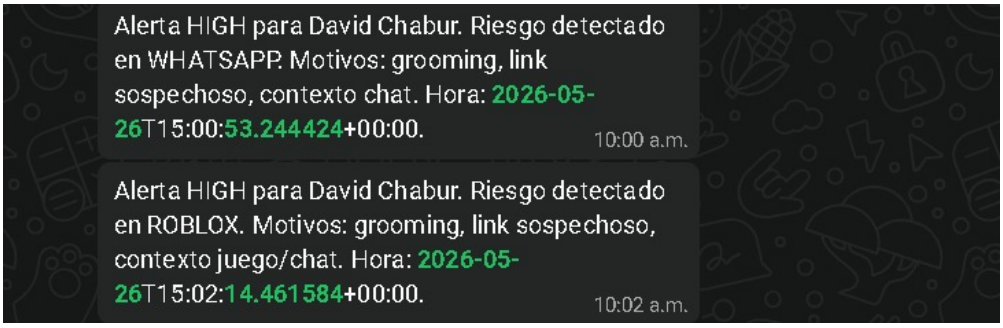
Alerta generada Twilio 1



Nota. Se observan los mensajes de alerta en tiempo real con N8N. Elaboración propia.

Figura 42

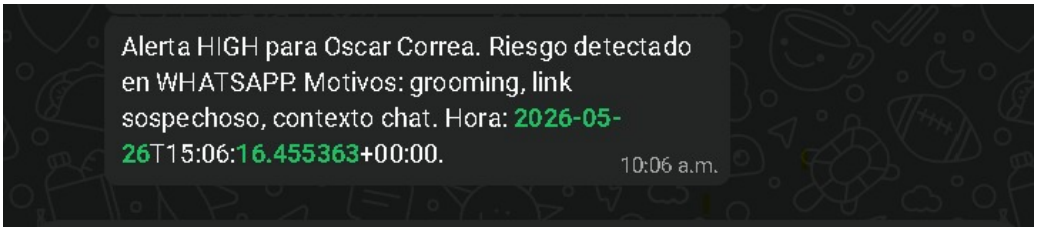
Alerta generada Twilio 2



Nota. Se observan los mensajes de alerta en tiempo real con N8N. Elaboración propia.

Figura 43

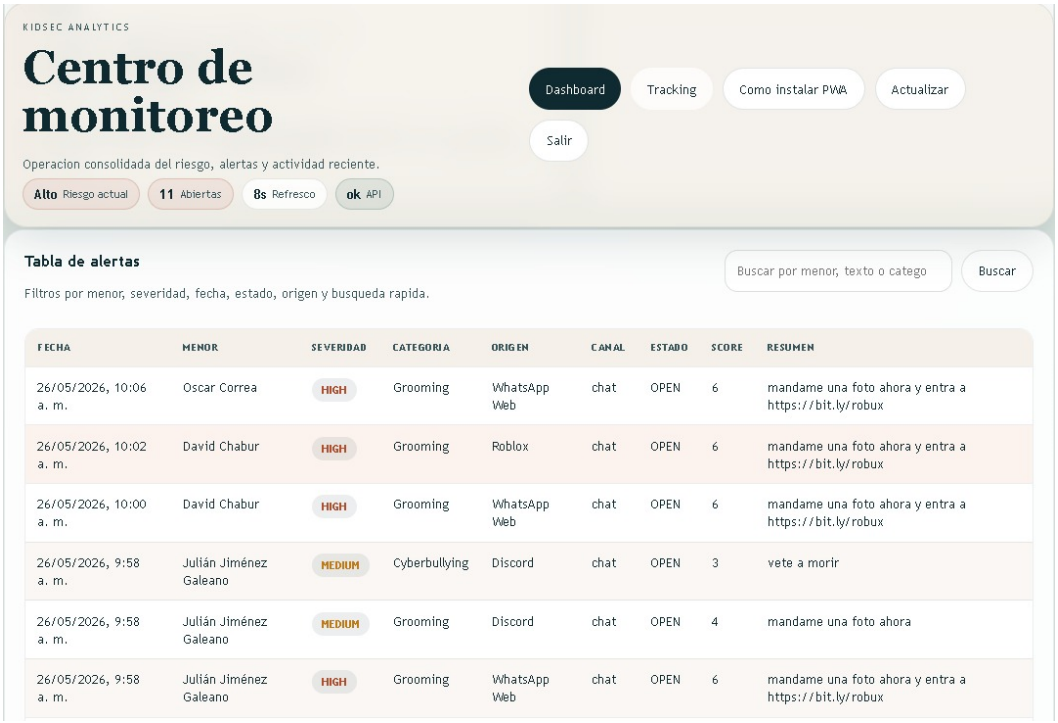
Alerta generada 3



Nota. Se observan los mensajes de alerta en tiempo real con N8N. Elaboración propia.

Figura 44

Historial de alertas registradas



Nota. Se observa la categoría y resumen de las alertas generadas en tiempo real. Elaboración propia.

Análisis de los Resultados

Los datos que se presentan en las encuestas realizadas permitieron identificar el grado de exposición digital de los niños y la existencia de malas experiencias en línea, se pueden ver gracias a las respuestas de los participantes. Se muestra la posibilidad de ser víctima de situaciones cuando se usa más tiempo el dispositivo móvil y más autonomía se tiene en el mismo, aumentando encontrarse con situaciones relacionadas con contactos sospechosos, robo de información y contacto con personas desconocías es mayor. Este descubrimiento está directamente relacionado con el objetivo de la investigación, que consiste en mitigar e identificar patrones peligrosos de comportamiento digital de manera temprana a través de un modelo basado en Machine Learning.

Los hallazgos de las encuestas se alienan con las opiniones de los 3 docentes entrevistados, quienes identifican una alta vulnerabilidad infantil debido a que los padres no hacen acompañamiento constante y exponen de forma temprana a sus hijos. Aunque los docentes ven viable y necesaria la implementación de un modelo de alertamiento automático en los colegios y mucho más en los hogares, consideran que su éxito normativo dependería de la capacitación comunitaria, una gran transparencia en el manejo de los datos y el respeto a la privacidad de los menores en Colombia.

La interpretación de los resultados permite afirmar que el componente más sólido de la propuesta fue la detección textual de riesgo. La combinación entre reglas léxicas y modelo estadístico mostró una capacidad alta para identificar señales asociadas con grooming, lenguaje abusivo, ciberbullying y solicitudes de información personal. Esto sugiere que las señales lingüísticas son más directas para la detección automática que las variables de perfil, ya que reflejan manifestaciones inmediatas del riesgo en la interacción digital.

No obstante, el análisis también evidencia una limitación metodológica importante, el conjunto textual estuvo fuertemente desbalanceado hacia la clase de riesgo. Aunque la

exactitud y el F1 fueron elevados, la cantidad de ejemplos seguros fue muy reducida. De acuerdo con la validación, en la partición de prueba hubo 130 casos de riesgo frente a solo 6 casos seguros, y estos últimos fueron clasificados erróneamente como riesgo. En consecuencia, aunque el recall de 99,23 % es muy alto, su interpretación debe hacerse con cautela, porque la capacidad real del modelo para reconocer casos seguros aún no está suficientemente validada con una base equilibrada.

Aun así, al analizar el comportamiento operativo del sistema con su umbral de alerta de 0,75, el desempeño siguió siendo favorable: 87,5 % de exactitud, 99,13 % de precisión, 87,69 % de recall y 93,06 % de F1. Esto indica que, en un escenario funcional, el sistema privilegia la precisión y reduce falsas alarmas, aunque asume el riesgo de no capturar la totalidad de los casos problemáticos. En términos prácticos, este equilibrio puede considerarse razonable en una herramienta de protección infantil, siempre que se complemente con supervisión humana y mejora continua del corpus textual.

En términos generales, los datos indican que, aunque una gran mayoría de los menores confirman no haber tenido experiencias negativas, hay un grupo importante que ha estado expuesto a situaciones potencialmente peligrosas, mayormente en las redes sociales y en videojuegos. Además, hay una gran diferencia entre lo que los padres identifican como un riesgo y lo que los niños realmente informan, lo cual demuestra las limitaciones de la supervisión tradicional.

Además, en los testimonios de las encuestas se revelan patrones que se repiten. Como lo son la solicitud de fotografías por parte de otras personas, la solicitud de información personal, las amenazas, los mensajes engañosos y él envió de enlaces sospechosos. Estos sucesos tienen conexión con las categorías de riesgo que se han identificado en el marco teórico, sobre todo en lo que hace al grooming digital, el ciberacoso, el malware y el phishing. Asimismo, los resultados muestran que las medidas que llevan los padres se enfocan sobre todo en el dialogo y las limitaciones, al mismo tiempo los problemas notificados demuestran

que los controles parentales tradicionales no siempre evitan los riesgos a los que el menor pueda estar expuestos. Por lo tanto, esto logra indicar que los métodos actuales son principalmente reactivos y necesitan una intervención antes que los sucesos ocurran.

En general, los patrones de uso digital (como la duración de la conexión, el tipo de actividad que se realiza en el dispositivo y el nivel de supervisión por parte de los padres) son variables importantes que pueden ser transformadas mediante la tecnología para clasificar niveles de riesgo. Por lo tanto, los resultados respaldan la facilidad de crear un sistema automatizado que permita la identificación de manera proactiva conductas digitales inusuales en los dispositivos móviles empelados por lo menores en Colombia.

Propuesta de Solución

A partir de los resultados, se propone consolidar un sistema de protección digital infantil de tres capas. La primera capa debe enfocarse en la detección inmediata de expresiones críticas mediante reglas léxicas y alertas automáticas. La segunda capa debe emplear clasificación textual con aprendizaje automático para reconocer expresiones de riesgo no previstas de forma literal en el léxico. La tercera capa debe incorporar una puntuación de perfil que permita identificar menores con mayor vulnerabilidad y orientar acciones preventivas por parte de padres, acudientes o instituciones educativas. Como mejora concreta, se recomienda ampliar la recolección de ejemplos textuales seguros, recalibrar el umbral operativo con una base más balanceada y mantener como variables prioritarias de intervención el control parental, la supervisión nocturna y el monitoreo de enlaces sospechosos.

Los resultados obtenidos permiten concluir que un modelo de Machine Learning tiene la capacidad de detectar rápidamente patrones relacionados con riesgos digitales en menores en Colombia. Se ha demostrado a través de la evidencia empírica que hay variables que se pueden observar como lo es, (tiempo de uso, clase de actividad que se realiza, interacción con personas en videojuegos, redes sociales y el grado de supervisión que tienen los padres que facilitan definir perfiles de riesgo.

Estos resultados indican que un sistema automatizado podría proporcionar una mejor detección en comparación con la supervisión parental actual, que solo se centra en bloquear cierto tiempo el dispositivo mas no identificar riesgos, aparte reducirá la dependencia total del control humano y reforzara la prevención de incidentes digitales.

Discusión de los Resultados

A partir de las encuestas realizadas en seminario de investigación, y la correspondiente búsqueda de artículos científicos, proyectos de grado y de más documentos pertinentes para el estado del arte se encontró que no hay ninguna solución aplicada en un contexto latinoamericano o específicamente al contexto colombiano. A partir de esto se planteo la pregunta de investigación ¿Hasta qué punto un modelo de Machine Learning autónomo, implementado en dispositivos móviles, hace posible la identificación de forma temprana de riesgos digitales en menores en Colombia y disminuir su exposición frente a la supervisión parental tradicional?

De acuerdo con Urbano (2024) afirma que la Inteligencia Artificial debe emplearse como una herramienta complementaria que no se encargue de tomar todas las decisiones, sino que se encargue de apoyar la correcta toma de decisiones sin sesgar el comportamiento del menor.

Con base en lo anterior, el responsable es quien toma las decisiones a partir de la información recolectada y organizada por el PWA. A diferencia de la metodología planteada en la literatura para el tratamiento y mitigación de este problema se seleccionó el método safety by design donde se prioriza la elaboración y correcto funcionamiento de módulos de seguridad, en este caso, el identificador de malas palabras o señales de riesgo.

En este sentido, se desarrollo una propuesta que funciona como un método alternativo a los ya existentes que no funcionan con efectividad al ser reactivos. Esta propuesta implementa machine Learning capaz de identificar señales de riesgo y malas palabras sin necesidad de la permanencia constate de los padres o responsables del menor, cuando este hace de dispositivos móviles y conexión a entornos digitales, como juegos online, redes sociales u otro tipo de aplicativos.

Finalmente, estos hallazgos permiten relacionar la supervisión humana y la inteligencia artificial como una solución asertiva en la prevención de riesgos para la seguridad de los menores en entornos digitales.

Plan de Divulgación

La viabilidad del plan de transferencia se fundamenta en la selección de la Revista Ontare como canal de divulgación, dada su especialización en ingeniería aplicada. Para asegurar la coherencia editorial, el manuscrito se adaptará a la tipología de artículo resultado de investigación, siguiendo la estructura IMRyD y el formato IEEE para la citación y referenciación. La gestión operativa contempla el proceso de evaluación doble ciego de la revista, estimando un periodo de tres meses para la revisión inicial del comité y otros tres meses adicionales para la evaluación por pares expertos. Durante esta etapa, se gestionará la carga de metadatos en la plataforma Open Journal Systems y se realizarán los ajustes documentales solicitados. La trazabilidad y el reconocimiento académico del autor se formalizan mediante el identificador ORCID 0009-0007-4433-6111, requisito necesario para integrar la producción intelectual con bases de datos especializadas. La preparación técnica del manuscrito incluye la transformación de la información del proyecto a un lenguaje académico, asegurando la precisión terminológica y la integridad de los resultados. El flujo de trabajo comprende la fase de alistamiento documental, la revisión de los componentes teóricos y el envío formal a través del sistema de gestión editorial.

Conclusiones

El diagnostico de las amenazas actuales en Colombia evidencio que la autonomía y el tiempo de exposición de los menores en aplicaciones cotidianas los expone a riesgos críticos como el contacto con personas desconocidas, cyberbullyng, grooming, sexting mientras se encuentran en línea. Las encuestas y entrevistas validaron la existencia de brechas críticas de conciencia y preparación digital en el entorno familiar y escolar en Colombia.

Dado esto, los modelos de aprendizaje automático evaluados, se logró determinar que el enfoque basado en procesamiento de lenguaje natural logra ofrecer una efectiva para identificar patrones de riesgo frente al análisis aislado de perfiles de usuario. No obstante, las pruebas metodológicas demostraron que, ante conjuntos desbalanceados en entornos reales, las métricas de precisión deben de interpretarse con cautela, priorizando algoritmos que equilibren un alto nivel de sensibilidad (recall) con una validación robusta en la clasificación de escenarios seguros.

Finalmente, se logró la consolidación de un prototipo funcional basado en Machine Learning optimizado para la detección temprana de patrones peligrosos en dispositivos móviles usados por lo menores. La viabilidad del aplicativo no solo se sustenta en la capacidad técnica de emitir alertas automáticas en tiempo real, sino alineándose con principios de transparencia y seguridad por diseño (safety by desing), lo que garantiza una herramienta de protección constante que respeta la privacidad de los menores y apoya la supervisión activa de los padres y educadores.

Referencias

Abarna, S., Sheeba, J. I., Jayasrilakshmi, S., & Devaneyan, S. P. (2022). Identification of cyber harassment and intention of target users on social media platforms. *Engineering Applications of Artificial Intelligence*. https://doi.org/10.1016/j.engappai.2022.105283

Abdullah, Latif, I., Hafeez, N., Ullah, F., Sidorov, G., Felipe-Riverón, E., & Gelbukh, A. (2025). Cyberbullying detection on social media using machine learning techniques. *Computación y Sistemas.* https://doi.org/10.13053/cys-29-3-5481

Ali, S., Elgharabawy, M., Duchaussoy, Q., Mannan, M., & Youssef, A. (2020, diciembre 7). *Betrayed by the Guardian: Security and Privacy Risks of Parental Control Solutions*. In Annual Computer Security Applications. https://dl.acm.org/doi/10.1145/3427228.3427287

Aliyeva, Ç. O., & Yağanoğlu, M. (2025). Deep learning approach to detect cyberbullying on Twitter. *Multimedia Tools and Applications.* https://doi.org/10.1007/s11042-024-19869-3

Alsuwaylimi, A. A., & Alenezi, Z. S. (2025). Leveraging transformers for detection of Arabic cyberbullying on social media: Hybrid Arabic transformers. *Computers, Materials & Continua.* https://doi.org/10.32604/cmc.2025.061674

Alvarez, D. G., Cardenoso, D. V. (s. f.). Sistemas de detección de intrusos basados en técnicas de Machine Learning. [Trabajo de fin de grado, Universidad vallolid]. Repositorio institucional.https://uvadoc.uva.es/handle/10324/44228

Argelich Comelles, C. (2025, mayo 29). La proyectada Ley Orgánica para la protección de las personas menores de edad en los entornos digitales: del control parental defectivo a las obligaciones a fabricantes [Artículo de investigación]. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5274283

Barbosa-Santillán, L. I., Guzman-Velazquez, B. P., Orozco-Aguilera, M. T., & Flores-Pulido, L. (2025). Unraveling cyberbullying dynamics: A computational framework empowered by artificial intelligence. *Information*. https://doi.org/10.3390/info16020080

Benaissa, A. R., Harbaoui, A., & Ben Ghezala, H. H. (2025). Class imbalance-sensitive approach based on pretrained language models for the detection of cyberbullying in English and Arabic datasets. *Behaviour & Information Technology*. https://doi.org/10.1080/0144929X.2024.2313142

Braga, P. R. V., & Tyrrell, K. R. (2025). Predicting cyberbullying victimisation in emerging markets and developing countries using the Global School-Based Health Survey. *Machine Learning with Applications*. https://doi.org/10.1016/j.mlwa.2025.100646

Caracol Radio. (2025, mayo 23). El 35% de niños entre 6 y 9 años en Colombia ya manejan un celular: Estudio CRC. Caracol Radio https://caracol.com.co/2025/05/24/el-35-de-ninos-entre-6-y-9-anos-en-colombia-ya-manejan-un-celular-estudio-crc/

Castro Chaparro, I. C., & Benítez Vizcaíno, P. I. (2025). *Estrategias con IA para la Psicología Infantil en Cergenial* [Trabajo de grado, Universidad de Santander]. Repositorio Institucional UDES. https://repositorio.udes.edu.co/server/api/core/bitstreams/64f50c5f-a893-49b5-8a24-9230a68bd77e/content

Chan, J., Kishore, S., & Yang, X. (2025). A real-time cyberbully checker. *Software Impacts.* https://doi.org/10.1016/j.simpa.2024.100710

Cirillo, S., Desiato, D., Polese, G., Solimando, G., Sugumaran, V., & Sundaramurthy, S. (2025). Exploring the ability of emerging large language models to detect cyberbullying in social posts through prompt-based classification approaches. *Information Processing & Management*. https://doi.org/10.1016/j.ipm.2024.104043

Coban, O., Ozel, S. A., & Inan, A. (2023). Detection and cross-domain evaluation of cyberbullying in Facebook activity contents for Turkish. ACM Transactions on Asian and Low-Resource *Language Information Processing*. https://doi.org/10.1145/3580393

Comité de los Derechos del Niño de las Naciones Unidas. (2021). Observación general núm. 25 (2021) relativa a los derechos de los niños en relación con el entorno digital (CRC/C/GC/25). Oficina del Alto Comisionado para los Derechos Humanos. https://www.ohchr.org/es/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

Daniel, R., Murthy, T. S., Kumari, C. D. V. P., Lydia, E. L., Ishak, M. K., et al. (2023). Ensemble learning with tournament selected glowworm swarm optimization algorithm for cyberbullying detection on social media. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3326948

Ejaz, N., Razi, F., & Choudhury, S. (2024). Towards comprehensive cyberbullying detection: A dataset incorporating aggressive texts, repetition, peerness, and intent to harm. *Computers in Human Behavior.* https://doi.org/10.1016/j.chb.2023.108123

Faraz, A., Ahsan, F., Mounsef, J., Karamitsos, I., & Kanavos, A. (2024). Enhancing Child Safety in Online Gaming: The Development and Application of Protectbot, an AI-Powered Chatbot Framework. *Information,* 15(4), 233. https://doi.org/10.3390/info15040233

Faraz, A., Mounsef, J., Raza, A., & Willis, S. (2022). *Child safety and protection in the online gaming ecosystem* [Artículo]. *IEEE Access*, PP (99), 10.1109/ACCESS.2022.3218415 https://www.researchgate.net/publication/365107295_Child_Safety_and_Protection_in_the_Online_Gaming_Ecosystem

Ferrer, R., Ali, K., & Hughes, C. (2024). Using AI-based virtual companions to assist adolescents with autism in recognizing and addressing cyberbullying. Sensors. https://doi.org/10.3390/s24123875

Fondo de las Naciones Unidas para la Infancia. (2020). *Investigating risks and opportunities for children in a digital world: A rapid review of the evidence on children's internet use and outcomes* (Innocenti Discussion Paper 2020-03). UNICEF Office of Research – Innocenti. https://www.unicef-irc.org/publications/1183

Gath, M., & Swit, C. (2024). Digital media in early childhood: Risk factors for online harm and psychosocial correlates. *Frontiers in Developmental Psychology*, *2*, https://doi.org/10.3389/fdpys.2024.1390276

Gómez Díaz, A. (2025, mayo 23). *El 61% de los menores ya tiene un celular propio*. Caracol Radio. https://caracol.com.co/2025/05/23/el-61-de-los-menores-ya-tiene-un-celular-propio

Gómez Varela, C. A. (2019, diciembre). *Detección de sitios web falsos en dispositivos móviles mediante algoritmos supervisados de aprendizaje automático.* [Tesis, Universidad CENFOTEC]. Repositorio Institucional CENFOTEC. http://20.55.226.204/handle/123456789/338

Guevara Tibaduiza, G. M. (2023). *Ciberseguridad: guía práctica para la protección de niños entre los 4 y 12 años de los riesgos en internet.* [Trabajo de grado - Especialización, Universidad Católica de Colombia]. RIUCaC. https://repository.ucatolica.edu.co/entities/publication/d8c6fc59-a42d-4e3f-b66c-ab3047b3faa5

Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.

Hinkley, T., Fisher, J., Iyer, P., & Birringa, L. (2024). Digital media in early childhood: risk factors for online harm and psychosocial correlates. *Frontiers in Developmental Psychology*, 3,

1390276. https://www.frontiersin.org/journals/developmental-psychology/articles/10.3389/fdpys.2024.1390276/full

Jerge, E. (2024, octubre 30). *Building Blocks: Investigating Child Exploitation and Safety Features of Minecraft* [Informe]. ClickSafe Intelligence. https://clicksafeintelligence.com/publications/building-blocks-investigating-child-exploitation-and-safety-features-of-minecraft

Jerge, E. (2024, septiembre 5). *Adverse Experiences in Online Gaming* [Informe]. ClickSafe Intelligence. https://clicksafeintelligence.com/publications/adverse-experiences-in-online-gaming

Jerge, E. (2025, julio 14). *Beyond the Virtual Playground* [Informe]. ClickSafe Intelligence. https://clicksafeintelligence.com/publications/beyond-the-virtual-playground

Livingstone, S. (2013). Online risk, harm and vulnerability: Reflections on the evidence base for child internet safety policy. *ZER: Journal of Communication Studies*, *18*, 13–28. https://eprints.lse.ac.uk/62278/

Lopes, R., Ta, V. T., & Korkontzelos, I. (2023, mayo). *On the conformance of Android applications with children's data protection regulations and safeguarding guidelines*. 10.48550/arXiv.2305.08492

Mane, S., Mandhan, K., Bapna, M., & Terwadkar, A. (2023). *Child Detection by Utilizing Touchscreen Behavioral Biometrics on Mobile*. In G. Ranganathan, G. A. Papakostas & Á. Rocha (Eds.), *Inventive Communication and Computational Technologies*.227– 244. https://doi.org/10.1007/978-981-99-5166-6_16

Mirabet Herranz, J. (2024). *Aplicación de técnicas de machine learning para la detección de malware en dispositivos Android.* [Trabajo fin de grado, Universitat Politècnica de

València]. Repositorio institucional UPV. https://riunet.upv.es/server/api/core/bitstreams/dfaef6bb-8a0d-4a94-8a68-ab4c0671a3a2/content

Nguyen, T., Roy, A., & Memon, N. (2019, julio). *Kid on The Phone! Toward Automatic Detection of Children on Mobile Devices.* Computers & Security, 84, 334-348, https://doi.org/10.1016/j.cose.2019.04.001

Nina-Gutiérrez, E. A., Pacheco-Alanya, J. E., & Morales-Arevalo, J. C. (2024). SocialBullyAlert: A web application for cyberbullying detection on minors' social media. *International Journal of Advanced Computer Science and Applications.* https://doi.org/10.14569/IJACSA.2024.0150776

Olguín Romero, A., & Arana-Llanes, J. Y. (2024). Ataques a teléfonos celulares mediante el uso de aplicaciones móviles: una revisión narrativa. *Tecnociencia Chihuahua*, sin paginacion. https://revistascientificas.uach.mx/index.php/tecnociencia/article/view/1584

Özaydın Aydoğdu, Ö. (2023, agosto 11). Designing, Developing and Examining the Effectiveness of a Machine Learning–Based Mobile Recommendation System for Parents' Digital Parenting Skills. *Child & Family Social Work.* https://doi.org/10.1111/cfs.13218

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. https://doi.org/10.1136/bmj.n71

Pellufó, I. C., Capobianco, M., Echaiz, J. (2014, mayo). Machine Learning aplicado en sistemas de detección de intrusos. XVI Workshop de Investigadores en Ciencias de la Computación (WICC). [Trabajo de fin de grado, Universidad Nacional de la Plata]. Repositorio institucional de la UNPL. https://sedici.unlp.edu.ar/handle/10915/43264

Peñuela Jiménez, L. C. (2023). Ciberseguridad: el peligro al acecho de los niños, niñas y adolescentes. Informática y Derecho: *Revista Iberoamericana de Derecho Informático* (segunda época), 15–28. https://dialnet.unirioja.es/servlet/articulo?codigo=9265273

Pérez, J., Castro, M., Awad, E., & López, G. (2023). Virtual Harassment, Real Understanding: Using a Serious Game and Bayesian Networks to Study Cyberbullying. https://www.researchgate.net/publication/376059077_Virtual_Harassment_Real_Understanding_Using_a_Serious_Game_and_Bayesian_Networks_to_Study_Cyberbullying

Ptaszek, G., & Pyżalski, J. (2023). Children's vulnerability to digital technology within the family: A scoping review. Societies, 13(1), 11. https://www.mdpi.com/2075-4698/13/1/11

Puetate Paucar, J. M., Toro Hernández, G. A., & Coka Flores, D. F. (2024). Deepfake: un desafío latente para la seguridad de los menores ecuatorianos en el entorno digital. Dilemas contemporáneos: Educación, Política y Valores. https://dilemascontemporaneoseducacionpoliticayvalores.com/index.php/dilemas/article/ view/4379

Ramírez, D. M., Garcés-Giraldo, L. F., Doria-Orozco, T., Franco-Castaño, S., Valencia-Arias, A., Rodríguez-Correa, P. A., & Espinoza Román, J. (2023). Tendencias investigativas en el uso de Machine Learning en la ciberseguridad. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 60–72. https://www.proquest.com/openview/c9bf3f3b2192c011f0620adce0649ab3/1?pq-origsite=gscholar&cbl=1006393

Saadi, M. Q., & Dhannoon, B. N. (2023a). Arabic cyberbullying detection using support vector machine with cuckoo search. Iraqi Journal of Science. https://doi.org/10.24996/ijs.2023.64.10.37

Saadi, M. Q., & Dhannoon, B. N. (2023b). Comparing random forest vs. extreme gradient boosting using cuckoo search optimizer for detecting Arabic cyberbullying. Iraqi Journal of Science. https://doi.org/10.24996/ijs.2023.64.9.40

Sahana, V., Anil Kumar, K. M., & Darem, A. A. (2023). A comparative analysis of machine learning techniques for cyberbullying detection on FormSpring in textual modality. International Journal of Computer Network and Information Security. https://doi.org/10.5815/ijcnis.2023.04.04

Salat Paisal, M. (2025). El uso de la inteligencia artificial en la investigación policial y judicial de delitos de online child sex grooming en España. *IDP. Revista de Internet, Derecho y Política*, (42), 1–12. https://www.raco.cat/index.php/IDP/article/view/430914

Sallahuddin, S., Hameed, M. F., Ismail, M., Ali, S., & Ahmed, A. (2025). SafeCon: AI-powered real-time cyber grooming detection system. VFAST Transactions on Software Engineering. https://doi.org/10.21015/vtse.v13i2.2118

Street, J., Ihianle, I. K., Olajide, F., & Lotfi, A. (2025). Enhanced online grooming detection employing context determination and message-level analysis. Intelligent Systems with Applications. https://doi.org/10.1016/j.iswa.2025.200607

Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., et al. (2023). A review of machine learning techniques in cyberbullying detection. Computers, Materials & Continua. https://doi.org/10.32604/cmc.2023.033682

Sun, R., Xue, M., Tyson, G., Wang, S., Camtepe, S., & Nepal, S. (2023, abril 30). *Not Seen, Not Heard in the Digital World! Measuring Privacy Practices in Children's Apps.* Association for Computing Machinery. 10.1145/3543507.3583327

Tolosana, R., Ruiz-Garcia, J. C., Vera-Rodriguez, R., Herreros-Rodriguez, J., Romero-Tapiador, S., Morales, A., & Fierrez, J. (2022). *Child-Computer Interaction with Mobile Devices: Recent Works, New Dataset, and Age Detection.* IEEE Transactions on Emerging Topics in Computing. 10.1109/TETC.2022.3150836.

UNAD, U. N. (2021). *Ciberataques, riesgos y consecuencias que han afectado a la población colombiana entre los años 2018 y 2020.* [Proyecto de grado, Universidad Nacional Abierta y a Dustancia]. Repositorio Institucional UNAD. https://repository.unad.edu.co/handle/10596/51582

Urbano, R. (2024). Strategies to protect children from violence in artificial intelligence. *Rotura – Revista de Comunicação, Cultura e Artes*, 40-49. https://publicacoes.ciac.pt/index.php/rotura/article/view/222

Zaccagnino, R., Capo, C., Guarino, A., Lettieri, N., & Malandrino, D. (2021, febrero 6). Techno-regulation and intelligent safeguards: Analysis of touch gestures for online child protection. *Multimedia Tools and Applications, 80*(21 -23), 32159–32186. https://link.springer.com/content/pdf/10.1007/s11042-020-10446-y