ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

IMPLEMENTACIÓN DE UN PROTOTIPO DE CONTROL PARENTAL ENFOCADO EN LA DETECCIÓN INTELIGENTE DE ATAQUES DE GROOMING

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

CHRISTIAN DAVID OÑA SALAZAR
e-mail: christian.ona@epn.edu.ec
JAIRO GEOVANNY PROAÑO CHAVIZNAN
e-mail: jairo.proano@epn.edu.ec

DIRECTOR: Ph.D. Pamela Catherine Flores Naranjo
e-mail: *pamela.flores@epn.edu.ec*

CO-DIRECTOR: Ph.D. Jenny Gabriela Torres Olmedo
e-mail: *jenny.torres@epn.edu.ec*

Quito, noviembre 2020

DECLARACIÓN

Nosotros, **Christian David Oña Salazar** y **Jairo Geovanny Proaño Chaviznan**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedemos nuestros derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Christian David Oña Salazar

Jairo Geovanny Proaño Chaviznan

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por **Christian David Oña Salazar** y **Jairo Geovanny Proaño Chaviznan**, bajo mi supervisión.

Pamela Catherine Flores Naranjo, Ph.D.

DIRECTOR DE PROYECTO

DEDICATORIA

Dedico este trabajo a Pablo, Lucia y Pablito mis padres y mi hijo, lo más importante en mi vida.

Christian Oña

Dedico este trabajo a mi madre, mi padre, y mis hermanas, por su apoyo y motivación a lo largo de mis estudios.

Jairo Proaño

AGRADECIMIENTO

Agradezco con gran afán a todas las personas que se cruzaron en mi camino estos años de crecimiento personal y profesional. A mis profesores por enseñarme a ser el mejor en esta área y en especial por enseñarme los valores que conllevan un título. A mis amigos por estar en cada trabajo, deber y juego junto a mi mientras avanzábamos juntos hasta el día de hoy, a todos les deseo los más grandes éxitos, en especial a mi amigo Jairo la mejor persona que he conocido en mi vida siguiéndome hasta en las ideas más ocurridas que he tenido y siendo un gran soporte en mi vida, con quien sé que puedo contar hasta en mis peores momentos a mi familia. A mis hermanos y mi hijo siempre mostrándome que no puedo rendirme por ellos, mis padres que me dieron todo para que pueda llegar con éxito donde estoy. A la Escuela Politécnica Nacional por acogerme en sus aulas donde nacieron muchos sueños, amistades, alegrías y tristezas mientras pasaban los años en ella, dándome algunos de los mejores momentos de mi vida. A Pamela Flores y Patricio Zambrano por su apoyo incondicional mientras trabajábamos en este proyecto. A Mishell quien se preocupó por mí en cada momento y que siempre quiso lo mejor para mi porvenir. Nunca podré terminar de agradecerles a todos por las experiencias que me dieron la oportunidad de vivir junto a ellos.

Christian Oña

Agradezco a mis padres Silvia y Joselo, por su apoyo, amor, cuidado y confianza a lo largo de mi vida, quienes a pesar de los problemas supieron encontrar la forma de darme lo necesario para seguir adelante. A mis abuelitos, tíos y tías por su paciencia y respaldo durante esta etapa. A Christian, por su amistad, quien conocí en la carrera, y desde entonces nos hemos apoyado durante tantos años, quien ha demostrado ser una persona confiable y capaz, y es la persona con quien he desarrollado este trabajo. A Pamela Flores y a Patricio Zambrano, por su guía, apoyo, conocimientos y tiempo brindado para que mi compañero y yo pudiéramos finalizar satisfactoriamente este proyecto.

Jairo Proaño

CONTENIDO

RESUMEN

Los controles parentales son herramientas que le permite a tutores de menores limitar el acceso a contenido especifico en Internet . Los accesos se definen previo al uso del dispositivo a ser monitoreado, sin embargo, su funcionalidad hasta el momento no ha permitido detectar ataques del tipo de ingeniería social como el Grooming. Donde Grooming se define como el proceso de acicalar a menores por internet con fines sexuales. Estos ataques de ingeniería social son difíciles de detectar y mitigar por sistemas de software ya que presentan semántica de lenguaje natural humano donde lo principal para su identificación sería entender las intenciones de un comunicador a través de un texto.

Para resolver este problema se propone el desarrollo de un sistema de control parental inteligente que permite detectar ataques de Grooming de textos extraídos de la red social Facebook en tiempo real y que alerte cuando uno de estos ataques se identifique. El proyecto se desarrollo usando el marco de trabajo Scrum que permitió ser flexibles y ordenados, ya que se contemplo que durante el desarrollo del control parental podrían existir cambios e integración de nuevas funcionalidades.

El presente trabajo de titulación basa su desarrollo en un estudio previo realizado por Zambrano et al. [1] en el cual se define al Grooming como un ataque de ingeniería social con un ciclo de vida de 6 estaciones. Con la definición de un proceso parecido se generó un clasificador lineal que logra identificar en cuál de estos niveles se encuentra un texto. El clasificador se integró dentro de un sistema web bajo una arquitectura basada en servicios REST. Se creó un complemento para Google Chrome que toma los mensajes recibidos a través de la red social Facebook y los envía al clasificador, este a su vez notifica a un tercer componente de escritorio de este sistema que se encarga de verificar, de acuerdo a las configuraciones del control parental si el nivel en el que fue clasificado el mensaje es aceptado o no, en caso de que el mensaje infrinja las configuraciones este tercer componente bloqueará la conexión con el sitio web Facebook y notificará vía Telegram que se detectó un potencial ataque y que la comunicación con el sitio ha sido bloqueada.

Una vez terminado el software se realizaron pruebas de efectividad al clasificador de texto usando 70% del conjunto de datos que se tenía para entrenar al clasificador y el 30% para probar su efectividad. Con estos datos, las pruebas devolvieron una precisión del 94% en la clasificación marcándose como más que aceptable. Se realizaron también pruebas de funcionalidad mediante encuestas de aceptación de uso a padres y tutores de menores de edad. Estas nos mostraron un alto índice de aceptación y también mostraron que hay cierta duda en la manera con la que se maneja este tipo de sistemas ya que al momento leer los mensajes del menor en texto plano se ve afectada la privacidad de los mensajes del menor,

poniendo en una balanza la privacidad del menor contra su seguridad, tema que debería ser tratado en futuros trabajos.

Palabras clave: Grooming, Clasificador lineal, Scrum.

ABSTRACT

Parental controls are tools that allow guardians of minors to limit access to specific content on the Internet. The accesses are defined prior to the use of the device to be monitored, however, its functionality so far has not allowed detecting attacks of the type of social engineering such as grooming. Where Grooming is defined as the process of grooming minors online for sexual purposes. These social engineering attacks are difficult to detect and mitigate by software systems since they present semantics of human natural language where the main thing for their identification would be to understand the intentions of a communicator through a text.

To solve this problem, it is proposed to develop an intelligent parental control system that allows detecting Grooming attacks on texts extracted from the social network Facebook in real time and that alerts when one of these attacks is identified. The project was developed using the Scrum framework that allowed it to be flexible and orderly, since it was contemplated that during the development of parental control there could be changes and integration of new functionalities.

The present degree work bases its development on a previous study carried out by Zambrano et al. [1] in which Grooming is defined as a social engineering attack with a 6 season life cycle. With the definition of a similar process, a linear classifier was generated that manages to identify in which of these levels a text is found. The classifier was integrated into a web system under an architecture based on REST services. A plugin was created for Google Chrome that takes the messages received through the Facebook social network and sends them to the classifier, this in turn notifies a third desktop component of this system that is in charge of verifying, according to the settings of parental control if the level at which the message was classified is accepted or not, in case the message violates the settings, this third component will block the connection with the Facebook website and will notify via Telegram that a potential attack was detected and that communication with the site has been blocked.

Once the software was finished, effectiveness tests were performed on the text classifier using 70% of the data set that was had to train the classifier and 30% to test its effectiveness. With these data, the tests returned a 94% accuracy in the classification, marking it as more than acceptable. Functionality tests were also carried out through user acceptance surveys to parents and guardians of minors. These showed us a high rate of acceptance and also showed that there is some doubt in the way in which this type of system is handled since when reading the child's messages in plain text, the privacy of the child's messages is affected, weighing the privacy of the minor against their safety, a topic that should be addressed in future work.

Keywords: Grooming, Linear Classifier, Scrum.

1. INTRODUCCIÓN

1.1. Planteamiento del problema

El Grooming, término definido como la acción de un adulto al acosar a los niños con finalidades sexuales, es uno de los ataques más efectuados en contra de niños de 5 a 17 años a través de Internet, con el objetivo de generar relaciones amistosas engañosas entre atacante y víctima [1] buscando disminuir las preocupaciones del menor y así poder abusar sexualmente de él.

Según el informe de UNICEF [2], "El Estado Mundial de la Infancia 2017: Niños en un mundo digital" el acceso de los niños a Internet les da enormes oportunidades, pero también hace que sean más susceptibles a sufrir daños en línea y fuera de línea. En la mayoría de los casos cada uno de estos niños tienen acceso libre a Internet en sus casas, escuelas, colegios y centros de cómputo sin un control adecuado para identificar cuáles son sus búsquedas y con quien hablan, lo que conlleva peligros de acceso a sitios de contenido para adultos, así como el peligro de ser víctimas de ataques como el Grooming o engaño pederasta.

Para el 2012 se estimaba que la cantidad de imágenes de abusos sexuales a niños en Internet, ascendían a los millones, y la cantidad de videos rondaban las decenas de millones [2]. En una encuesta realizada en Inglaterra se reportó que tan solo el 22% de los padres encuestados eran totalmente conscientes del tipo de ataques que podían recibir sus hijos a través de Internet. Además, se reportaron 5 mil ofensas por Grooming en menos de 18 meses en el mismo país, en contraste con los 3 mil casos reportados en el 2017 [4]; por lo que se puede entender que no es un caso aislado.

El incremento del uso de teléfonos y dispositivos inteligentes, hacen que el acceso a Internet en lugares privados sea más accesible, haciendo complicado a los padres vigilar la actividad de los niños en línea. Adicionalmente, estudios demuestran que cuando los niños reciben ofensas a través de Internet, los niños tienden a confiar más en sus amigos que en los padres [5]. Esto evidencia la falta de mecanismos de control a nivel tecnológico que le permitan a los padres monitorear las actividades que hacen sus hijos sobre Internet. Actualmente, las aplicaciones de control parental orientado al uso de Internet solo se limitan a restringir la cantidad de tiempo que los niños pueden estar navegando, y otras características como, gestión de gastos, notificación de cuando los niños agregan nuevos

amigos, restringen la información que se puede hacer pública, entre otras [3]; pero ninguna de esos controles se ha centrado en identificar ataques de Grooming.

Las investigaciones relacionadas a estudiar este fenómeno (Grooming) son escasas. Un trabajo relacionado encontrado en la literatura es el de Bogdanova et al. [9], en donde se realizaron experimentos usando cadenas léxicas, y análisis de sentimientos pertenecientes al área de Procesamiento de Lenguaje Natural (PLN), con lo que obtuvieron una precisión de hasta el 97% en la determinación del perfil psicológico del pedófilo, en donde, en base a características especialmente de inestabilidad emocional del atacante lograron identificarlo en las conversaciones. De manera complementaria, en la investigación de Zambrano et al. [1] se define el Grooming como un ataque informático con un ciclo de vida (6 estaciones). Esto permite que después de un procesamiento de texto, las conversaciones se puedan ubicar dentro de una de las estaciones del ciclo de vida del Grooming e identificar las intenciones comunicacionales del actor o atacante en la conversación.

Aspectos o factores como las prohibiciones legales de cada país para la obtención, tratamiento y recopilación de datos relacionados a los niños y adolescentes, acceso a bases de conocimientos específicas del Grooming que permitan entrenar sistemas inteligentes, unificación de criterios científicos que evalúen la naturaleza del problema desde el punto de vista de la seguridad de la información y el acceso a nuevas tecnologías propietarias, podrían ser las causas del desinterés de los investigadores para generar aportes relevantes en beneficio de prevenir y mitigar esta problemática social.

El presente trabajo de titulación plantea esquematizar y desarrollar un aplicativo de control parental funcional, apoyado en aprendizaje de máquina, que tenga las capacidades de capturar conversaciones de Facebook en tiempo real desde un explorador (Google Chrome). Luego de la captura de éstas, el sistema procederá a analizarlas y ubicarlas en una de las estaciones del ciclo de vida del Grooming a través de un motor de inteligencia artificial. En caso de que una conversación sea ubicada dentro de una estación previamente especificada por el tutor del niño, el sistema tendrá la capacidad de bloquear automáticamente la comunicación y notificará al tutor sobre este incidente.

En el primer semestre del 2019 se ha visto un aumento de casi un 50% en los delitos relacionados con el Grooming en comparación con el mismo período del 2018 [4], por lo que podemos aseverar que este crecimiento podría aumentar en futuros años si no se implementan sistemas inteligentes como el propuesto en este trabajo de titulación.

1.2. Objetivos

1.2.1. Objetivo General

Implementar un prototipo de control parental enfocado en la detección inteligente de ataques de Grooming.

1.2.2. Objetivos Específicos

- Desarrollar un Plug-in de extracción de datos para el Navegador Google Chrome.
- Desarrollar un API de clasificación de texto por tópicos.
- Desarrollar la aplicación de escritorio para control parental y alerta al tutor.
- Evaluar la funcionalidad del sistema con todos sus componentes integrados.

1.3. Alcance

En el presente proyecto se implementará un prototipo de control parental que demuestre la viabilidad de un sistema inteligente que clasifique con la mayor efectividad posible mensajes de conversaciones reales dentro de la red social Facebook, capaz de detectar ataques de Grooming en estas conversaciones y restringir el acceso de la víctima al sitio web de la red social.

Por lo tanto, el alcance de este proyecto es tomar el contenido de los mensajes que recibe un menor, y clasificar el contenido dentro de uno de los 6 estados que pertenecen al ciclo de vida del Grooming definidos dentro del trabajo de investigación de Zambrano et al. [1]. El prototipo estará formado por 5 componentes principales. Un interceptor de mensaje, un clasificador de Grooming, un sistema de cola de mensajes, una aplicación de gestión de bloqueos, y un bot de Telegram.

El producto final consta de los siguientes componentes:

- Una aplicación de escritorio que levanta un sistema local para que los tutores del menor puedan gestionar los estados aceptables en el ciclo de vida del Grooming y puedan desbloquear al sitio si este fue bloqueado.

- Un servidor de clasificación de mensajes, centralización de datos e interconexión de componentes.

- Una extensión para Google Chrome que recolectará los mensajes de la red social Facebook y los enviará al servidor de clasificación.

2. MARCO TEÓRICO

Este capítulo hace referencia al marco teórico, donde se recopila la definición de varios términos relacionados a la problemática de este proyecto. Estas definiciones se presentan de una manera inductiva tomando como punto de partida otras literaturas que estudian problemáticas relacionadas, adentrando progresivamente al tema con las definiciones de pedofilia, abuso sexual infantil, delincuentes en Internet y el Grooming. En base a esto se genera una definición a usar en el resto del trabajo que permita al lector familiarizarse con el tema.
Posteriormente se presenta la funcionalidad de los sistemas de control parental, de los clasificadores lineales y de como estos se pueden acoplar para crear un prototipo de control parental enfocado en la detección inteligente de ataques de Grooming.

2.1. SCRUM

Scrum fue presentado por Takeuchi, DeGrace, Schwaber y otros a fines de la década de 1990 [37]. Es presentado como un marco de gestión ágil para el desarrollo incremental de productos, que utiliza uno o más equipos autoorganizados y multifuncionales de aproximadamente siete personas cada uno, en donde este numero puede variar de acuerdo a las necesidades del proyecto. Es un método de gestión de proyectos iterativo, cronometrado e incremental basado en un marco simple de "inspección y adaptación" [38].

El marco de trabajo de Scrum se basa en un conjunto de valores, principios y prácticas que prestan la base para la gestión del proyecto. Se concentra en la toma de decisiones a partir de resultados del mundo real en lugar de suposiciones es decir en sucesos reales en el proyecto. Proporciona una estructura de roles, reuniones, reglas y artefactos. Los equipos son responsables de crear y adaptar sus procesos dentro de este marco. Scrum usa iteraciones de longitud fija, llamadas Sprints, que duran desde dos semanas a 30 días. Los Sprints se limitan a este tiempo de manera muy estricta, es decir, terminan en una fecha específica, ya sea que el trabajo se haya completado o no, y nunca se extienden [39].

El proceso que sigue scrum inicia con la recopilación de requisitos en forma de historias de usuarios de los clientes, los equipos y otras partes interesadas. Todas estas historias de usuarios están organizadas en orden de prioridad en el Product Backlog. Los elementos de alta prioridad se seleccionan del Product Backlog para implementarse en el próximo Sprint. Durante el Sprint, no se pueden agregar nuevos elementos del Product Backlog. Al comienzo del Sprint, se lleva a cabo una reunión de planificación en la que los miembros

del equipo calculan con cuántas tareas pueden comprometerse y luego crean un Sprint Backlog, una lista de las tareas a realizar durante el Sprint [39]. Durante cada Sprint, el equipo de Scrum trabaja en el desarrollo del requisito, incluido su diseño, codificación y pruebas para obtener una funcionalidad totalmente implementada. Al final, estas características se codifican, prueban e integran en el producto o sistema en desarrollo.

Para cada Sprint se hace una reunión diaria en la que se verifica avances para su revisión y limitaciones de los participantes para poder ayudar a superarlas. Al finalizar el Sprint se hace una reunión con todas las partes interesadas y para presentarle los resultados del Sprint y obtener una retroalimentación, esta retroalimentación se suma al siguiente Sprint es decir se realiza una adaptación del Product Backlog incluyendo las actividades terminada y de ser necesario se aumentan más [39][40].

Schwaber et al. en [37] dividen estas prácticas en: Scrum Master, Product Backlog, Scrum Teams, Daily Scrum Meetings, Sprint Planning Meeting, Sprint y Sprint Review. Prácticas que son adoptadas en el presente trabajo para su ejecución.

2.1.1. Product Backlog

El Product backlog es una lista priorizada de requisitos del proyecto con tiempos estimados para desarrollar una funcionalidad completa del producto. El Product backlog se actualiza continuamente para reflejar los cambios en las necesidades del cliente, nuevas ideas, nuevas necesidades, obstáculos técnicos que aparecen, entre otros. [39]. El dueño del producto es responsable de priorizar los elementos del Product backlog después de consultar con el equipo y las partes interesadas. Todas las entradas del Product Backlog deben estimarse en términos de esfuerzo. Esta estimación se puede utilizar para priorizar las entradas en el Product Backlog y para planificar los Sprints.

El equipo y el propietario del producto deciden la estimación de esfuerzo y las estimaciones de riesgo técnico para cada punto del Product Backlog. Los elementos más grandes en el Product Backlog se dividen en elementos más pequeños y se les asignan prioridades individuales. Esto se conoce como refinamiento del Product Backlog [39].

2.1.2. Sprint Backlog

El Sprint Backlog es una lista de tareas identificadas por el equipo que se completarán durante el siguiente Sprint. Durante la Sprint Planning Meeting, el equipo selecciona los elementos de mayor prioridad del Product backlog, generalmente en forma de historias de

usuarios, e identifica las tareas necesarias para completar cada historia de usuario. La mayoría de los equipos también calculan cuántas horas llevará completar cada tarea a alguien en el equipo [40] [37]. Solo el equipo está autorizado a cambiar la selección de elementos dentro del Sprint [40]. Las tareas deben detallarse en términos de horas hombre. Las tareas incluyen información sobre el trabajo que debe realizarse. A veces, el Sprint Backlog está vinculado con el Product Backlog para poder identificar el progreso de los elementos del Product Backlog [40].

2.1.3. Sprint Burn-down Chart

Sprint Burn-down Chart es una gráfica utilizada por los equipos para identificar el esfuerzo de desarrollo que queda en un Sprint. Muestra, cada día, una nueva estimación del trabajo restante hasta que el equipo finalice el Sprint. Se lo llama gráfico burn (quemado), ya que es un gráfico con pendiente descendente y llega a cero el último día del Sprint [39]. Todos los días, los miembros del equipo trabajan en sus tareas y el trabajo debe disminuir todos los días. Por lo general, en el eje horizontal, se trazan el tiempo o el número de Sprints, mientras que, en el eje vertical, se traza el trabajo restante (esfuerzo u horas-hombre).

2.1.4. Reunión de planificación de Sprint

Al comienzo de cada Sprint, se lleva a cabo la reunión de planificación de Sprint. Las entradas para la reunión de planificación de Sprint son la capacidad del equipo y el Product Backlog [41]. La reunión está dividida en partes, cada una de ellas con un calendario. Durante la primera parte, el Product Owner presenta los elementos del Product Backlog de mayor prioridad al equipo. El equipo y el Product Owner colaboran para ayudar al equipo a estimar la cantidad de Product Backlog que puede desarrollar en el próximo Sprint. La segunda parte de la reunión se centra en cómo implementar los elementos que el equipo decide tomar para ese Sprint. El equipo estima la cantidad de elementos del Product Backlog que pueden completar al final del Sprint. Los elementos de mayor prioridad se seleccionan primero para la implementación.

2.1.5. Reunión diaria de Scrum

Una reunión diaria de Scrum es una reunión corta de 15 minutos. En la reunión diaria de Scrum, cada miembro del equipo informa tres cosas a los otros miembros del equipo [37]:

1. ¿Qué avances se tiene desde la última reunión?
2. ¿Qué se pretende hacer antes de la próxima reunión? y
3. ¿Qué obstáculos se ha tenido en el avance de alguna tarea?

Las reuniones diarias de Scrum mejoran la comunicación, la colaboración y el conocimiento sobre el progreso actual. El Scrum Master es responsable de organizar eficazmente las reuniones diarias de Scrum.

2.1.6. Revisión del Sprint

Después de que finaliza el Sprint, hay una reunión donde las personas revisan el Sprint. El Product Owner, el Scrum Master, el equipo y todas las partes interesadas participan en esta reunión. En esta revisión se discute lo que se ha logrado en el Sprint. La Revisión de Sprint es una actividad de inspección y adaptación para el producto [39]. La revisión de Sprint no debe durar más de 30 minutos [39].

2.1.7. La Retrospectiva de Sprint

La Retrospectiva de Sprint, que sigue a la Revisión, es una actividad de inspección y adaptación relacionada con el proceso y el entorno. En esta reunión, el equipo discute qué prácticas están demostrando ser beneficiosas y cuáles no [39]. Aquí se analiza el BurnDown Chart para verificar que la manera en la que se trabajó permitió una eficiente ejecución de las tareas del Sprint.

2.1.8. Roles de Scrum

El marco trabajo de Scrum incluye tres roles, Product Owner o propietario del producto, Team o equipo y Scrum Master.

- **El propietario del producto (Product Owner)** es la parte interesada clave y es responsable del retorno de la inversión. El propietario del producto prioriza la acumulación de productos durante la reunión de planificación de Sprint. Ella/Él es responsable de que Product Backlog se actualice, sea visible y se priorice todo el tiempo [39].

- **El equipo (Team)** en Scrum es multifuncional; para un producto de software, el Equipo podría incluir personas con habilidades en análisis, desarrollo, pruebas, diseño de interfaces, diseño de bases de datos, arquitectura, documentación, etc. [39]. Incluye toda la experiencia necesaria para entregar el producto integrable en cada Sprint y debe ser autogestionado [39]. El equipo decide cuántos elementos del Product Backlog se seleccionarán para implementar en un Sprint. El Equipo en Scrum está conformado de 5 a 9 personas [39]. El Equipo trabaja en los requisitos

para desarrollar un producto funcional y también proporciona ideas al propietario del producto para mejorar la productividad.

- **El Scrum Master** ayuda al equipo a aprender y aplicar Scrum para alcanzar el valor comercial [39]. Como el equipo ya es un equipo autoorganizado, el Scrum Master mejora la autogestión, la funcionalidad, la creatividad y el empoderamiento [43]. El Scrum Master es el facilitador de las reuniones y es responsable de garantizar que los miembros del equipo puedan continuar con sus tareas.

2.2. Grooming

La fase inicial de la explotación sexual infantil es el Grooming o proceso de acicalamiento. Según la *National Society for the Prevention of Cruelty to Children* (NSPCC) el Grooming se define actualmente como [22]:

> *"Cuando alguien construye una conexión emocional con un niño para ganar su confianza con el propósito de abuso sexual, explotación sexual o tráfico. Los niños y los jóvenes pueden ser acicalados en línea o cara a cara, por un extraño o por alguien que conozca, por ejemplo, un familiar, amigo o profesional.*
> *Los Groomers pueden ser hombres o mujeres. Podrían ser de cualquier edad. Muchos niños y jóvenes no entienden que han sido acicalados o que lo que sucedió es abuso ".*

De manera condensada, Olsen et al. [23] definieron al Grooming como un proceso de desviación social, donde el resultado deseado es el abuso sexual del menor que está siendo atacado. Adicional a esto hay que aclarar que un Groomer es un término usado para referirse a una persona que ejecuta el Grooming.

2.2.1. Ciclo de vida Grooming

En la literatura de Zambrano et al. [1], desde el punto de vista de la seguridad informática mediante el uso de algoritmos de aprendizaje computacional, se formaliza al Grooming como un ataque informático dentro del campo de la ingeniería social contrastando sus

etapas o fases con los ciclos de vida asociados con las llamadas amenazas persistentes avanzadas (APT)

La definición formal del ciclo de vida del Grooming se puede ver en Figura 1, el mismo que se tomará en el presente trabajo como referencia teórica durante el desarrollo.

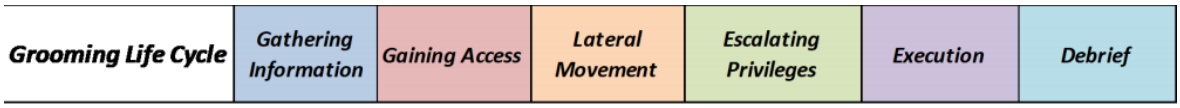


Figura 1. Conceptos operacionales (definición) relacionados con las intenciones comunicacionales [1]

Como última aclaración hay que decir que dentro de los resultados presentados por Zambrano et al. [1] es una colección con las frases de los chats reales de ciber-pedófilos ya clasificadas en una de las fases del ciclo de vida del Grooming. Un ejemplo se puede observar en Figura 1. Este documento se lo usará más adelante en el presente trabajo como base para generar un clasificador lineal.

2.2.2. Otras definiciones

En la presente sección se presentarán las definiciones relacionadas con el Grooming, las mismas que permitirán al lector entrar en contexto sobre el tema que se trata en el presente proyecto.

- **Pedofilia**

Para definir la pedofilia usaremos una perspectiva basada en fenómenos, es decir, más orientada por los preceptos que están relacionados con la investigación y descripción de estos como tal y como se experimentan conscientemente y lo más libremente posible de teorías, presuposiciones y preconceptos respecto a su origen [13].
Así, basándonos en investigaciones relacionadas se pudo determinar que la pedofilia puede interpretarse como una orientación sexual [12]. Al igual que la heterosexualidad, la homosexualidad y la bisexualidad, representa una categoría distinta de interés sexual y, en distintos grados, deseos de amor y apego romántico.
Las personas con Pedofilia se sienten sexualmente atraídos por niños que aún no han alcanzado la pubertad. En algunos casos, este interés es exclusivo, es decir, la persona experimenta motivación sexual solo hacia niños sin signos de desarrollo puberal y no tiene atracción sexual hacia personas sexualmente maduras [13].

Figura 2. Vestimenta de personas en diferentes etapas de Tanner (Clothed pictures of people at different Tanner stages from the Virtual People Set [15]).

Usando la escala de Tanner para valorar la maduración sexual a través del desarrollo físico de personas como se indica en Figura 2, se observa que la prepubescencia generalmente se define como el puntaje más bajo en la escala de maduración física, al ser esta la fase en la que la persona no encuentra aún definida o desarrollada ninguna característica sexual física es la que más muestra atención a la persona con Pedofilia como se dijo antes [14].

Es así que podemos definir a la pedofilia como una atracción sexual hacia los niños prepúberes, indicando características como pensamientos sexuales, fantasías, impulsos, excitación o comportamiento persistentes y recurrentes hacia ellos.

- **Abuso sexual infantil**

En la literatura existe una variedad de definiciones que explican el abuso sexual infantil. Sin embargo, para poder crear una definición consolidada que se pueda reconocer de manera global la agencia de aplicación de la ley internacional de la Unión Europea, define al abuso sexual infantil como cualquier abuso sexual de una persona menor de 18 años incluida la fabricación y distribución de imágenes que representan dicho abuso [16].

Si de manera clara esta definición consolidada globalmente integra características específicas como la pornografía infantil, una crítica a esta definición se la puede basar en su simplicidad, ya que esta definición de abuso sexual infantil se centra principalmente en

el concepto de abuso, un punto clave que puede o no ser aplicable a la explotación sexual. Sin embargo, la definición no incorpora aspectos importantes de la explotación sexual, como el concepto de consentimiento verdadero, el engaño y la manipulación, que son las características del delito y el papel principal que el delincuente tiene en la dinámica de la relación. La irregularidad en la edad del consentimiento sexual en los países europeos pudo ser un posible factor contribuyente a la elección simplista de esta definición por parte de la EUROPOL [17].

En complemento en la literatura de Daigneault et al. [18], se decidió aplicar una definición propia de "abuso sexual" a su estudio, en lugar de restringirla a la de EUROPOL. Su definición establece que el abuso sexual es cualquier gesto de naturaleza sexual, con o sin contacto físico, cometido por un individuo sin el consentimiento de la persona o, en algunos casos, particularmente el de los niños, a través de la manipulación emocional o el chantaje [18]. Esta es una definición más clara y específica, que se usará en el presente trabajo y que se relaciona con el Grooming.

- **Delincuentes de Internet**

En la literatura de Aslan [19] se publicó una revisión de varias literaturas relacionadas, donde se identifican diferentes tipologías de delincuentes sexuales en Internet. Allí se puede ver claramente que se toma en primer plano como delincuentes en internet a aquellos que ejecutan los delitos de pornografía infantil, sin embargo, algunos investigadores han distinguido entre la pornografía infantil y otros comportamientos de explotación sexual infantil en línea. Estas tipologías se presentan en Anexo 1 [19].

En Anexo 1 [19] los autores hablan de términos específicos y únicos para delincuentes de internet. Por las diferencias presentadas en los comportamientos, la motivación del delito, el uso que se le da a la información obtenida del menor e incluso el uso del material pornográfico infantil obtenido por cualquiera de los medios, llevan a deducir que los Groomers o acicaladores sexuales pueden ser una población infractora distinta. Por ejemplo, se ha encontrado que los Groomer presentan menos interés pedófilo, pero un mayor uso de pornografía infantil y mayores niveles de enfoque sexual que los delincuentes sexuales de contacto o, como se los conoce en varias literaturas, viajeros, que conocen a su víctima fuera de línea [20].

Con respecto a las diferencias en comportamientos, se ha descubierto que una muestra de Groomers tiene menos probabilidades de tener antecedentes penales, incluido el abuso de sustancias, que los delincuentes sexuales infantiles que no son de Internet [21].

Existen varios comportamientos similares entre todo este tipo de delincuentes de internet más en los estudios de la literatura se puede observar que estos se centran únicamente en delitos relacionados con pornografía infantil y como se habló anteriormente el tipo de delincuente conocido como Groomer no necesariamente hace uso de este tipo de material para efectuar el delito. En consecuencia, podemos decir que existe una escasez de literatura que evalúa el fenómeno del Grooming o acicalamiento en línea.

2.3. Sistema de Control Parental

Para el presente trabajo se usará la siguiente definición de control parental [32]:

> *"Un sistema de control parental es una herramienta que permite a los padres controlar y/o limitar el contenido a los que sus hijos puedan acceder a internet desde sus dispositivos, ya sean ordenadores, móviles o tabletas.".*

Los padres utilizan las aplicaciones de control parental para realizar un seguimiento de las actividades de sus hijos en Internet. Por lo general, el padre instala la aplicación en el dispositivo electrónico del niño y luego puede configurar un conjunto de reglas para dictar lo que su hijo puede y no puede hacer con su dispositivo.

2.3.1. Privacidad en Control Parental

Uno de los principales temas tratados cuando se habla de controles parentales es la privacidad. Llevando a un debate ético sobre el control de los padres sobre los datos privados de sus hijos. Algunas personas dicen que los padres deben tener control total sobre la información de uso de los dispositivos tecnológicos de sus hijos. Otros sostienen que la mejor manera de cubrir este problema es instruir a los niños sobre los riesgos de Internet y la tecnología, y dejar que aprendan a comportarse de manera segura por sí mismos.

Westin definió la privacidad como "El derecho del individuo a decidir qué información sobre sí mismo debe comunicarse a los demás y bajo qué circunstancias" [33], en otras palabras, esto significa tener el control sobre la información de cada uno.

Los controles parentales manejan y procesan los datos de actividad, información, mensajes u otros de un menor, para poder controlar o limitar el contenido al que accede el menor en internet. Al realizar esta manipulación sin autorización del propietario de esta información (el menor), este proceso, puede llegar a considerarse como una violación de su privacidad

En contraste a esto, una de las justificaciones presentadas a este tipo de violaciones a la privacidad es el temor que tiene el tutor ya que los niños pueden acceder a una gran cantidad de contenido inapropiado en Internet, como violencia, armas, imágenes sexuales, lenguaje fuerte, drogas e incluso por la posibilidad de que sus hijos participen en relaciones con extraños de Internet, que pueden ser personas peligrosas que intentan aprovecharse de los niños, es decir sean víctimas de Grooming,

Otorgándosele a estas aplicaciones la autorización a la información del menor por parte del tutor. Esto nos lleva a decir que la privacidad del menor quedaría ubicada en un segundo plano si el tutor considera que la seguridad en línea del menor es más importante.

2.4. Clasificadores Lineales

La clasificación estadística o clasificación lineal es el enfoque amplio de aprendizaje supervisado que capacita a un programa para categorizar información nueva y no etiquetada en función de su relevancia para datos conocidos y etiquetados, debido a estas características es ampliamente usado en la clasificación de texto. El proceso de entrenamiento de un modelo de inteligencia artificial que clasifica texto dentro de un grupo basado en sus características es el que se puede observar en Figura 3.

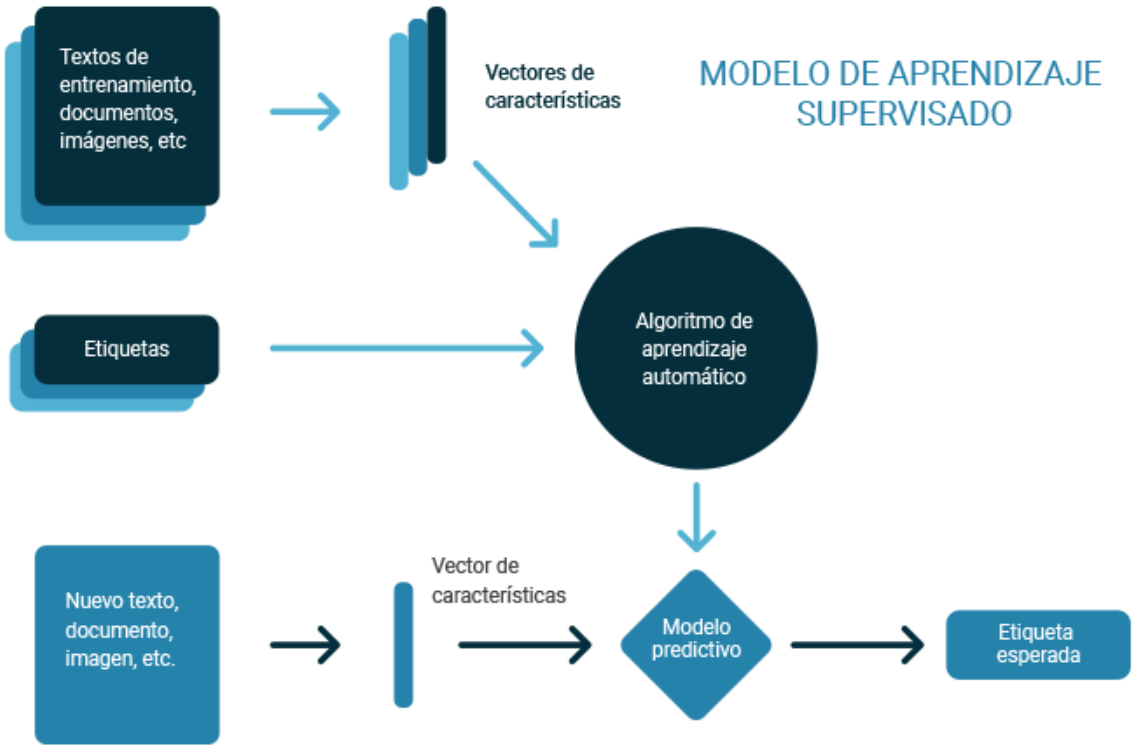


Figura 3. Diagrama de flujo del aprendizaje supervisado

2.5. Herramientas de desarrollo

A continuación, en Tabla 1, se describen la lista de herramientas utilizadas en el trabajo desarrollado.

Nombre	Descripción	Utilizado en
LucidChart	Herramienta de diagramación que permite trabajar de manera colaborativa en tiempo real	Diagrama de arquitectura y flujos de todo el sistema
API de Extensiones de Google Chrome	Conjunto de funciones, objetos y librerías que permiten manipular la información en una página que se ejecuta sobre Google Chrome.	Interceptor de mensajes
Python	Lenguaje de programación interpretado que permite el desarrollo de sistemas rápidamente.	Servidor y aplicación de escritorio.
Flask	Micro framework desarrollado en Python que permite el desarrollo rápido de aplicaciones web.	Aplicación de escritorio, servidor.
	Biblioteca desarrollada en Python con uso especial en machine learning debido a su gran gamma de algoritmos y módulos de inteligencia artificial.	Servidor.
RabbitMQ	Message Broker que permite la comunicación asíncrona a través de cola de mensajes.	Comunicación entre servidor y aplicación de escritorio.

 Visual Studio Code	Editor de código, desarrollado por Microsoft.	Interceptor de mensajes, servidor y aplicación de escritorio.
 Git	Sistema de control de versiones de código.	Interceptor de mensajes, servidor y aplicación de escritorio.
 Gitlab	Servicio web para el control de versiones de código, Gestor centralizado de repositorios digitales.	Interceptor de mensajes, servidor y aplicación de escritorio.
 Postman	Herramienta que permite probar servicios web.	Servidor, y aplicación de escritorio.

Tabla 1. Herramientas de desarrollo.

3. METODOLOGÍA

En este capítulo se detallan los artefactos, eventos, roles y procesos usados en la ejecución de este proyecto, de acuerdo con la propuesta metodológica de Scrum, descrita en la planificación de este trabajo.

Por lo tanto, de acuerdo con Scrum, se procedió a identificar los respectivos roles de las personas involucradas en el proyecto. Entonces, considerando esto, se procedió a identificar los respectivos roles de las personas involucradas en el proyecto, los cuales se muestran en Tabla 2.

Roles de Scrum	
Scrum Master	Pamela Flores
Product Owner	Patricio Zambrano
Scrum Team	Christian Oña
	Jairo Proaño

Tabla 2. Roles de Scrum.

3.1. Definición de requerimientos

Para el desarrollo de este proyecto se realizó una primera reunión con el Product Owner donde se manifestó la necesidad de crear un prototipo de control parental que demuestre la viabilidad de implementar un sistema inteligente capaz de clasificar conversaciones dentro de las estaciones estandarizadas en el ciclo de vida del Grooming.
Para el desarrollo del proyecto, se usa como fuente principal la investigación de Zambrano et al. [1], que otorga las pautas más importantes para implementar el clasificador de Grooming, motor del sistema de control parental a desarrollar.
Contemplando la definición de un control parental, y tomando en cuenta los requerimientos expresados por el Product Owner en la reunión, la Tabla 3 de requerimientos mínimos queda de la siguiente forma.

ID	REQUERIMIENTO
RF01	El sistema será capaz de leer los mensajes que el tutelado reciba en su cuenta de Facebook.
RF02	El sistema será capaz de ubicar los mensajes dentro de una de las 6 estaciones del ciclo de vida del Grooming.
RF03	El sistema será capaz de bloquear la interacción que el tutelado tiene en Facebook, cuando se haya detectado una conversación maliciosa.
RF04	El sistema notificará a los tutores del menor cuando una incidencia sea detectada.
RF05	El sistema permitirá al tutor ingresar al sistema por medio de un usuario y contraseña.
RF06	El sistema permitirá al tutor cambiar el usuario y contraseña.
RF07	El sistema permitirá que el tutor revise las incidencias de ataque de grooming de su tutelado.
RF08	El sistema permitirá que el tutor configure las fases desde las que se considere maliciosas.
RF09	El sistema permitirá que el tutor desbloquee la cuenta del tutelado.

Tabla 3. Requerimientos funcionales.

3.2. Historias épicas de usuario

A continuación, en la Tabla 4 se presentan las épicas de usuario formadas en base a los requerimientos mínimos antes mencionados. El detalle de las épicas de usuario se las presenta en el Anexo I.

ID	Épica de usuario
US01	Como tutor de un menor requiero un sistema capaz de identificar cuando mi tutelado está siendo acosado por medio de prácticas de Grooming para

	ayudar a reducir el peligro al que está expuesto al usar Facebook
US02	Como tutor requiero un sistema que sea capaz de bloquear la interacción que mi tutelado tiene con Facebook cuando se ha detectado que es víctima de Grooming, para así evitar consecuencias más graves y yo poder tomar las acciones mientras tanto
US03	Como tutor requiero un sistema capaz de notificarme cuando mi tutelado ha sido expuesto a un ataque de Grooming, para poder tomar acciones al respecto
US04	Como tutor requiero que el sistema me permita configurar las fases que considero son inadecuadas, y así ser tener más control sobre lo que mi tutelado puede ver en redes sociales.
US05	Como tutor requiero una aplicación que me permita instalar el sistema de control parental en la computadora de mi tutelado, y así poder ejecutar el control parental.
US06	Como tutor requiero que el sistema registre los ataques de Grooming detectados para de esta forma monitorear las acciones del tutelado.
US07	Como tutor requiero que el sistema implemente un inicio de sesión con usuario y contraseña, para que la única persona que pueda administrar la cuenta sea yo.
US08	Como tutor requiero que el sistema me permita gestionar la cuenta para así desbloquear la computadora de mi tutelado cuando haya definido que está libre de peligro.

Tabla 4. Épicas de usuario.

3.3. Product Backlog

Una vez identificados las épicas de usuario se ha procedido a armar el Product Backlog inicial, en donde se expone la historia de usuario, la estimación en días y la prioridad. La valoración de la prioridad se ha definido de acuerdo con Tabla 5.

VALORACIÓN	PRIORIDAD DEL NEGOCIO

1	Baja
2	Mediana
3	Alta
4	Muy alta

Tabla 5. Valoración de prioridad.

Entonces, el Product Backlog inicial queda de acuerdo con la siguiente tabla:

PRODUCT BACKLOG				
ÉPICA DE USUARIO	ID	HISTORIA DE USUARIO	ESTIMACIÓN (días)	PRIORIDAD
US01	US01-01	Extracción de mensaje recibido para cada chat	5	4
	US01-02	Limpieza de mensajes	1	4
	US01-03	Clasificador de Grooming	7	4
US02	US02-01	Bloqueador de dominio	1	4
US03	US03-01	Alerta de incidencia	3	3
	US03-02	Notificación de incidencia	4	3
US04	US04-01	Configuración de estado aceptable	2	3
US05	US05-01	Despliegue de sistema	1	4
US06	US06-01	Historial de incidentes	1	2

US07	US07-01	Inicio de sesión de usuario	1	3
	US07-02	Gestión de usuario y contraseña	1	3
US08	US08-01	Desbloqueo de cuenta	1	3

Tabla 6. Product backlog inicial.

3.4. Planificación del Lanzamiento

En la planificación de este proyecto se definió previamente una duración de desarrollo de 8 semanas, divididas en 3 Sprints de 2 semanas de duración por cada Sprint, pero debido a requerimientos nuevos, actividades extras y después de un análisis realizado por el equipo, se determinó establecer 4 Sprints de 2 semanas cada uno; los cuales han sido definidos como se muestra a continuación.

HISTORIA DE USUARIO	**ESTIMACIÓN**
SPRINT 1	
US01-01	5
US01-02	1
TOTAL	6
SPRINT 2	
US01-03	5
US03-01	3
TOTAL	8
SPRINT 3	
US02-01	1
US07-01	1
US07-02	1
US04-01	2
US06-01	1
TOTAL	6
SPRINT 4	
US03-02	4
US08-01	1

US05-01	2
TOTAL	7

Tabla 7. Planificación del Release.

Como se puede apreciar, se han definido Sprints con una duración promedio de 6 días cada Sprint, a excepción del segundo Sprint en donde la incertidumbre es mayor debido a que el clasificador es el motor de todo el proyecto. Cada día de estimación implica un promedio de 4 horas de trabajo para la ejecución de las tareas.

A demás se hace uso de un "Sprint 0", denominado así por el equipo debido a que es un espacio que se uso para planificar las actividades previas al desarrollo de los demás Sprints. Este "Sprint 0" realmente no consta en la documentación de Scrum, ya que no generara entregregables al finalizar las actividades, pero es importante porque aquí es donde se definen las herramientas a usar, se prepara el ambiente de desarrollo, la arquitectura base del sistema, y el equipo se documenta con la información necesaria para continuar. El equipo definió una duración de 1 semana para la ejecución de este Sprint, el cual no tendrá historias de usuario ni entregables.

Por último, en la retrospectiva de cada Sprint, se multiplicó por 4 los días totales estimados por cada historia de usuario, esto con el objetivo de que la gráfica presentada muestre mejor la distribución de trabajo a lo largo de los días.

3.5. Sprint 0

3.5.1. Objetivos

Preparar al equipo y ambiente de desarrollo

- Documentar al equipo con los recursos necesarios acerca del tema a desarrollar.
- Diseñar la arquitectura base del sistema.
- Instalar las herramientas necesarias para el desarrollo del sistema.
- Crear los repositorios para los componentes del sistema.

3.5.2. Ejecución del Sprint 0

En la ejecución de este Sprint, se pueden mencionar los siguientes puntos como los único ejecutados:

- **Documentación**

En esta fase se ejecutaron actividades de lectura para entender todo lo que se necesitó para desarrollar el proyecto:

- En la arquitectura que se detalla más adelante, se requiere desarrollar una extensión de Google Chrome, por lo que, accedimos a la documentación oficial de Chrome para desarrolladores.
- El trabajo se desarrolla tomando como base el modelo de clasificación explicado en el trabajo de Zambrano et al. [1]. Así que, el equipo procedió a entender y hacer pruebas con esa información.

- **Diseño arquitectura de la aplicación**

El prototipo está formado por 3 componentes principales y por 2 componentes de apoyo para el correcto funcionamiento de este, tal y como se muestra en Figura 4.

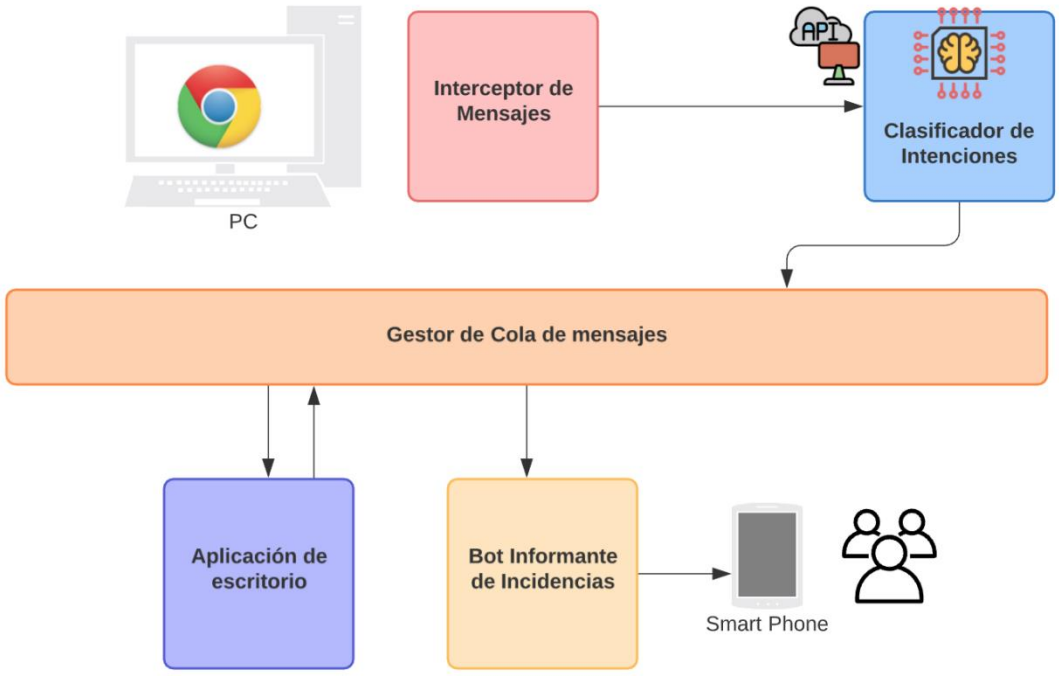


Figura 4. Arquitectura de componentes del sistema.

Estos componentes son:

- **Interceptor de mensajes**

Este componente se encarga de leer constantemente los mensajes del usuario, y de transmitirlos al servidor.

- **Servidor:** Este componente está integrado de algunos módulos que cumplen las siguientes funciones:

    - Limpiar los mensajes que se receptará del componente interceptor.
    - Clasificar los mensajes dentro de las fases del Grooming estandarizadas en el trabajo de investigación de Zambrano et al. [1]
    - Alertar a la aplicación de escritorio con la clasificación obtenida por mensaje.
    - Establecer la comunicación respectiva con el bot de Telegram.
    - Registrar los usuarios de Telegram.

- **Aplicación de escritorio:**

Este componente es un programa que se ejecutará en el computador que está usando el menor. Su principal función es almacenar las configuraciones del tutor. Y actuar con las clasificaciones enviadas por servidor.
Cuando la frase clasificada sea considerada como peligrosa, es decir se presente una incidencia, inmediatamente se bloqueará el acceso al sitio web Facebook en la computadora que el tutelado está usando, y los respectivos tutores serán notificados.

- **Gestor de Colas:**

Es un componente auxiliar, su principal trabajo es el de asegurarse que los componentes del sistema reciban los mensajes. Esto lo logrará gracias a la gestión de colas de mensajes que maneja.

- **Bot:**

Componente externo al sistema. El sistema se conectará a este bot y será usando principalmente para notificar a los tutores por medio de Telegram de los incidentes presentados.

- **Ambiente de trabajo**

El equipo de trabajo está formado por 2 integrantes. Cada integrante cuenta con un computador. El editor de código seleccionado es Visual Studio Code, debido a flexibilidad del editor de adaptarse a cualquier lenguaje de programación, y todas las herramientas extras que nos ofrece para realizar pruebas y depuración.
El lenguaje de programación seleccionado fue Python, debido a la afinidad que tiene el equipo de desarrollado con ese lenguaje de programación, la robustez de la plataforma, y la versatilidad, ya que se puede usar tanto en ambientes web, de escritorio, inteligencia artificial, etc.

- **Control de versiones y desarrollo de software colaborativo Gitlab.**

Para alojar el proyecto se usa Gitlab, un servicio centralizado de repositorios basado en git. Cuando se usa Gitlab se tiene ventajas respecto a otros servicios similares, la primera es que nos permite crear grupos ya sean privados o públicos de forma gratuita, los grupos los podemos usar para asociar varios repositorios bajo un mismo lugar, además de esto, de ser requerido se puede usar herramientas de DevOps fácilmente. Entonces, de acuerdo a la arquitectura anteriormente presentada, se crea un grupo de Gitlab el que contiene 3 repositorios que corresponden a cada uno de los componentes del sistema componente del sistema.

Nombre	Descripción
fb-message-interceptor	Repositorio que contiene el código del componente que interceptará los mensajes de Facebook
ParentIAl	Repositorio que contiene el código del componente que cumplirá la función de servidor central del sistema
parential-desktop	Repositorio que contiene el código de la aplicación de escritorio.

Tabla 8. División de repositorio de Git.

3.5.3. Revisión del Sprint

Al finalizar este Sprint se concluye que se alcanzaron exitosamente los objetivos. La arquitectura se ha definido, los repositorios han sido creados, las herramientas establecidas, y los proyectos de cada componente iniciados. Todo esto permite una mayor fluidez al continuar con los siguientes Sprints.

3.6. Sprint 1

3.6.1. Objetivo

- Desarrollar una extensión de Google Chrome que sea capaz de leer los mensajes de los chats abiertos del tutelado.
- Desarrollar un API en el servidor para procesar los mensajes.

3.6.2. Historias de Usuario

En Tabla 9, se muestran las historias de usuario que forman parte de este Sprint. El detalle completo de las historias de usuario del Sprint 1 se presentan en Anexo III.

ID	TÍTULO	ESTIMACIÓN	ESTADO
US01-01	Extracción de mensaje recibido para cada chat	5	Por implementar
US01-02	API para limpieza de mensajes	1	Por implementar

Tabla 9. Historias de usuario para Sprint 1.

3.6.3. Sprint Backlog

HISTORIA DE USUARIO	TAREAS
US01-01	Implementar la estructura de la extensión de Google Chrome de acuerdo con la documentación oficial.
	Identificar las etiquetas HTML que encapsulan el contenido de interés

	Desarrollar las funciones que extraerán los mensajes que nos interesan
US01-02	Identificar del bloque HTML que recibiremos los mensajes que pertenecen al emisor del mensaje
	Implementar un API que procese la información proporcionada por el Interceptor.
	Extraer los mensajes del emisor y generar un objeto que contenga id del emisor y el mensaje respectivo

Tabla 10. Sprint Backlog para Sprint 1.

3.6.4. Ejecución del Sprint

En la planificación del proyecto de titulación se analizó el orden de ejecución de los Sprints. En esta fase de planificación se optó primero por buscar una forma de extraer los mensajes de las conversaciones que el tutelado tiene, se llegó a esta conclusión debido a la incertidumbre que existía al momento de tomar estos datos, y porque es muy importante para la ejecución del sistema primero poseer esta la información.
Una vez iniciado el Sprint en la fase de análisis se presentaron algunas opciones, para la ejecución de este proyecto integrador se optó por el siguiente método:
Se desarrollará una extensión para el navegador, la extensión constantemente leerá los mensajes de las conversaciones abiertas que tiene el tutelado.

- **Análisis**

El análisis para la obtención de los mensajes de las conversaciones de Facebook fue el siguiente:

El primer paso fue identificar todo un bloque de mensajes, tanto de receptor (Tutelado), como de emisor (Posible atacante), es decir un chat completo, se encontró que este bloque estaba bajo una etiqueta HTML “div” que tenía la clase “fbNubFlyoutOuter”, un bloque de este código se muestra en Figura 5.

```
alignh="right">…</a>
▼<div class="fbNubFlyout uiToggleFlyout" role="dialog"
aria-label="Keyboard Shortcut Help">
  ▶<div class="fbNubFlyoutOuter">…</div>
  </div>
</div>
<div id="u_0_2s"></div>
```

Figura 5. Bloque de mensajes.

Con esta información, el siguiente paso fue identificar los mensajes pertenecientes únicamente al emisor. Se detectó que Facebook encapsula el texto de cada mensaje bajo una etiqueta que contiene la clase de estilo CSS llamada "_5yl5 ", más los mensajes bajo esta etiqueta pertenecen a ambas partes del chat (Emisor-Receptor), un bloque de este código se muestra en Figura 6.

```
▼<div class="_4gx_">
  ▼<div class="_1aa6">
    ▼<div class>
      ▼<span class="_5yl5">
          <span>Hello</span> == $0
        </span>
      </div>
      ::after
    </div>
```

Figura 6. Espacio que contiene el texto del emisor.

Continuando con el proceso de identificación se detectó que los mensajes del remitente se encapsulan bajo un cuadro con las clases "_5wd4 _1nc7". Con esas dos condiciones encontradas se logró identificar los mensajes del remitente, un bloque de este código se muestra en Figura 7.

```
▶<div class="_4tdt _ua0">…</div>
▼<div class="_4tdt _ua1">
  ▶<div class="_31o4">…</div>
  ▼<div class="_ua2">
    ▼<div class="_4tdv">
      ▶<div class="_5wd4 _1nc7 _2cnu">…</div> == $
      ▶<div class="_5wd4 _1nc7 _2cnu">…</div>
      </div>
    </div>
  </div>
▶<div class="_4tdt _ua0">…</div>
```

Figura 7. Espacio de mensajes de emisor.

También se identificó que el nombre del emisor se ubicaba bajo la etiqueta "_2mgq" dentro del mismo bloque de mensaje, un bloque de este código se muestra en Figura 8.

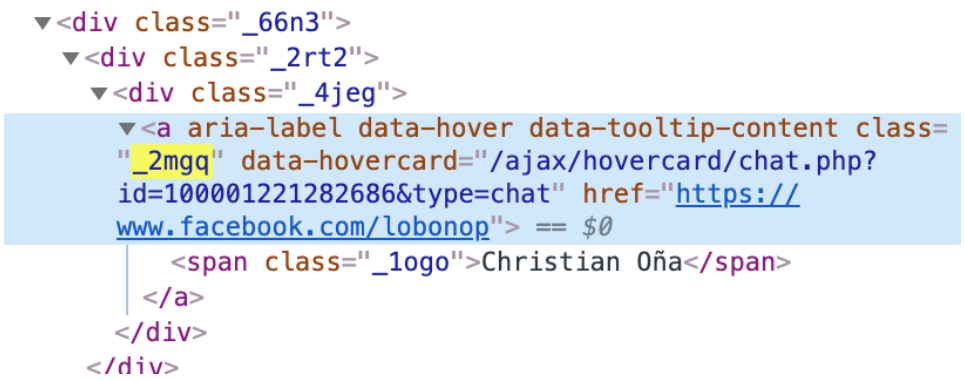


Figura 8. Espacio del nombre del emisor.

Con esto se finaliza la fase de análisis del Sprint, se había identificado las diferentes etiquetas HTML y estilos que contiene la información que requerimos para continuar.

- **Implementación**

Para obtener esta información, fue necesario desarrollar un script que limpie los mensajes automáticamente considerando el análisis previo. El script fue desarrollado en el componente del servidor debido a la carga y complejidad del proceso de limpieza. Entonces se divide el flujo para la limpieza de estos mensajes en dos partes:

La primera parte corresponde a la extensión de Google Chrome, esta extensión se ejecuta en el navegador del tutelado, y extrae el bloque de mensajes en HTML, después mediante de una petición HTTP se transmite a un API que se ejecuta en el servidor.
La segunda parte corresponde a esta API, aquí se realiza realmente la limpieza de los mensajes usando el análisis comentado en la sección previa.
Cuando la limpieza es realizada, el resultado se pasa al motor del sistema, que actualmente corresponde al clasificador de Grooming y que será desarrollado en próximos Sprints.
A continuación, en Figura 9 se muestra el flujo resultante, que se aplica para la limpieza de los mensajes.

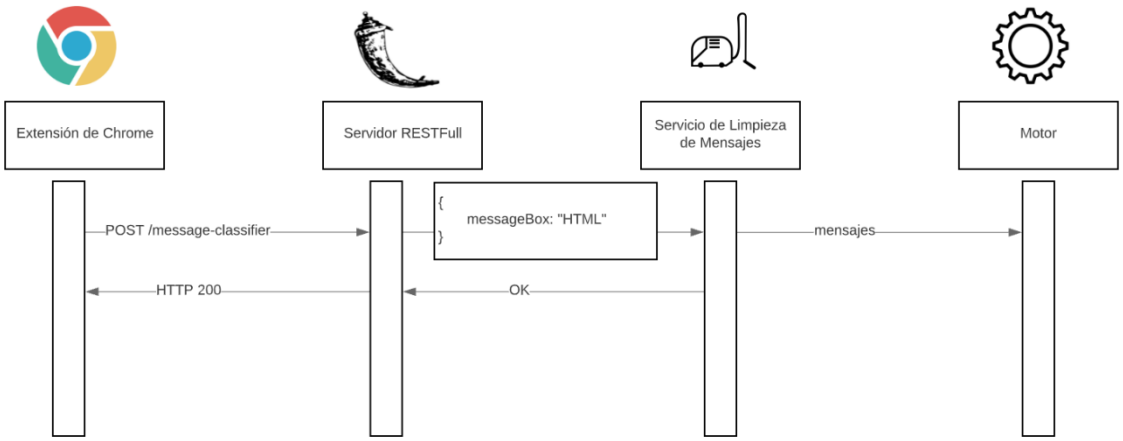


Figura 9. Flujo de limpieza de mensajes

- **Cuerpo de petición y respuesta**

La comunicación entre los componentes se realiza por medio de peticiones HTTP. Por lo tanto, se establecen estándares para las peticiones y respuestas que sirven para una comunicación más simple entre los componentes.
Cuando se transmite el bloque de mensajes desde el componente Interceptor, se realiza una petición *POST / HTTP/1.1*, el cuerpo JSON formado tiene el siguiente formato.

- messageBox: Campo obligatorio

Por el otro lado se espera una respuesta *HTTP/1.1 200 OK* con contenido vacío, ya que al interceptor realmente no le interesa un resultado.

3.6.5. Reunión Diaria

En la ejecución de este Sprint se realizaron 8 reuniones diarias, de aproximadamente 10 minutos cada una. Las primeras 3 reuniones sirvieron para identificar problemas en la investigación. Las siguientes reuniones fueron usadas principalmente para mantener actualizado al equipo.

3.6.6. Revisión del Sprint

El objetivo del primer Sprint se lo alcanzo con éxito. El principal reto en este Sprint fue encontrar una forma factible y óptima de extraer los mensajes para poder avanzar con los Sprints posteriores.

3.6.7. Pruebas de Aceptación

HISTORIA DE USUARIO	CRITERIO DE ACEPTACIÓN	CUMPLIDO
US01-01	Se tendrá una extensión que se podrá instalar en un navegador de Google Chrome	SI
	La extensión de Google Chrome leerá cada 10 segundos los bloques de mensajes de cada conversación abierta del tutelado	SI
	Desde la extensión de Chrome se emitirá una petición HTTP POST por cada chat abierto del tutelado, cuyo cuerpo tendrá el bloque de mensajes de cada chat.	SI
US01-01	Existirá un API Rest que tome un bloque de mensajes el cual estará en un formato HTML y obtendrá los mensajes en un formato manejable por la aplicación.	SI

Tabla 11. Pruebas de aceptación Sprint 1.

3.6.8. Adaptación del Product Backlog

Una vez finalizado el primer Sprint se realizó un análisis por el equipo que permitió llegar a la conclusión de que es necesario que se maneje un identificador para cada instalación de la aplicación es decir un ID para cada instancia de instalación. Este ID de instalación ayudará a identificar a un canal que se usará para la transferencia de mensajes a una instancia específica.

Ya que previo a la implementación de este ID es necesario realizar las tareas del Sprint 3, se agrega una historia de usuario adicional a este Sprint.

El proceso de generación de este ID será el siguiente:

- Cuando se ejecute por primera vez el servicio sobre el sistema del tutelado, se generará un ID único para esa instalación, este ID estará disponible para la extensión de Google Chrome por medio de un API REST que se ejecutará en un servidor HTTP local.

PRODUCT BACKLOG				
ÉPICA DE USUARIO	ID	HISTORIA DE USUARIO	ESTIMACIÓN (días)	ESTADO

US01	US01-01	Extracción de mensaje recibido para cada chat	5	Terminado
	US01-02	Limpieza de mensajes	1	Terminado
	US01-03	Clasificador de Grooming	7	Por implementar
US02	US02-01	Bloqueador de dominio	1	Por implementar
	US02-02	Generador de ID	1	Por implementar
US03	US03-01	Alerta de incidencia	3	Por implementar
	US03-02	Notificación de incidencia	4	Por implementar
US04	US04-01	Configuración de estado aceptable	2	Por implementar
US05	US05-01	Despliegue de sistema	1	Por implementar
US06	US06-01	Historial de incidentes	1	Por implementar
US07	US07-01	Inicio de sesión de usuario	1	Por implementar
	US07-02	Gestión de usuario y contraseña	1	Por implementar

US08	US08-01	Desbloqueo de cuenta	1	Por implementar

Tabla 12. Product Backlog del Sprint 1.

3.6.9. Retrospectiva del Sprint

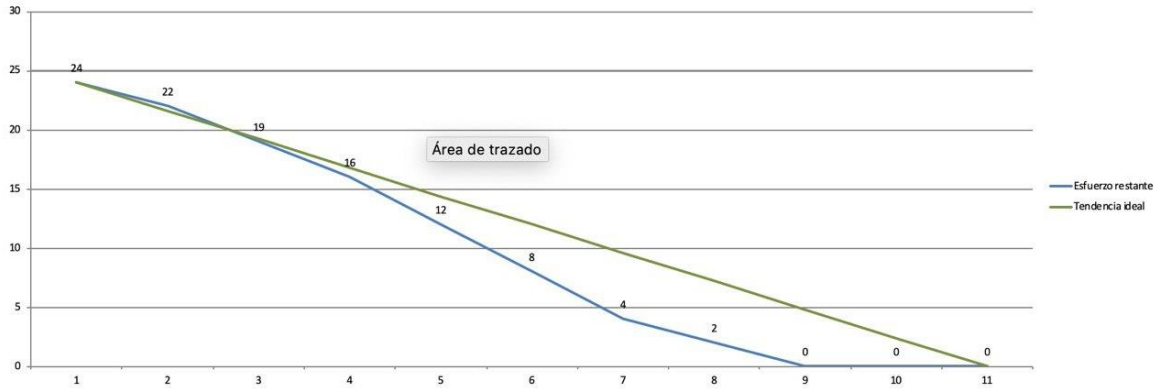


Figura 10. Gráfico Burndown del Sprint 1

3.7. Sprint 2

3.7.1. Objetivo

Desarrollar un modelo inteligente que sea capaz de clasificar los mensajes recibidos y ubicarlos dentro de las fases del ciclo de vida del Grooming.

3.7.2. Historias de Usuario

En Tabla 13, se muestran las historias de usuario que forman parte de este Sprint. El detalle completo de las historias de usuario del Sprint 2 se presentan en Anexo IV.

ID	TÍTULO	ESTIMACIÓN	ESTADO
US01-03	Clasificador de Grooming	5	Por implementar
US03-01	Alerta de incidencia	3	Por implementar

Tabla 13. Historias de usuario del Sprint 2.

3.7.3. Sprint Backlog

HISTORIA DE USUARIO	TAREAS
US01-03	Elegir un método para el entrenamiento de un clasificador de Grooming
	Entrenar un modelo de clasificación lineal que permita predecir la fase de un mensaje de texto
	Probar el modelo de entrenamiento con un set de datos 70-30
	Generar un modelo de datos único para las predicciones
US03-01	Integrar el API de limpieza de datos, con el módulo de predicción de fases
	Implementar un bus de cola de mensajes (Message Broker) por donde se transmita los resultados de las fases clasificadas
	Implementar una comunicación entre el clasificador de Grooming y el Message Bróker.

Tabla 14. Sprint Backlog del Sprint 2.

3.7.4. Ejecución del Sprint

- **Análisis**

Para proceder con el Sprint 2 se usa el documento de salida del trabajo de investigación de Zambrano et al. [1] que se muestra en Figura 9 en la sección del Marco teórico. El documento de salida de esta investigación es un archivo CSV que fue provisto por el Product Owner, en donde se detallan dos columnas importantes para el presente trabajo, **message_text** y **message_stage**, estas columnas contienen los mensajes y su respectiva clasificación, esto se explica de una manera más detallada en el ciclo de vida del Grooming ubicado en el Marco teórico de este trabajo.

Después de un análisis de estos datos se decide que es ideal usar este archivo para entrenar un clasificador lineal que pueda ser puesto en un ambiente de producción. Para lo que se realizó una prueba con la que se obtuvo una precisión del 94% para este prototipo. Si bien no se alcanza el 97.61% que se detalla en el experimento de la

investigación de Zambrano et al. [1], se llegó a la conclusión que es un porcentaje aceptable para continuar con el desarrollo de este proyecto.

- **Clasificador de Grooming**

En la implementación del clasificador de Grooming se usa la librería llamada Scikit Learn, esta librería nos provee de los módulos, clases y funciones necesarios para desarrollar el modelo de clasificación. Los módulos que se usan son:

- TfidfVectorizer: Convierte una colección de documentos a una matriz de características TF-IDF.
- train_test_split: Separa los arreglos o matrices en subconjuntos aleatorios de los datos que lo conforman.
- LinearSVC: Clase que implementa el algoritmo Linear Support Vector Classification y la que se usó para entrenar el modelo de clasificación lineal.

TF-IDF es una fórmula que ayuda a definir la importancia de palabras dentro de los documentos. El primer paso que se realizo fue, extraer los mensajes y fases clasificadas previamente en el archivo CSV.
A este conjunto de datos se los separó con una función en dos subconjuntos de datos 70-30, es decir, 70% para entrenamiento y 30% para pruebas. El modelo de clasificación lineal LinearSVC finalmente es entrenado con el subconjunto del 70%. Un bloque de este código se muestra en Figura 18.

```
self.model = LinearSVC()
X_train, X_test, y_train, y_test, indices_train, indices_test = \
    train_test_split(self.features, labels, data.index,
                     test_size=0.3, random_state=0)
self.model.fit(X_train, y_train)
```

Figura 11. Entrenamiento de clasificador.

- **Integración**

Una vez que el modelo de clasificación ha sido entrenado, se empezó a desarrollar las capas que cubrirán los demás requerimientos.
El primer paso fue desarrollar un módulo que permite manipular al clasificador. El bloque de código que se usa para predecir las fases de nuevos mensajes se muestra en Figura 12.

```
def predict(self, text):
    texts = [text]
    text_features = self.tfidf.transform(texts)
    predictions = self.model.predict(text_features)
    for text, predicted in zip(texts, predictions):
        return self.id_to_category[predicted]
```

Figura 12. Clasificador.

La predicción de los mensajes se realiza después de que estos son limpiados. Se realiza la clasificación e inmediatamente se alerta el resultado a los demás componentes del sistema.

Un gestor de colas de mensajes ayuda a que componentes que se han suscrito a una cola sean notificados cuando otro componente produce un mensaje. Esto lo podemos ver ejemplificado en la Figura 13. Una vez que el mensaje es entregado al nodo este se destruye.

Figura 13. Cola de mensajes [34].

Esta cola de mensajes nos ayudará a que la aplicación de escritorio, que se desarrollará en próximos Sprint, conozca el resultado de la clasificación de un mensaje producido por el clasificador. Relacionando los componentes del sistema que se desarrolla con los componentes que se muestran en Figura 14. el **producer** es el clasificador, el mensaje es la clasificación realizada, y el **consumer** es la aplicación de escritorio. En este proyecto se usará como gestor de la cola de mensajes a RabbitMQ.

El flujo para esta sección del proyecto quedaría como se muestra en Figura 19:

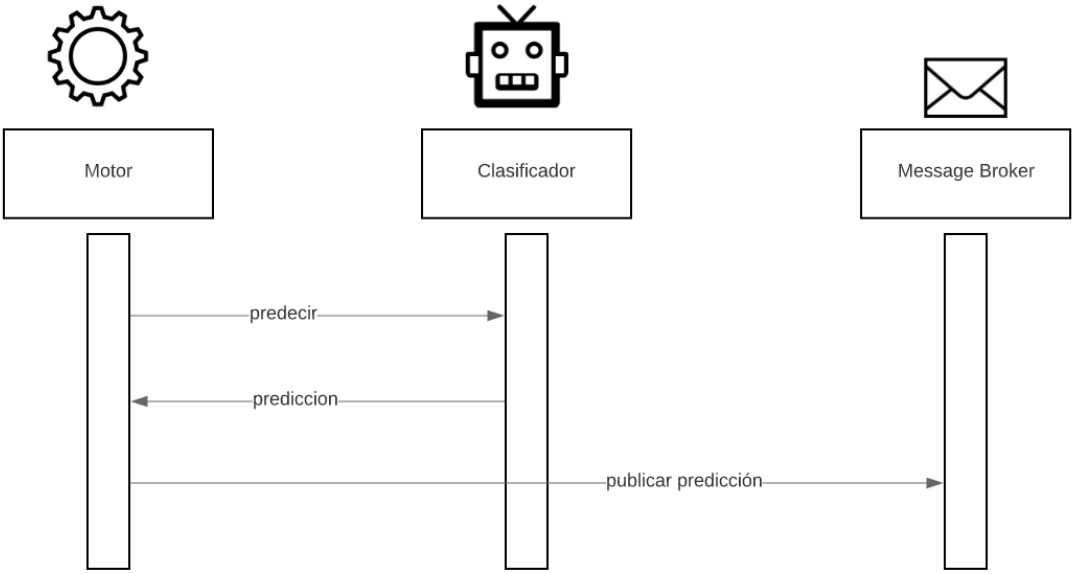


Figura 14. Arquitectura de módulo de clasificación

3.7.5. Reunión Diaria

En la ejecución de este Sprint se realizaron 9 reuniones diarias, de aproximadamente 10 minutos cada una. En las primeras reuniones se analizaron los avances en la investigación acerca clasificador y en las siguientes reuniones se revisaron los avances en la implementación del modelo y sus respectivas pruebas.

3.7.6. Revisión del Sprint

El objetivo del Sprint fue alcanzado con éxito. Hay que considerar que en este Sprint se desarrollaba el clasificador, que es el motor de todo el sistema. Por lo que era importante, la investigación e implementación adecuada para el modelo clasificador.

3.7.7. Pruebas de Aceptación

HISTORIA DE USUARIO	CRITERIO DE ACEPTACION	CUMPLIDO
US01-03	Se realizará una matriz de confusión con una precisión mayor al 90%	SI

US03-01	Existirá un método que retornará la predicción del clasificador	SI
	El sistema publicará el resultado en una cola de mensaje.	SI

Tabla 15. Pruebas de aceptación del Sprint 2.

3.7.8. Adaptación del Product Backlog

Una vez finalizado el segundo Sprint el equipo se reunió y llegó a la conclusión de que es necesario que los mensajes que son producidos por el clasificador sean publicados en una cola específica para cada ID de instancia.
Así que, posteriormente se cambiará la conexión al gestor de cola de mensajes. Cada cola será identificada por el ID de instancia que se implementará en el siguiente Sprint. Esto no afectará el desarrollo del siguiente Sprint ya que son cambios mínimos en el funcionamiento actual del sistema.

El Product Backlog una vez finalizada la adaptación queda como se muestra en la siguiente Tabla:

PRODUCT BACKLOG				
ÉPICA DE USUARIO	ID	HISTORIA DE USUARIO	ESTIMACIÓN (días)	ESTADO
US01	US01-01	Extracción de mensaje recibido para cada chat	5	Terminado
	US01-02	Limpieza de mensajes	1	Terminado
	US01-03	Clasificador de Grooming	7	Terminado
US02	US02-01	Bloqueador de dominio	1	Por

				implementar
	US02-02	Generador de ID	1	Por implementar
US03	US03-01	Alerta de incidencia	3	Por corregir.
	US03-02	Notificación de incidencia	4	Por implementar
US04	US04-01	Configuración de estado aceptable	2	Por implementar
US05	US05-01	Despliegue de sistema	1	Por implementar
US06	US06-01	Historial de incidentes	1	Por implementar
US07	US07-01	Inicio de sesión de usuario	1	Por implementar
	US07-02	Gestión de usuario y contraseña	1	Por implementar
US08	US08-01	Desbloqueo de cuenta	1	Por implementar

Tabla 16. Product Backlog del Sprint 2.

3.7.9. Retrospectiva del Sprint

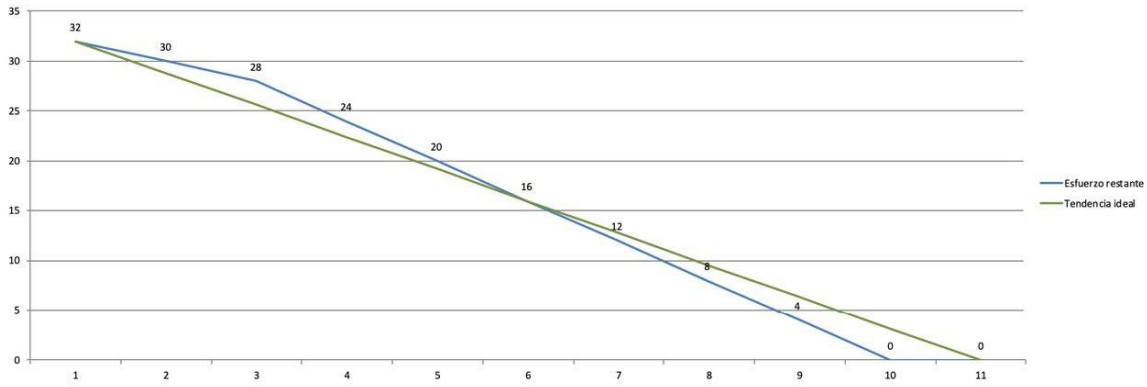


Figura 15. Gráfico Burndown del Sprint 2

3.8. Sprint 3

3.8.1. Objetivo

Implementar funcionalidades para la gestión de la aplicación.

- Implementar un inicio de sesión para la autenticación de los tutores.
- Implementar una sección para la gestión de usuario y contraseña.
- Implementar un panel para la administración de incidencias.
- Implementar un panel para la administración de la aplicación.

3.8.2. Historias de Usuario

En Tabla 17, se muestran las historias de usuario que forman parte de este Sprint. El detalle completo de las historias de usuario del Sprint 3 se presentan en Anexo V.

ID	TÍTULO	ESTIMACIÓN	ESTADO
US02-01	Bloqueador de dominio	1	Por implementar
US02-02	Generador de ID	1	Por implementar
US07-01	Inicio de sesión de usuario	1	Por

			implementar
US07-02	Gestión de usuario y contraseña	1	Por implementar
US04-01	Configuración de estado aceptable	2	Por implementar
US06-01	Historial de incidentes	1	Por implementar
US03-01	Alerta de incidencia	1	Por corregir

Tabla 17. Historias de usuario del Sprint 3.

3.8.3. Sprint Backlog

HISTORIA DE USUARIO	TAREAS
US02-01	Desarrollar una aplicación que corra sobre el ordenador del tutelado.
	Implementar un método para bloquear el acceso del tutelado cuando se presente un ataque de Grooming (Incidencia)
US02-02	Generar un ID único para cada aplicación de escritorio cuando se ejecute por primera vez el programa
	Desarrollar un API Rest que se ejecute en la aplicación de escritorio, la cual retorne el ID de la instalación
	Leer desde el API Rest de la aplicación de escritorio el ID de instalación y adjuntar en el cuerpo de la petición HTTP al clasificador.
US07-01	Desarrollar un servidor web que se levante sobre el ordenador del tutelado, cuando la aplicación de escritorio sea ejecutada.
	Implementar un método de autenticación por medio de usuario y contraseña para posteriormente consumir los demás servicios.
US07-02	Desarrollar un formulario para cambiar el usuario y contraseña del tutelado.
US06-01	Desarrollar una página principal para cuando el usuario este autenticado en la aplicación de escritorio
	Mostrar un historial de incidentes en la página principal del mentor

US04-01	Desarrollar una sección donde el mentor pueda configurar los estados aceptables del ciclo de vida de Grooming.
US03-01	Modificar la lectura del cuerpo de la petición del clasificador, para identificar el ID de instalación que se usará para la conexión a una cola de RabbitMQ.

Tabla 18. Sprint Backlog del Sprint 3.

3.8.4. Ejecución del Sprint

Este Sprint incluye tareas que quedaron pendientes en Sprint anteriores, estas tareas han sido agregadas para mejorar la comunicación entre los componentes. De los cambios realizados se puede comentar los siguientes puntos:

- **Generación de ID**

El nuevo cuerpo de la petición que se realizará al servidor ahora contiene un campo adicional, el campo tiene el nombre de ID, y lo podemos ver en Figura 16.

```
{
    "messageBox": "HTML",
    "ID": "ID"
}
```

Figura 16. Nuevo cuerpo de petición a servidor

Este ID estará formado por 20 caracteres numéricos, los primeros 4 dígitos se obtendrán de un número pseudoaleatorio, y los siguientes 16 dígitos se obtendrán de una marca de tiempo o TIMESTAMP del momento en el que se ejecutó por primera vez la aplicación de escritorio.

La extensión de Google Chrome obtendrá el ID de instancia cuando realice la primera transmisión al clasificador. El ID obtenido se almacenará en el navegador para que la consulta no se realice siempre, y se adjuntará en cada transmisión que se realice hacia el servidor. Para obtener este ID se desarrolló un API que retorna este ID y que está ejecutándose en la aplicación de escritorio. El flujo se modifica de acuerdo con Figura 17.

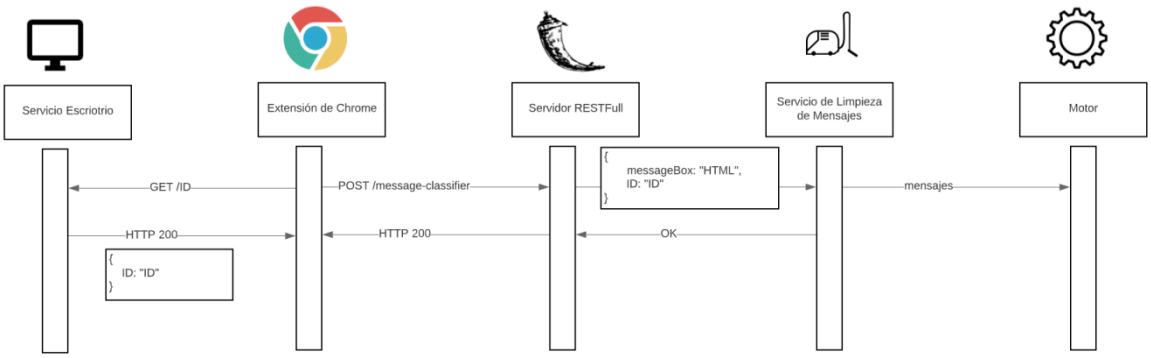


Figura 17. Flujo de mensajes, Aplicación de Escritorio, Extensión, Clasificador

Una vez que el servidor recepte la petición, encontraremos que el cuerpo incluye el ID de la instancia. Este ID será usado para generar una nueva cola de ser necesario, en el gestor de cola de mensajes.

Después de clasificados los mensajes se los publicará en la cola de mensajes con nombre igual al ID recibido. Gráficamente, podemos ver que el “motor” ahora comprende al clasificador y al Producer tal y como se lo representa en Figura 18.

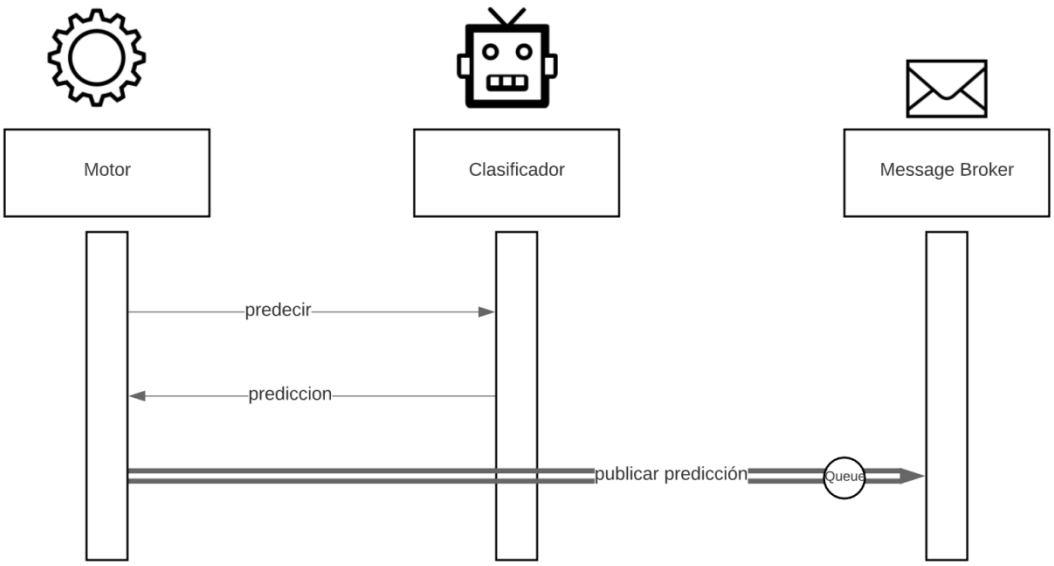


Figura 18. Flujo de mensajes para módulo de clasificación para un ID.

NOTA: La flecha con doble línea representa el canal dedicado de la cola de mensaje.

- **Aplicación de escritorio**

En este Sprint se desarrolló la aplicación de escritorio que se ejecuta en el ordenador del tutelado. Se realizó una reunión y se decidió que la aplicación debería ser un servidor web que se ejecuta sobre el computador, esto principalmente debido a las habilidades del equipo de desarrollo, además de que esto ayuda en la comunicación entre la extensión de Google Chrome y la aplicación de escritorio.

Los requerimientos faltantes de este Sprint se implementaron sobre este servidor web, de lo que se puede mencionar los siguientes puntos.

Cuando la aplicación es ejecutada, se abre un navegador web, ubicándose en la página para iniciar sesión en la aplicación, para que el tutelado pueda autenticarse por medio de usuario y contraseña.
La arquitectura genérica de este componente para que pueda responder a las peticiones fue diseñada como se muestra en Figura 19.

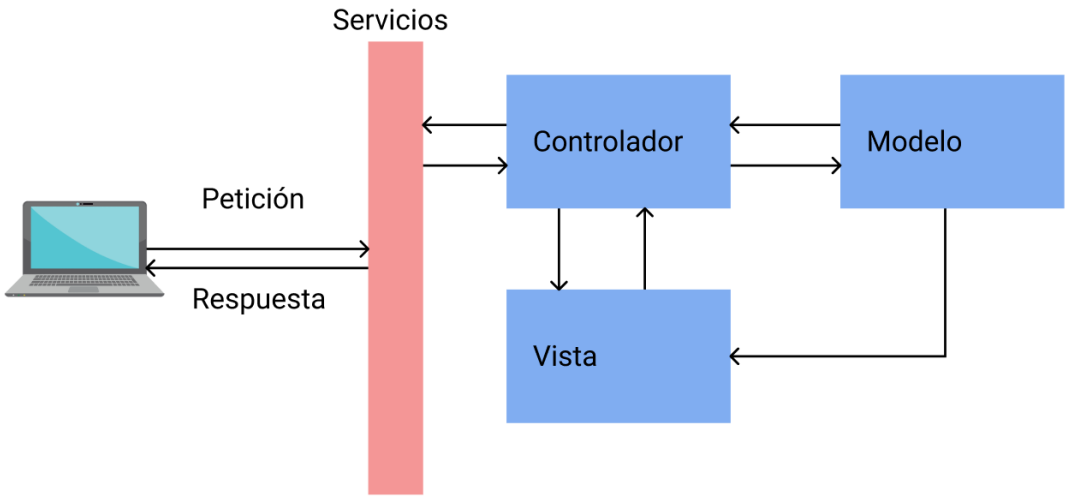


Figura 19. Arquitectura de servicios.

Como se puede apreciar, se optó por un patrón de diseño MVC. Esto nos ayuda a organizar el código, para que sea mantenible y escalable.

Con esta arquitectura están desarrollados los servicios que responden a las siguientes rutas:

- / - Pantalla principal, muestra el inicio de sesión.
- /dashboard – Tablero del usuario. Muestra el listado de incidencias, y presenta las opciones de edición para el tutor.
- /usuario – Gestión de la información del usuario.
- /configuraciones – Gestión de niveles de Grooming

3.8.5. Reunión Diaria

En la ejecución de este Sprint se realizaron 10 reuniones diarias, de aproximadamente 10 minutos cada una. Todas se usaron para revisar el avance en las tareas asignadas.

3.8.6. Revisión del Sprint

El objetico del Sprint fue alcanzado con éxito. Las tareas del Sprint fueron finalizadas y probadas. Este Sprint se caracterizó por la cantidad de tareas que se debía desarrollar, pero no resultaron con mayores inconvenientes.

3.8.7. Pruebas de Aceptación

HISTORIA DE USUARIO	CRITERIO DE ACEPTACION	CUMPLIDO
US02-01	La aplicación bloqueará el dominio cuando un ataque de grooming sea detectado	SI
US02-02	La aplicación generará un ID único cuando la aplicación de escritorio sea ejecutada por primera vez	SI
	La aplicación retorna el ID cuando se realice una petición HTTP /ID	SI
US07-01	La aplicación tendrá un formulario para autenticarse con usuario y contraseña.	SI
	La aplicación redirigirá al usuario autenticado exitosamente a una página principal o dashboard	SI
US07-02	La aplicación tendrá un formulario para que el tutor cambie el usuario y contraseña	SI

US04-01	La aplicación tendrá una pantalla y formulario para que el tutor edite el estado aceptable del ciclo de vida del Grooming	SI
US06-01	La aplicación tendrá una sección donde se listarán las incidencias presentadas en la computadora	SI
US03-01	En el bus de mensajes se generará una cola de mensajes por cada ID de instalación, el cual se podrá observar en el panel de administración de RabbitMQ	SI

Tabla 19. Pruebas de aceptación del Sprint 3.

3.8.8. Adaptación del Product Backlog

Al momento de probar el flujo con todos los componentes funcionando se notó que es importante tener un dominio con un *https*, esto debido a las políticas del navegador contra orígenes sin esta seguridad, por lo que se aumenta una historia de usuario dentro de la épica US05 para generar certificados validos que puedan ser usados en producción.
Entonces, el Product Backlog queda de la siguiente forma.

PRODUCT BACKLOG				
ÉPICA DE USUARIO	ID	HISTORIA DE USUARIO	ESTIMACIÓN (días)	ESTADO
US01	US01-01	Extracción de mensaje recibido para cada chat	5	Terminado
	US01-02	Limpieza de mensajes	1	Terminado
	US01-03	Clasificador de Grooming	7	Terminado
US02	US02-01	Bloqueador de dominio	1	Terminado

	US02-02	Generador de ID	1	Terminado
US03	US03-01	Alerta de incidencia	3	Terminado
	US03-02	Notificación de incidencia	4	Por implementar
US04	US04-01	Configuración de estado aceptable	2	Terminado
US05	US05-01	Despliegue de sistema	1	Por implementar
	US05-02	Generación de certificador SSL validos	1	Por implementar
US06	US06-01	Historial de incidentes	1	Terminado
US07	US07-01	Inicio de sesión de usuario	1	Terminado
	US07-02	Gestión de usuario y contraseña	1	Terminado
US08	US08-01	Desbloqueo de cuenta	1	Terminado

Tabla 20. Product Backlog del Sprint 3.

3.8.9. Retrospectiva del Sprint

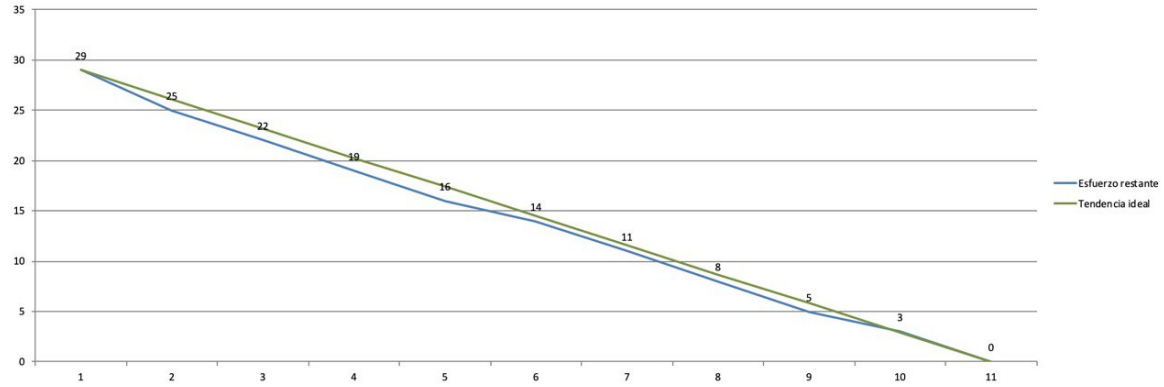


Figura 20. Gráfico Burndown del Sprint 3

3.9. Sprint 4

3.9.1. Objetivo

Concluir el desarrollo del sistema, y generar paquetes de instalación y despliegue final.

3.9.2. Historias de usuario

En Tabla 21, se muestran las historias de usuario que forman parte de este Sprint. El detalle completo de las historias de usuario del Sprint 4 se presentan en Anexo VI.

ID	TÍTULO	ESTIMACIÓN	ESTADO
US03-02	Notificación de incidencia	5	Por implementar
US08-01	Desbloqueo de cuenta	1	Por implementar
US05-01	Despliegue de sistema	1	Por implementar
US05-02	Generación de certificador SSL válidos	1	Por

			implementar

Tabla 21. Historias de usuario para Sprint 4.

3.9.3. Sprint Backlog

HISTORIA DE USUARIO	TAREAS
US03-02	Implementar un bot de Telegram
	Implementar un API dentro del clasificador que notificará a los tutores
	Desarrollar un método para notificar desde la aplicación de escritorio al bot cuando una incidencia se presente
US08-01	Desarrollar un formulario dentro de las opciones del tutor para desbloquear la cuenta del menor
US05-02	Comprar dominios públicos para los servicios de la aplicación
	Generar certificados SSL válidos para los dominios de la aplicación
US05-01	Generar un instalador de la aplicación de escritorio
	Alojar el instalador de la aplicación de escritorio en un sitio web público
	Generar un paquete de instalación del plugin interceptor
	Alojar el paquete de instalación del plugin interceptor en un sitio público
	Desplegar el clasificador sobre un dominio público

Tabla 22. Sprint Backlog del Sprint 4.

3.9.4. Ejecución del Sprint

Durante la ejecución de este último Sprint se puede mencionar los siguientes puntos

Notificación de incidencia

Para integrar el sistema con un Bot de Telegram, el primer paso es registrar el mismo en la aplicación, para lograr esto se usa otro Bot llamado BotFather. Cuando registramos nuestro Bot en la plataforma BotFather nos entrega un token que usaremos para crear la conexión.

La conexión hacia el Bot de Telegram se realizará desde el servidor web que aloja actualmente al clasificador de Grooming. Para lograr esto usaremos la librería *telebot*, la cual ayuda en la comunicación entre el sistema y el Bot de una forma bastante sencilla.

La creación del Bot se lo realizo como se muestra en Figura 21.

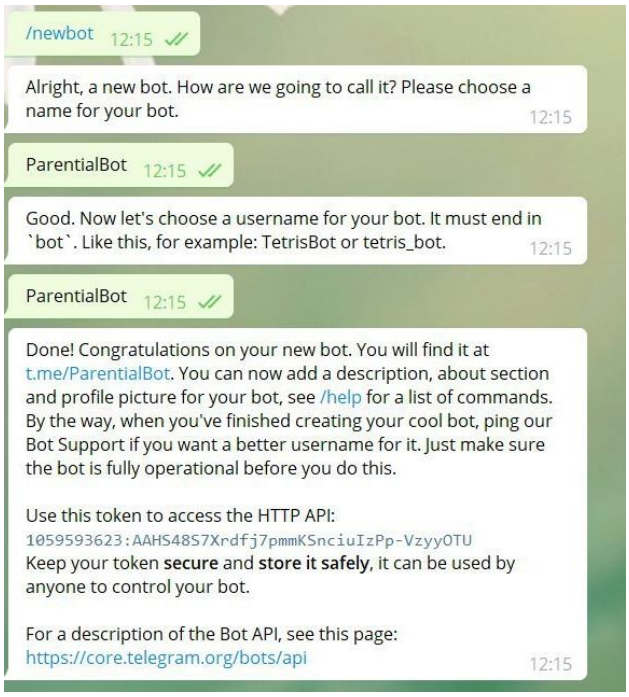


Figura 21. Creación de chatbot mediante BotFather

En Figura 22 se puede observar el flujo de mensajes usado para integrar el sistema con el Bot de Telegram.

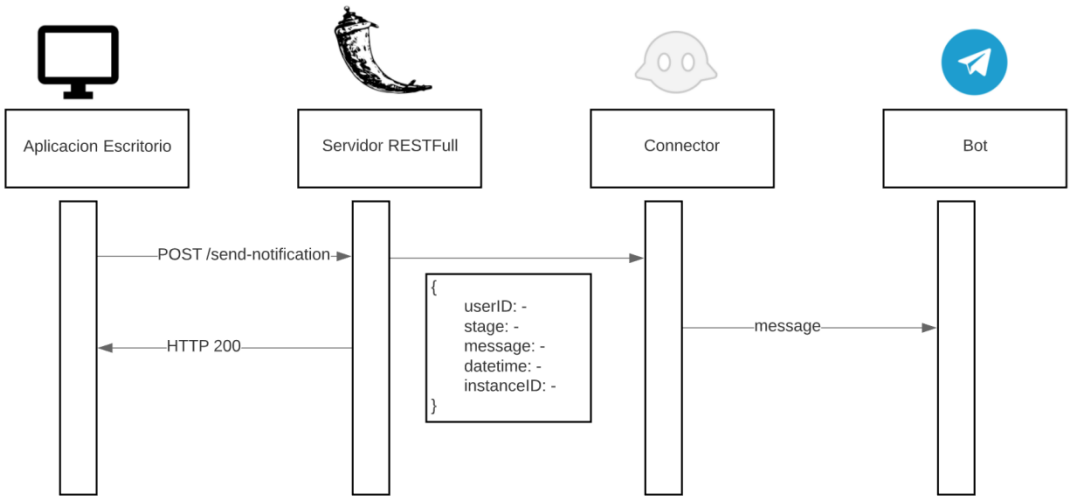


Figura 22. Arquitectura de integración con Telegram

Entonces, cuando la aplicación de escritorio detecte una incidencia, se ejecutará una llamada a un API que se estará ejecutando en el servidor web del clasificador. El servidor

web tendrá almacenado una conexión al Bot de Telegram, y cuando reciba una señal de alerta, enviará un mensaje a los tutores que se hayan registrado.

Despliegue

Un dominio es un nombre único que tiene un espacio dentro de internet, su finalidad es cubrir la dirección IP de un activo en la red en un nombre que las personas puedan recordar fácilmente. Los subdominios son nombres usados para la agrupación de espacios bajo el mismo nombre de dominio. Estos subdominios son usados típicamente con fines de administración.

Para la realización de este proyecto se está usando el dominio thepixcloud.com, un dominio adquirido anteriormente por uno de los miembros de equipo, y sobre este dominio se han creado 2 subdominios parential-classifier.thepixcloud.com y parential.thepixcloud.com, sobre el primero se alojará el servidor web del clasificador de Grooming, y el segundo tendrá un uso especial, ya que se editará el archivo hosts de las computadoras huésped para que resuelva este subdominio hacia el servidor local 127.0.0.1 donde se levantará el servidor web que alojará todos los servicios de la aplicación de escritorio.

La finalidad de realizar este proceso con el subdominio parential, es poder generar unos certificados SSL válidos, ya que en realidad se puede incluir un dominio falso en el archivo hosts el cual resuelva a la IP local, e incluir unos certificados auto firmados, pero el navegador bloquearía las peticiones por no ser un dominio confiable.

Una vez hecha las modificaciones para que funcione con los subdominios elegidos se procedió a generar el instalador para la aplicación de escritorio. Para lograr este objetivo se usó pynsist, el cual generará un ejecutable para Windows, el instalador al ejecutarse como administrador copiará todo el contenido a una carpeta del sistema y generará lo accesos directos. Para la instalación del Plug-in de Google Chrome se generó únicamente un zip, debido a que es aún un prototipo se deicidio no publicarla en la Chrome Web Store de Google.

Los servicios del clasificador fueron levantados sobre un Servidor Privado Virtual (VPS), en el subdominio parential-classifier, este subdominio tiene los certificados SSL que los habíamos generado anteriormente.
Entonces para que el proceso de instalación de los componentes que se ejecutaran de lado del usuario no sea tan complejo se subieron los archivos a un servidor dentro del mismo subdominio, y donde se entregan los manuales de instalación respectivos

3.9.5. Reunión Diaria

En la ejecución de este Sprint se realizaron 9 reuniones diarias, de aproximadamente 10 minutos cada una. En cada reunión se evaluó el progreso de las tareas, y se propuso formas de resolver problemas en la integración con Telegram que fue el mayor conflicto en este Sprint.

3.9.6. Revisión del Sprint

Al terminar este Sprint se puede concluir que se alcanzó el objetivo exitosamente

3.9.7. Pruebas de Aceptación

HISTORIA DE USUARIO	CRITERIO DE ACEPTACION	CUMPLIDO
US08-01	La aplicación permitirá al tutor desbloquear la cuenta del menor desde el panel de administración de la aplicación	SI
US03-02	La aplicación notificará a los tutores por medio de un mensaje de Telegram cuando el menor a sufrido un ataque de Grooming	SI
US05-02	Se ingresará a la aplicación de escritorio por medio de la dirección https://parential.thepixcloud.com	SI
US05-01	Se podrá instalar la aplicación de escritorio por medio de un instalador.	SI
	Se podrá descargar el plugin de Google Chrome y el instalador de la aplicación de escritorio desde un sitio web público.	SI

Tabla 23. Pruebas de aceptación del Sprint 4.

3.9.8. Adaptación del Product Backlog

En la ejecución de los Sprints de este proyecto se cumplieron todas las historias de usuario planificadas al iniciar el desarrollo del trabajo. El proyecto finalmente fue presentado al Product Owner y las historias de usuario aprobadas, debido a que al inicio del proyecto se dejó abierta la posibilidad de crear nuevos requerimientos o adaptar algunos existentes de ser necesario para la optimización del software, el sistema pasó por un proceso de pruebas y después de unas semanas, se agregaron nuevas historias de usuario las cuales fueron estimadas y delegadas a los miembros del equipo. Esto generó la necesidad de un nuevo Sprint, el mismo que es nombrado Sprint 5.

Por lo tanto, el Product Backlog queda como se muestra en Tabla 24.

PRODUCT BACKLOG				
ÉPICA DE USUARIO	ID	HISTORIA DE USUARIO	ESTIMACIÓN (días)	ESTADO
US01	US01-01	Extracción de mensaje recibido para cada chat	5	Terminado
	US01-02	Limpieza de mensajes	1	Terminado
	US01-03	Clasificador de Grooming	7	Terminado
US02	US02-01	Bloqueador de dominio	1	Terminado
	US02-02	Generador de ID	1	Terminado
	US02-03	Bloqueo de dominio, a través de criterio de repetición de incidencias.	1	Por implementar
US03	US03-01	Alerta de incidencia	3	Terminado
	US03-02	Notificación de incidencia	4	Terminado

	US03-03	Registro de clasificaciones en servidor centralizado	1	Por Implementar
US04	US04-01	Configuración de estado aceptable	2	Terminado
US05	US05-01	Despliegue de sistema	1	Terminado
	US05-02	Generación de certificador SSL validos	1	Terminado
US06	US06-01	Historial de incidentes	1	Terminado
US07	US07-01	Inicio de sesión de usuario	1	Terminado
	US07-02	Gestión de usuario y contraseña	1	Terminado
US08	US08-01	Desbloqueo de cuenta	1	Terminado
	US08-02	Estado de cuenta de usuario	1	Por Implementar

Tabla 24. Product Backlog del Sprint 4.

3.9.9. Retrospectiva del Sprint

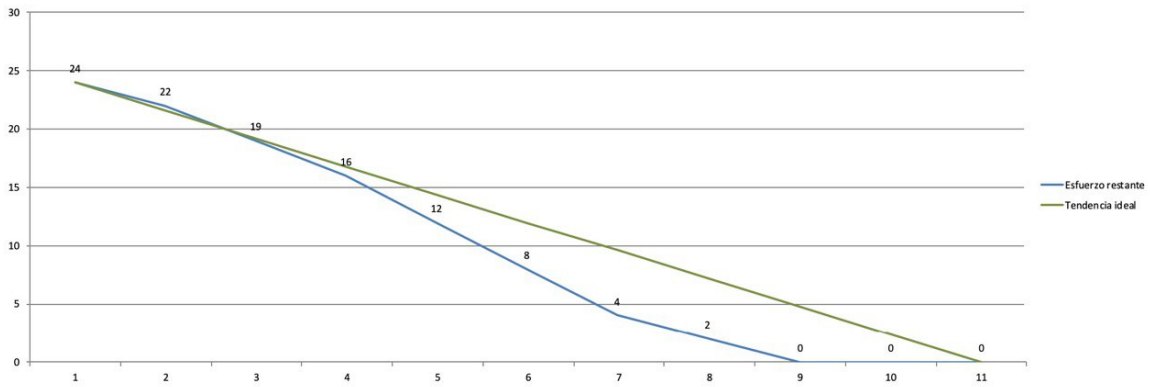


Figura 23. Gráfico Burndown del Sprint 4

3.10. Sprint 5

3.10.1. Objetivos

- Centralizar la información en la base de datos del servidor.
- Implementar un bloqueo de la red social Facebook usando un criterio de repetición de incidencias de clasificaciones.
- Implementar un método para gestionar los usuarios que pueden ser bloqueados.

3.10.2. Historias de Usuario

En Tabla 25, se muestran las historias de usuario que forman parte de este Sprint. El detalle completo de las historias de usuario del Sprint 5 se presentan en Anexo VII.

ID	TÍTULO	ESTIMACIÓN	ESTADO
US02-03	Bloqueo de la red social Facebook, a través de criterio de repetición de incidencias.	1	Por implementar
US03-03	Registro de clasificaciones en servidor centralizado.	1	Por Implementar
US08-02	Estado de bloqueo del usuario.	1	Por implementar

Tabla 25 Historias de usuario de Sprint 5

3.10.3. Sprint Backlog

HISTORIA DE USUARIO	TAREAS
US02-03	Implementar un módulo que permita optimizar el proceso de bloqueo de una cuenta
US03-03	Crear en la base de datos del servidor central las tablas para registrar las clasificaciones
	Registrar las clasificaciones en la base de datos del servidor central

US08-02	Modificar la pantalla principal del usuario para mostrar el estado “Bloqueado” ó “Activo” de la red social Facebook en el ordenador del tutelado.

Tabla 26 Sprint Backlog de Sprint 5

3.10.4. Ejecución de Sprint

En la ejecución de este Sprint se pueden mencionar los siguientes puntos:

- **Estado de bloqueo de la aplicación**

La página principal en la aplicación de escritorio fue cambiada, para que muestre el estado del ordenador del menor. Los estados posibles que se definieron son: Activo para cuando Facebook funciona normalmente, y Bloqueado para cuando se ha detectado una incidencia y la red social ha sido bloqueada.

- **Centralización de base de datos**

El Product Owner indicó que es necesario que ahora las clasificaciones se registren en el servidor principal, estas clasificaciones ayudarían a mejorar el modelo de clasificación para futuros trabajos. Por lo tanto, ahora existe una tabla que registra el mensaje en la base de datos del servidor, el nivel clasificado, la fecha y hora del registro se muestra en Figura 26.

Clasificacion	
Texto	mensaje
Entero	fase
Texto	hash
Fecha y Hora	fecha

Figura 24. Entidad Relación.

Optimización de bloqueo de dominio

En los anteriores Sprints se había definido un sistema de bloqueo bastante rígido, ya que se definió un nivel con el que se bloquearía la red social Facebook. Un gran inconveniente de esto es que un Groomer puede estar deslizándose continuamente sobre una fase menor para no levantar sospechas en el menor. Para detectar esto, se ha implementado un proceso inicial, el cual puede ser mejorado en trabajos futuros, para detectar estos

comportamientos y etiquetarlos como maliciosos. Este proceso lo podemos ver en Figura 25.

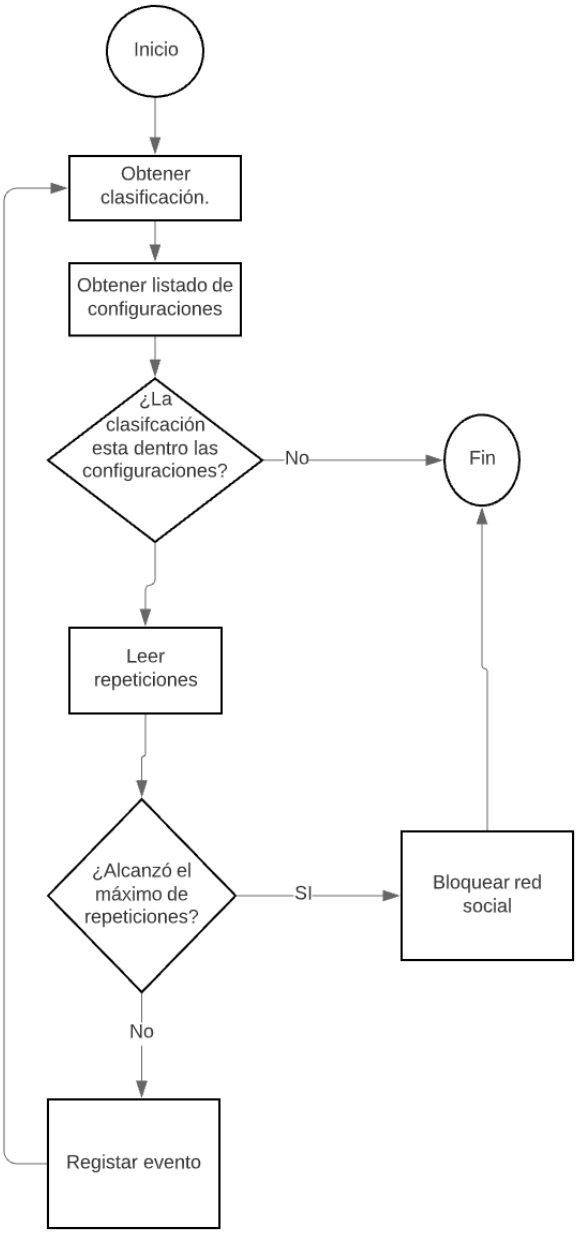


Figura 25. Diagrama de flujo para bloqueo de la red social.

El flujo inicia cuando la aplicación de escritorio recibe la clasificación realizada en el servidor, ahí se obtiene desde la base de datos la configuración de cada fase, máximo índice de repetición, y se compara si la clasificación obtenida está dentro de ese índice, si se encuentra dentro del índice, lee la repetición actual para esa fase, y de llegar al límite establecido en las configuraciones, la aplicación bloquea la red social Facebook. Si no llega al límite, entonces se suma una repetición a la fase y se repite el proceso con la próxima

clasificación. La ejecución de este flujo finaliza, cuando no existe configuración, o cuando el sistema se bloquea.

3.10.5. Reunión Diaria

En la ejecución de este Sprint se realizaron 5 reuniones diarias, de aproximadamente 10 minutos cada una. En las reuniones se revisó principalmente el avance sobre las tareas asignadas.

3.10.6. Revisión del Sprint

El objetivo de este Sprint adicional fue alcanzado exitosamente. El Sprint fue bastante corto ya que solo se realizaron pocos cambios. Con la ejecución de estas historias de usuario, el proyecto ha finalizado.

3.10.6. Pruebas de Aceptación

HISTORIA DE USUARIO	CRITERIO DE ACEPTACION	CUMPLIDO
US02-03	Cuando se presente 5 veces la clasificación S4 para un mismo usuario, bloquear la red social Facebook en el ordenador del menor.	SI
	Cuando se presente 8 veces la clasificación S3 para un mismo usuario, bloquear la red social Facebook en el ordenador del menor.	SI
US03-03	Cuando el administrador del sistema ingrese a la base de datos del servidor, observará los mensajes y sus respectivas clasificaciones.	SI
US08-02	Cuando el tutor ingrese a la página principal de la aplicación de escritorio, podrá observar el estado de bloqueo del ordenador.	SI

Tabla 27. Pruebas de aceptación de Sprint 5.

3.10.7. Adaptación del Product Backlog

Al finalizar el proyecto, el Product Backlog queda de como se muestra en Tabla 28.

PRODUCT BACKLOG				
ÉPICA DE USUARIO	ID	HISTORIA DE USUARIO	ESTIMACIÓN (días)	ESTADO
US01	US01-01	Extracción de mensaje recibido para cada chat	5	Terminado
	US01-02	Limpieza de mensajes	1	Terminado
	US01-03	Clasificador de Grooming	7	Terminado
US02	US02-01	Bloqueador de dominio	1	Terminado
	US02-02	Generador de ID	1	Terminado
	US02-03	Bloqueo de dominio, a través de criterio de repetición de incidencias.	1	Terminado
US03	US03-01	Alerta de incidencia	3	Terminado
	US03-02	Notificación de incidencia	4	Terminado
	US03-03	Registro de clasificaciones en servidor centralizado	1	Terminado
US04	US04-01	Configuración de estado aceptable	2	Terminado
US05	US05-01	Despliegue de sistema	1	Terminado
	US05-02	Generación de certificador SSL validos	1	Terminado

US06	US06-01	Historial de incidentes	1	Terminado
US07	US07-01	Inicio de sesión de usuario	1	Terminado
	US07-02	Gestión de usuario y contraseña	1	Terminado
US08	US08-01	Desbloqueo de cuenta	1	Terminado
	US08-02	Estado de cuenta de usuario	1	Terminado

Tabla 28. Adaptación de Product Backlog de Sprint 5.

3.10.8. Retrospectiva del Sprint

En Figura 28, podemos observar cómo se logró al alcanzar el objetivo un día antes de lo planificado.

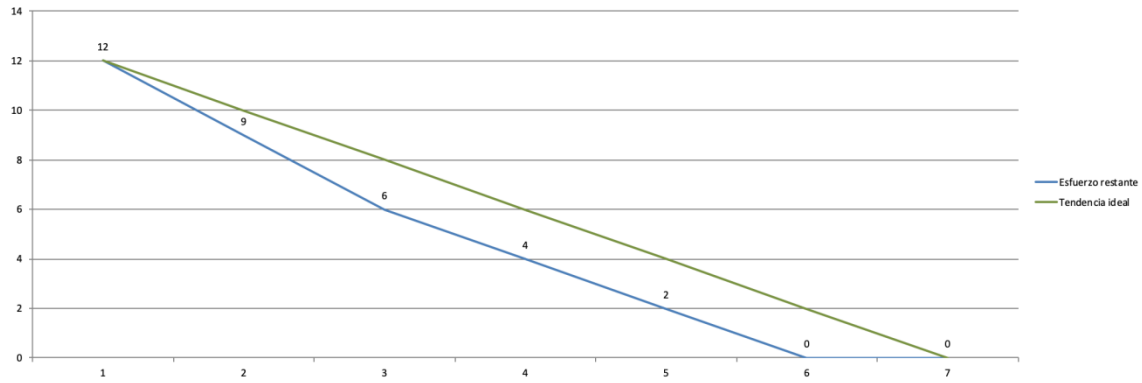


Figura 26. Gráfico Burndown del Sprint 5.

4. RESULTADOS Y DISCUSIÓN

4.1. Producto (Prototipo)

En esta subsección se muestra el producto final de este proyecto, el cual consta de 5 pantallas, las cuales serán descritas más adelante. En Figura 27 se muestra en el lado derecho, a quien se le ha denominado *"Victima"*, entablando una conversación con el que se ha denominado "*Groomer*" al lado izquierdo de la figura. El fragmento mostrado a continuación es una conversación establecida en la red social Facebook.

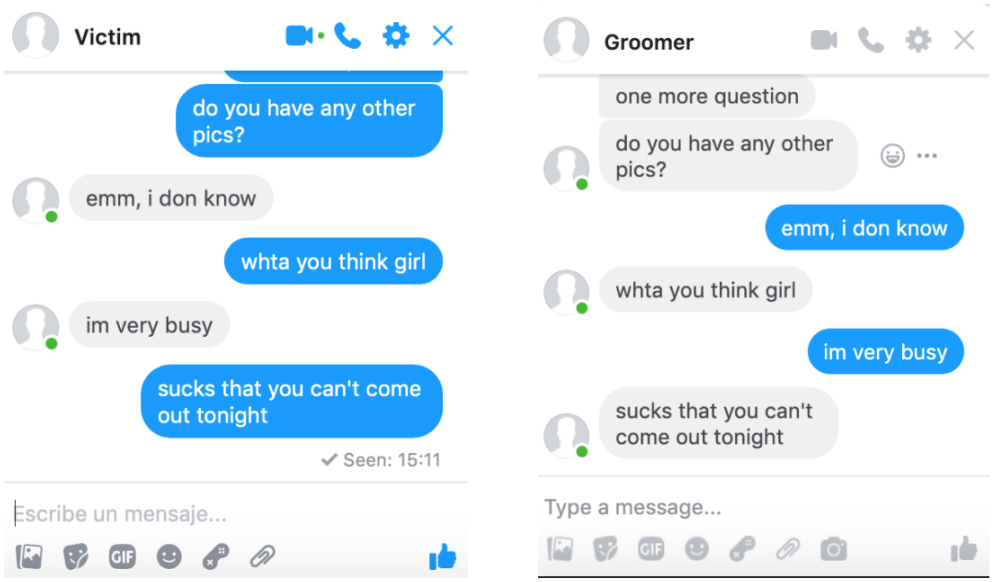


Figura 27: Conversación entre la Víctima y el Groomer

El resultado de esta conversación la red social Facebook es bloqueada, tal y como se muestra en Figura 28.

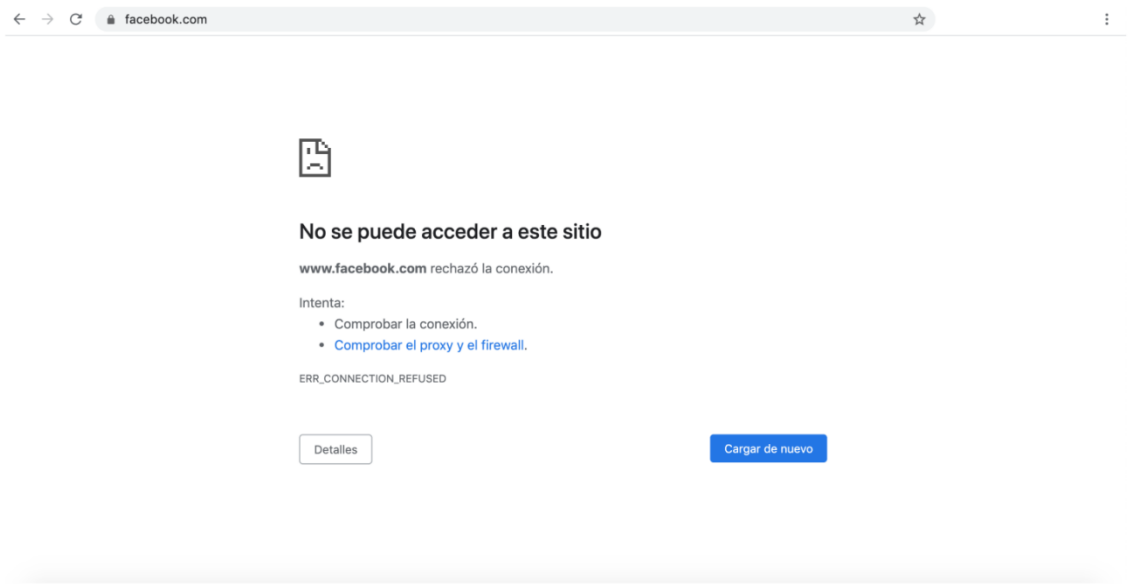


Figura 28: Facebook bloqueado por la aplicación

Posteriormente, los tutores del menor son notificados vía Telegram, un ejemplo del mensaje lo podemos ver en Figura 29.

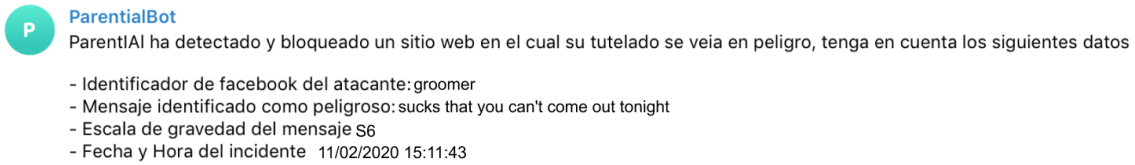


Figura 29: Mensaje que le llega al Tutor

4.1.1. Pantalla de ingreso

Esta pantalla contiene un campo para ingresar el usuario, y un campo para ingresar la contraseña. Cuando el tutor ingresa por primera vez el usuario por defecto es admin y la contraseña por defecto admin. Esto para facilitar el progreso de registro del usuario. Esta pantalla se puede observar en Figura 30.

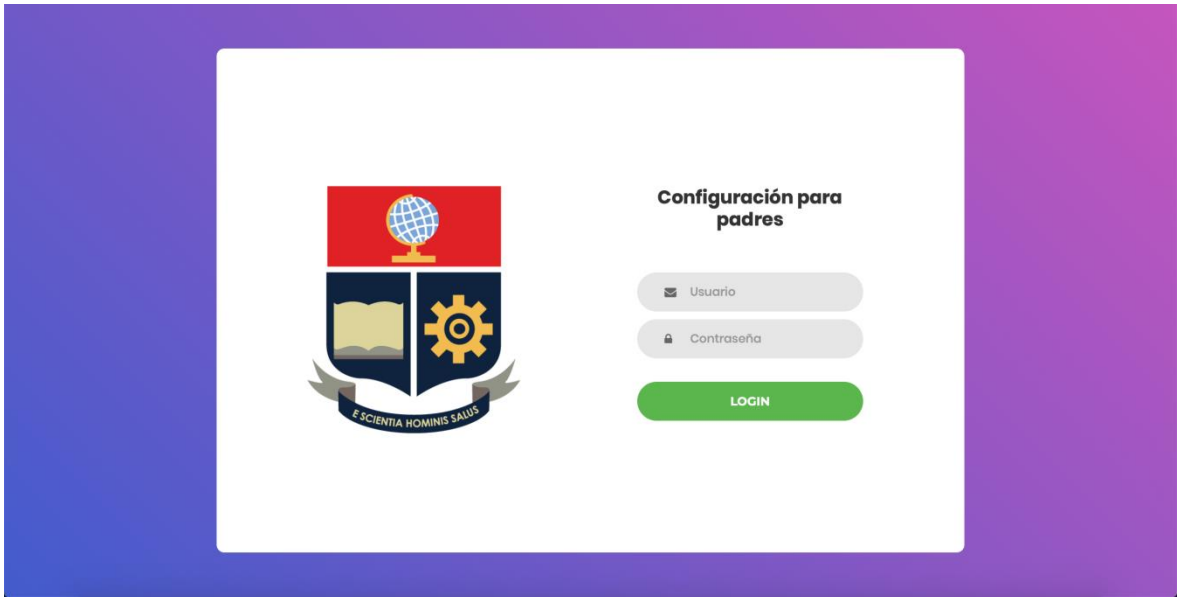


Figura 30. Pantalla de ingreso.

4.1.2. Tablero de la aplicación

Cuando el usuario se autentica en la aplicación se dirige a la pantalla principal donde se muestra el historial de incidentes de la aplicación. En la parte superior de la pantalla se puede apreciar 2 opciones, la primera es un enlace hacia la sección para editar el usuario y la contraseña de la aplicación, y la segunda opción es un link hacia las configuraciones de niveles de Grooming.

A continuación, se muestra el estado de la aplicación, las opciones son Activo o Bloqueado. Esta pantalla se puede observar en Figura 31.

Figura 31. Tablero de la aplicación.

Cuando un ataque se ha detectado, como es el caso en el ejemplo antes presentado, el estado de la aplicación cambia, aparece la opción para desbloquear la cuenta, y en el historial de incidentes se agrega el mensaje y sus detalles. Esta pantalla se puede observar en Figura 32.

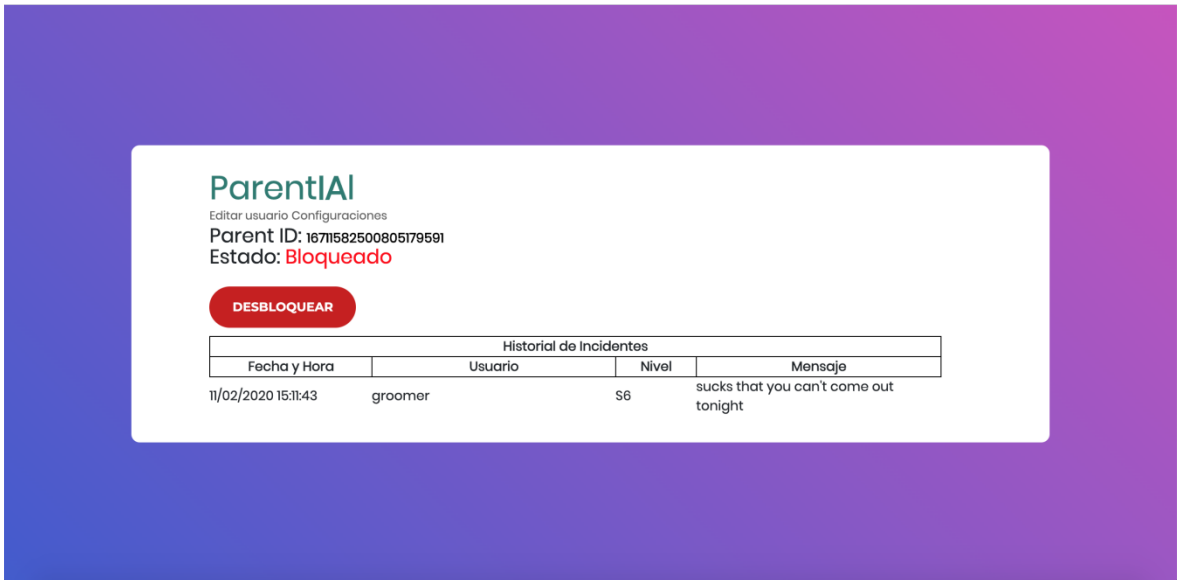


Figura 32. Tablero de la aplicación con estado bloqueado.

4.1.3. Configuración de niveles de Grooming

En la pantalla de niveles de Grooming nos encontramos con dos partes. La primera se refiere a un formulario para actualizar el nivel aceptable que el usuario desea configurar. Las opciones van de S1 a S6 que son las estaciones definidas en el ciclo de vida del Grooming que se explicó en el marco teórico de este trabajo.

La segunda parte se refiere a un formulario que ayudará a configurar la mayor cantidad de veces que se puede repetir un mismo nivel cuando este no cae en el máximo configurado en la parte 1. Esta sección se puede observar en Figura 33.

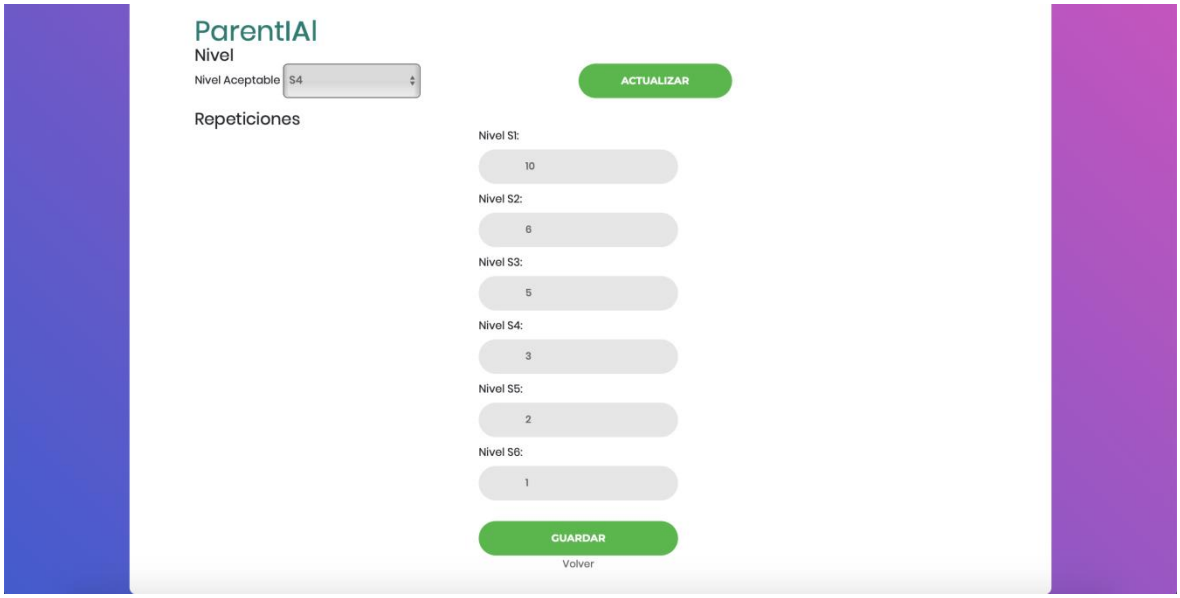


Figura 33. Pantalla de configuración de los niveles de Grooming

4.1.4. Configuración de la cuenta de usuario

El formulario para cambiar las configuraciones del usuario cuenta con el cambio de nombre de usuario, de tipo texto donde se puede modificar el usuario por defecto que es admin. Y los campos para cambiar la contraseña de tipo password. Para cambiar la contraseña se pide ingresar dos veces la nueva contraseña para asegurarse de que la escribió bien, y se agregó el campo de contraseña en donde se ingresará la contraseña antigua para estar seguro de que solo los tutores puedan cambiar esta información. Esta sección se puede observar en Figura 34.

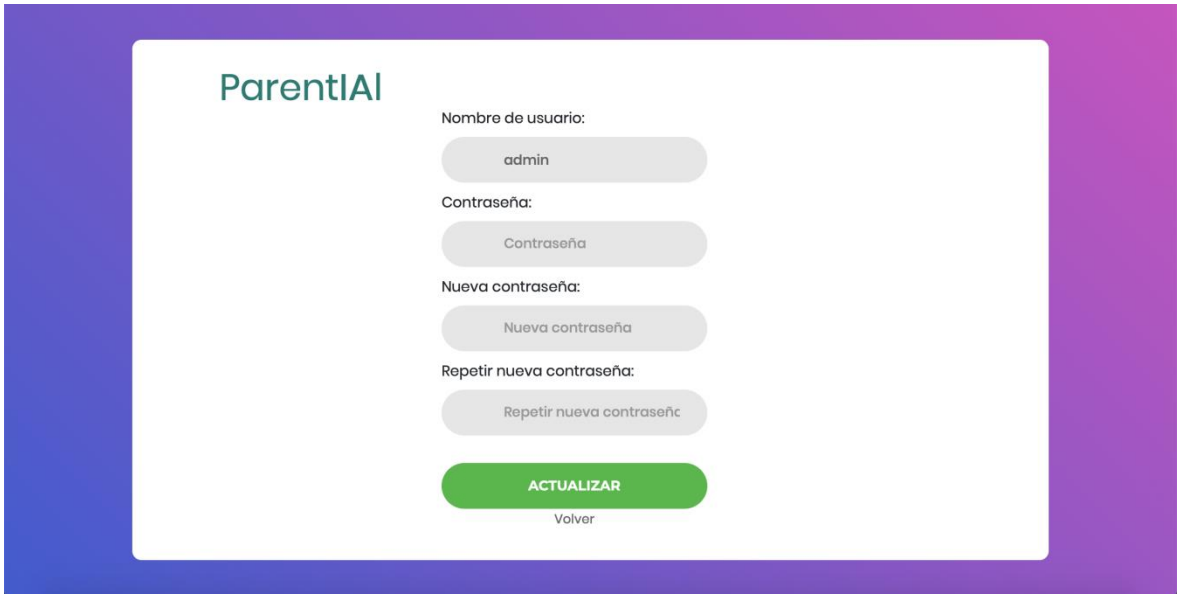


Figura 34. Pantalla de configuración de la cuenta de usuario.

4.2. Prueba de Efectividad

4.2.1. Clasificador de texto

El clasificador de Grooming fue entrenado con un conjunto de 74.642 conversaciones etiquetadas anteriormente, la distribución de las etiquetas la podemos observar en Figura 37. Se realizaron pruebas de efectividad donde el 70% de las conversaciones fueron usadas para entrenar el modelo y el otro 30% se usó para probar el mismo. Este proceso nos mostró que nuestro modelo tiene una precisión promedio del 94%. La precisión obtenida se consideró satisfactoria en comparación al 97% que se obtuvo en la investigación de Zambrano et al. [1]. El reporte generado de precisión se lo puede observar en Figura 35.

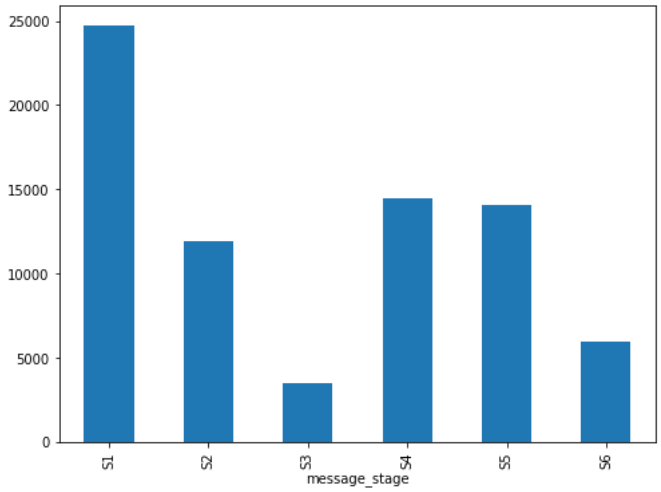


Figura 35: Distribución de etiquetas clasificadas

```
M↓
from sklearn import metrics
print(metrics.classification_report(y_test, y_pred,
                                    target_names=data['message_stage'].unique()))
```

```
precision    recall  f1-score   support

          S4       0.97      0.94      0.95      4828
          S1       0.92      0.97      0.94      8235
          S6       0.98      0.81      0.88      1945
          S2       0.86      0.95      0.90      3885
          S5       0.99      0.95      0.97      4565
          S3       0.92      0.80      0.86      1174

    accuracy                           0.94     24632
   macro avg       0.94      0.90      0.92     24632
weighted avg       0.94      0.94      0.94     24632
```

Figura 36. Reporte de clasificación.

4.3. Prueba de funcionalidad

Para verificar la aceptación del aplicativo se realizó una encuesta de 3 preguntas 2 Obligatorias y una opcional, encuesta realizada a 40 padres y tutores que tienen a cargo un menor de edad, las preguntas no se enfocaban en el uso del aplicativo sino en la funcionalidad que este le otorgaría al cuidado del menor a través del uso de la tecnología, los resultados son los siguientes:

Pregunta 1

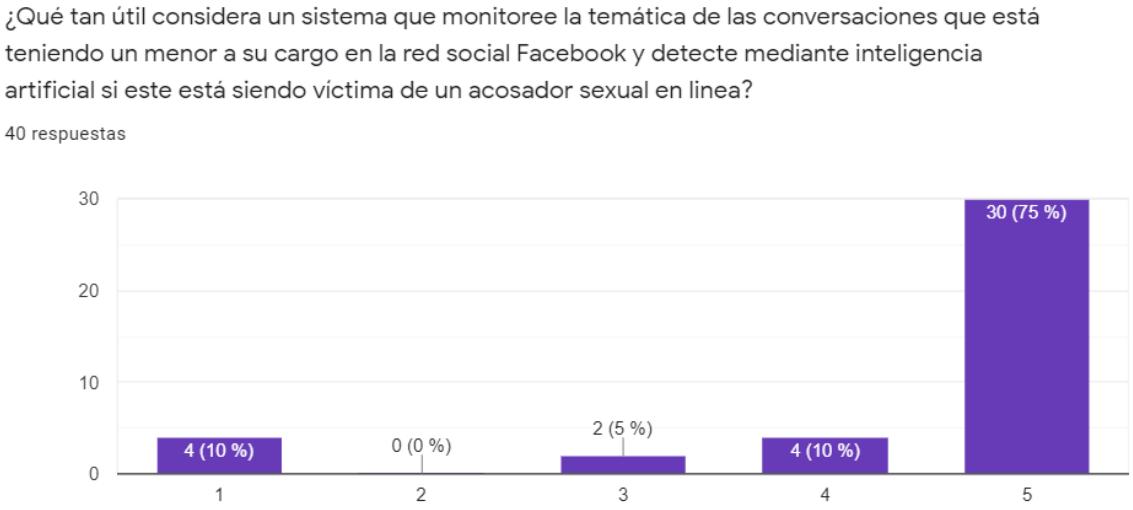


Figura 37: Porcentaje en utilidad del sistema percibida por los tutores

Esta pregunta fue orientada a explicar la funcionalidad del aplicativo en pocas palabras y entender la utilidad que los usuarios perciben que les dará al usarla dándoles una escala del 1 al 5 donde 1 es considerado nada útil y 5 es considerado extremadamente útil. Un 75% de los encuestados consideran el concepto del software extremadamente útil lo que nos indica una alta aceptación en comparación al 10% que indicó que lo considera nada útil.

Pregunta 2

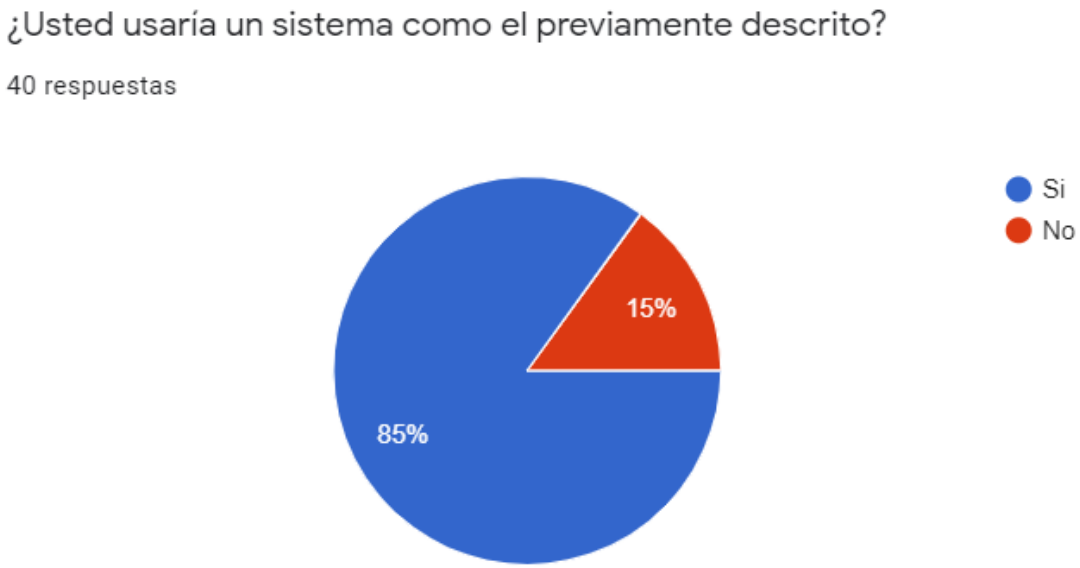
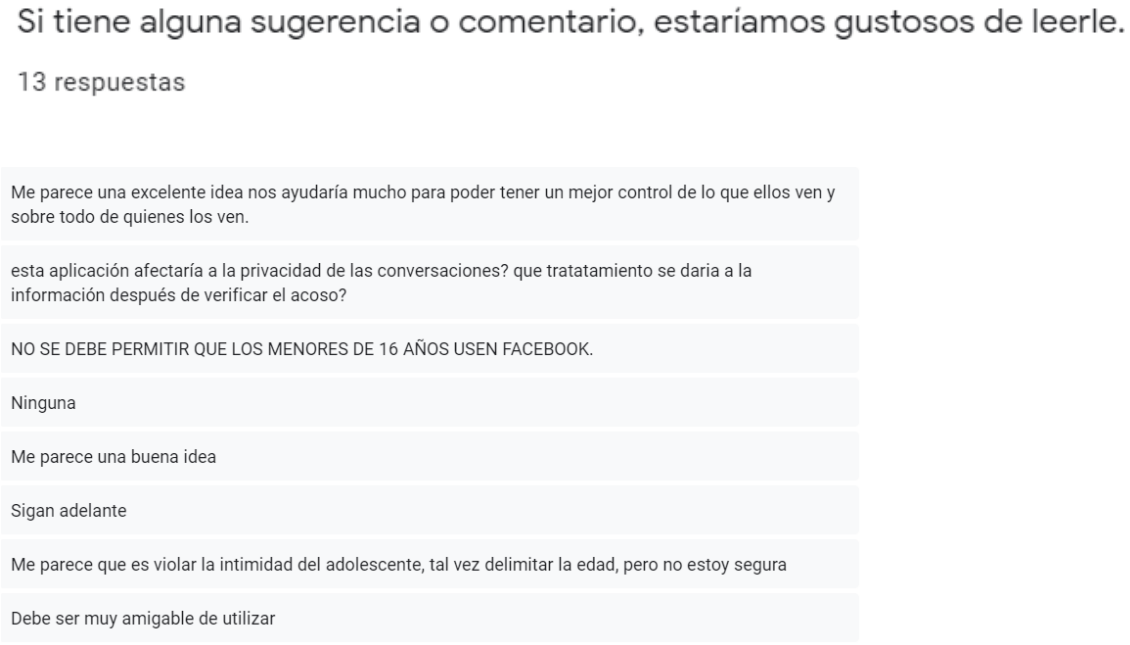


Figura 38: Distribución de respuesta para la factibilidad de usar el sistema

En la segunda pregunta se consultó si los tutores o padres estarías dispuestos a usar este sistema, la encuesta mostró una aceptación del 85% y un 15% afirmó que no lo usaría, lo que también nos muestra una gran aceptación por parte de los encuestados.

Pregunta 3



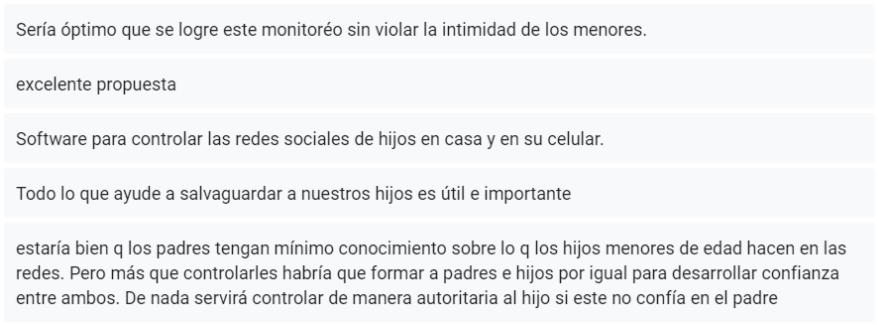


Figura 39: Comentarios acerca del sistema

En la última pregunta se pidió comentarios sobre el sistema especificado para poder entender las respuestas negativas, a primera vista se logró identificar que las respuestas negativas de uso fueron generadas en función de las preocupaciones personales sobre privacidad y manejo de datos que no fueron definidas, siendo este un gran punto de estudio a futuro si el aplicativo llega a pasar de la fase de prototipo.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Se concluye que la implementación de un prototipo de control parental enfocado en la detección inteligente de ataques de Grooming se pudo realizar gracias a los aportes científicos e investigaciones como la de Zambrano et al. [1]. Este tipo de trabajos investigativos siempre son necesarios previo a la ejecución de cualquier implementación práctica de un software u otros, lo que permitió al presente trabajo basarse en teoría científica que respalde la ejecución del mismo.

Adicionalmente, la extracción de datos o información personal de sitios web en tiempo real como la que se realiza en el presente proyecto para obtener las conversaciones del menor, con la autorización del tutor del mismo, es una técnica de hackeo ético que se tuvo que implementar mediante una extensión para el navegador de Google Chrome. Esta técnica única nos ayudó a obtener los datos de entrada para probar el prototipo sin necesidad de solicitar autorizaciones al dueño de la red social Facebook. Por esto se concluye que el uso de este tipo de técnicas es útil para pruebas y prototipos de software mas si se desea implementar este tipo de software en ambientes reales se debe analizar a profundidad las implicaciones legales con la red social y la información del menor.

Hay que tomar en cuenta que gracias a la generalización con la que se realiza el sistema prototipo, se puede modificar el modelo entrenado de entrada, a un modelo que cubra otro tipo de problema social como el cyberbulling, cyberacoso, estafas en línea, entre otros, y el mismo sistema permitiría detectar y bloquear el acceso al sitio donde se está realizando esta actividad ilegítima.

Finalmente, el desarrollo de la aplicación de escritorio que servirá para el control parental y, a su vez, alertará al tutor, en conjunto con las encuentras enfocadas a padres y tutores de menores de edad, determinó que es importante que los padres o tutores tengan conocimiento en tiempo real de los incidentes que ocurren en los dispositivos de los menores a su cargo.

5.2. Recomendaciones

Se recomienda que para un trabajo futuro el sistema sea capaz de detectar cuando la conversación se realiza entre dos personas adultas. Para lograr esto la aplicación debería permitir registrar las cuentas de los adultos que serán ignoradas. Entonces si es una cuenta que no está dentro de la lista blanca se aplicará el proceso que se ha definido en el desarrollo de este trabajo. Si la conversación se ha realizado entre dos personas adultas entonces se ignoran los mensajes.

Adicionalmente, se recomienda que para trabajos futuros el modelo sea entrenado con diccionarios de lengua española. Debido a que este proyecto actualmente está limitado a conversaciones en inglés, ya que se desarrolló basándose en la investigación de Zambrano et al. [1]. Y este a su vez desarrollo su investigación con conversaciones de habla inglesa.

Finalmente, se recomienda que se realicen prueba para optimizar la cantidad de repeticiones que requiera un nivel para considerar una conversación como peligrosa, y bloquee el acceso a la red social Facebook.

6. REFERENCIAS BILIOGRÁFICAS

[1] P. Zambrano, “Technical Mapping of the Grooming Anatomy using Machine Learning Paradigms: An Information Security Approach”,

[2] G. Lansdown, "La seguridad de los niños en línea", Unicef-irc.org, 2012. [En línea]. Disponible: https://www.unicef-irc.org/publications/pdf/ict_spa.pdf. [Accedido: 18- Oct- 2019].

[3] A. Omar, “Control Parental de Facebook”, Famisafe, 2018, [En línea]. Disponible: https://famisafe.wondershare.com/es/parental-control/facebook-parental-controls.html

[4] P. Wanless, "Over 5,000 online Grooming offences recorded in 18 months", www.nspcc.org.uk, 2019. [En línea]. Disponible: https://www.nspcc.org.uk/what-we-do/news-opinion/over-5000-grooming-offences-recorded-18-months/ . [Accedido: 18- Oct- 2019].

[5] M. Nieto, "¿Cómo usan los niños Internet?", elpais.com, 2016. [En línea]. Disponible: https://elpais.com/tecnologia/2016/11/17/actualidad/1479376724_135426.html. [Accedido: 18- Oct- 2019].

[6] T. Satpathy. “A Guide to the Scrum Body Of Knowledge (SBOK™ Guide) – 3rd Edition.” Avondale, Arizona, 2017. [Accedido: 18- Oct- 2019].

[7] DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. Psychological bulletin, 129(1), 1974.

[8] Joachims, T. Learning to Classify Text Using Support Vector Machines. Boston, MA: Springer US, 2002.

[9] D. Bogdanova, P. Rosso, and T. Solorio, “Exploring high-level features for detecting cyberpedophilia,” Computer Speech and Language, vol. 28, no. 1, 2014. [Online]. Disponible: http://dx.doi.org/10.1016/j.csl.2013.04.007

[10] R. Pressman, “Ingeniería del Software. Un enfoque práctico”. McGraw Hill Interamericana editores. 7ma Edición.

[11] GCFGlobal. 2020. Seguridad En Internet: ¿Qué Es El Control Parental?. [En Línea] Disponible: https://edu.gcfglobal.org/es/seguridad-en-internet/que-es-el-control-parental/1/ [Accedido 15 Marzo 2020].

[12] M. C. Seto, «Is Pedophilia a Sexual Orientation?», Archives of Sexual Behavior, vol. 41, n.º 1, pp. 231-236, ene. 2012.

[13] Sergio Sánchez-Migallón Granados, «Fenomenología», Philosophica, n.º 79, feb. 2014, doi: 10.17421/2035_8326_2014_SSM_1-1.

[14] American Psychiatric Association, «Diagnostic and Statistical Manual of Mental Disorders». American Psychiatric Association.

[15] Dombert, B., 2020. [imagen] Disponible: https://www.researchgate.net/figure/Clothed-example-pictures-of-the-Virtual-People-Set-for-females-and-males-at-different_fig1_234084241. [Accedido 15 Marzo 2020].

[16] EUROPOL (2017). «Child Sexual Exploitation». [En línea] Disponible: *https://www.europol.europa.eu/crime-areas-and-trends/crime-areas/childsexual-exploitation.* [Accedido 08 Febrero, 2020]

[17] Graupner, H. (2000) Sexual consent: The criminal law in Europe and overseas. Archives of Sexual Behavior, 29(5), pp. 415-461.

[18] Daigneault, I., Vézina-Gagnon, P., Bourgeois, C., Esposito, T. and Hébert, M. (2017) Physical and mental health of children with substantiated sexual abuse: gender comparisons from a matched-control cohort study. Child abuse & neglect, 66, pp. 155-165

[19] Aslan D. (2011). Critically evaluating typologies of internet sex offenders: A psychological perspective. Journal of Forensic Psychology Practice, 11(5), 406–431. doi:10.1080/15228932.2011.588925.

[20] M. C. Seto, J. M. Wood, K. M. Babchishin, y S. Flynn, «Online solicitation offenders are different from child pornography offenders and lower risk contact sexual offenders.», Law and Human Behavior, vol. 36, n.º 4, pp. 320-330, ago. 2012.

[21] J. Wolak y D. Finkelhor, «Are Crimes by Online Predators Different From Crimes by Sex Offenders Who Know Youth In-Person?», Journal of Adolescent Health, vol. 53, n.º 6, pp. 736-741, dic. 2013, doi: 10.1016/j.jadohealth.2013.06.010.

[22] NSPCC (2018). Grooming: What it is, signs and how to protect children. [En linea] Disponible from https://www.nspcc.org.uk/what-is-child-abuse/types-of-abuse/grooming/. [Disponible 16 Febrero, 2020]

[23] Olson, L. N., Daggs, J. L., Ellevold, B. L. and Rogers, T. K. (2007). Entrapping the innocent: Toward a theory of child sexual predators' luring communication. Communication Theory, 17(3), pp. 231-251.

[24] Winters, G. M. and Jeglic, E. L. (2017). Stages of sexual grooming: recognizing potentially predatory behaviours of child molesters. Deviant behaviour, 38(6),pp. 724-733.

[25] Mooney, J. L., & Ost, S. (2013). Group localised grooming: What is it and what challenges does it pose for society and law. Child & Fam. LQ, 25, 425.

[26] L. N. Olson, J. L. Daggs, B. L. Ellevold, y T. K. K. Rogers, «Entrapping the Innocent: Toward a Theory of Child Sexual Predators? Luring Communication», Communication Theory, vol. 17, nº 3, pp. 231-251, ago. 2007, doi: 10.1111/j.1468-2885.2007.00294.x.

[27] R. Williams, I. A. Elliott, y A. R. Beech, «Identifying Sexual Grooming Themes Used by Internet Sex Offenders», Deviant Behavior, vol. 34, nº 2, pp. 135-152, feb. 2013.

[28] D. M. Blei, «Probabilistic topic models», Communications of the ACM, vol. 55, nº 4, p. 77, abr. 2012, doi: 10.1145/2133806.2133826.

[29] B. I. Messaoud, K. Guennoun, M. Wahbi, and M. Sadik, “Advanced persistent threat: New analysis driven by life cycle phases and their challenges,” in 2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS). IEEE, 2016, pp. 1–6.

[30] P. J. Black, M. Wollis, M. Woodworth, and J. T. Hancock, “A linguistic analysis of grooming strategies of online child sex offenders: Implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world,” Child Abuse & Neglect, vol. 44, pp. 140–149, 2015.

[31] BBVAOpen4U. 2020. API REST: Qué Es Y Cuáles Son Sus Ventajas En El Desarrollo De Proyectos. [En línea] Disponible: https://bbvaopen4u.com/es/actualidad/api-rest-que-es-y-cuales-son-sus-ventajas-en-el-desarrollo-de-proyectos. [Accedido 15Marzo 2020].

[32] Grupo Deidev, “¿Qué es el control parental y para qué sirve? - SecureKids", Control parental Android para móvil y tablet | SecureKids, 2015. [En linea]. Disponible: https://securekids.es/que-es-el-control-parental-y-para-que-sirve/. [Accedido: 08 Mar 2020]

[33] Alan F. Westin, Privacy And Freedom, 25 Wash. & Lee L. Rev. 166 (1968), [En línea]. Disponible: https://scholarlycommons.law.wlu.edu/wlulr/vol25/iss1/20. [Accedido: 08-Mar- 2020]

[34] Amazon Web Services, Inc. 2020. ¿Qué Es Una Cola De Mensajes?. [En línea] Disponible: https://aws.amazon.com/es/message-queue/. [Accedido 15 Marzo 2020].

[35] G.-X. Yuan, C.-H. Ho, y C.-J. Lin, «Recent Advances of Large-Scale Linear Classification», Proceedings of the IEEE, vol. 100, n.º 9, pp. 2584-2603, sep. 2012.

[36] Gonzalez, J., 2018. Tipos De Aprendizaje Automático. [En línea] Medium. Disponible: https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2. [Accedido 15 Marzo 2020].

[37] Schwaber K., Beedle M. (2001). Agile software development with scrum. Prentice Hall PTR, Upper Saddle River, NJ, USA

[38] Herbsleb, J. D., & Moitra, D. (2001). Global software development. Software, IEEE, 18(2), 16-20

[39] Deemer,P., Benefield, G., Larman, C., Vodde, B.(2008). The Scrum Primer Version 2.0, Scrum Training Institute, available at http://www.scrumprimer.com ,last visited on August 19, 2015.

[40] Kniberg, H. (2007). Scrum and XP from the Trenches: How we do Scrum. [En línea] Disponible: http://www.lulu.com/content/899349. [Accedido: 16 Marzo 2020].

[41] Schwaber K.(2008), It's Not Scrum If..., Presentation at Stockholm Scrum Gathering Fall 2008,[En línea] Disponible: http://www.scrumalliance.org/resource_download/441. [Accedido: 16 Marzo 2020.]

[42] Schwaber K., Beedle M. (2001). Agile software development with scrum. Prentice Hall PTR, Upper Saddle River, NJ, USA

[43] Schwaber K., 2009, Scrum Guide, Online Guide, [En línea] Disponible: *http://www.scrumalliance.org/resource_download/598*. [Accedido: 16 Marzo 2020].

7. ANEXOS

7.1. Anexo I: Comparación de definiciones de Grooming

Lanning, 2001		Durkin, 1997		Alexy, Burgess & Baker, 2005		Beech et al., 2008	Mclaughlin, 2000	
Situacional	Oportunista	**Comerciantes**	Tráfico de pornografía infantil	**Comerciantes**	Recopila e intercambiar pornografía infantil	Acceso a la pornografía infantil impulsivamente / por curiosidad.	**Coleccionista**	Recopila e intercambiar pornografía infantil
Preferencial	Tiene un propósito y una víctima objetivo	**Networkers**	Comunicación con otros con intereses similares	**Viajeros**	Usa Internet para conocer y preparar a los niños para luego abusar de ellos sin conexión	Acceso e intercambio de imágenes para satisfacer el interés sexual desviado	**Viajeros**	Uso de Internet para conocer y preparar a los niños para luego abusar de ellos sin conexión
Miscelaneos	Puede acceder de manera inapropiada a material como parte de investigaciones, bromistas y adultos jóvenes.	**Groomers**	Comunicación sexual con niños	**Viajeros comerciantes**	Recopila e intercambia pornografía infantil, así también viajar para encontrarse fuera de línea para más abusos	Uso del Internet como parte del comportamiento ofensivo, ya sea para preparar a las víctimas o para distribuir pornografía infantil producida	**Fabricante**	Productor y distribuidor de pornografía infantil.
		Viajeros	Usa Internet para conocer y preparar a los niños para luego abusar de ellos sin conexión			Acceso a imágenes de abuso con fines no sexuales. p.ej. Ganancia financiera	**Platicador**	Recopila información erótica infantil (no pornografía) puede entablar comunicación sexual en línea con niños

7.2. Anexo II: Historias Épicas de Usuario

HISTORIA EPICA DE USUARIO	
ID: US01	**Prioridad:** 4
Descripción: Como tutor de un menor requiero un sistema capaz de identificar cuando mi tutelado está siendo acosado por medio de prácticas de Grooming para ayudar a reducir el peligro al que está expuesto al usar Facebook	

HISTORIA EPICA DE USUARIO	
ID: US02	**Prioridad:** 4
Descripción: Como tutor requiero un sistema que sea capaz de bloquear la interacción que mi tutelado tiene con Facebook cuando se ha detectado que es víctima de Grooming, para así evitar consecuencias más graves y yo poder tomar las acciones mientras tanto	

HISTORIA EPICA DE USUARIO	
ID: US03	**Prioridad:** 3
Descripción: Como tutor requiero un sistema capaz de notificarme cuando mi tutelado ha sido expuesto a un ataque de Grooming, para poder tomar acciones al respecto	

HISTORIA EPICA DE USUARIO	
ID: US04	**Prioridad:** 3
Descripción: Como tutor requiero que el sistema me permita configurar las fases que considero son inadecuadas, y así ser tener más control sobre lo que mi tutelado puede ver en redes sociales.	

HISTORIA EPICA DE USUARIO	
ID: US05	**Prioridad:** 4
Descripción: Como tutor requiero una aplicación que me permita instalar el sistema de control parental en la computadora de mi tutelado, y así poder ejecutar el control parental.	

HISTORIA EPICA DE USUARIO	
ID: US06	**Prioridad:** 2

Descripción: Como tutor requiero que el sistema registre los ataques de Grooming detectados para de esta forma monitorear las acciones del tutelado.

HISTORIA EPICA DE USUARIO	
ID: US07	**Prioridad:** 3
Descripción: Como tutor requiero que el sistema implemente un inicio de sesión con usuario y contraseña, para que la única persona que pueda administrar la cuenta sea yo.	

HISTORIA EPICA DE USUARIO	
ID: US08	**Prioridad:** 2
Descripción: Como tutor requiero que el sistema me permita gestionar la cuenta para así desbloquear la computadora de mi tutelado cuando haya definido que está libre de peligro.	

7.3. Anexo III: Historias de usuario de Sprint 1

HISTORIA DE USUARIO		
ID: US01-01	**Días estimados:** 5	**Prioridad:** 4
Título: Extracción de mensaje recibido para cada chat		**Usuario:** Tutor
Descripción: Como tutor de un menor necesito que el sistema analice las conversaciones y así conocer cuando está siendo acosado.		
Responsable: Christian Oña		

HISTORIA DE USUARIO		
ID: US01-02	**Días estimados:** 1	**Prioridad:** 4
Título: Limpieza de mensajes		**Usuario:** Administrador
Descripción: Como administrador necesito que el sistema extraiga solo el texto de las conversaciones debido a que el algoritmo de clasificación solo puede analizar texto plano.		
Responsable: Jairo Proaño		

7.4. Anexo IV: Historias de usuario de Sprint 2

HISTORIA DE USUARIO		
ID: US01-03	**Días estimados:** 7	**Prioridad:** 4

Título: Clasificador de Grooming		**Usuario:** Administrador
Descripción: Como administrador necesito que exista un clasificador inteligente de Grooming debido a la cantidad de conversaciones que puede haber.		
Responsable: Jairo Proaño		
HISTORIA DE USUARIO		
ID: US03-01	**Días estimados:** 3	**Prioridad:** 3
Título: Alerta de incidencia		**Usuario:** Administrador
Descripción: Como administrador necesito que el sistema alerte a los ordenadores asociados sobre las incidencias presentes para que cada ordenador realice las operaciones necesarias en base a sus configuraciones		
Responsable: Christian Oña		

7.5. Anexo V: Historias de usuario de Sprint 3

HISTORIA DE USUARIO		
ID: US02-01	**Días estimados:** 1	**Prioridad:** 4
Título: Bloqueador de dominio		**Usuario:** Tutor
Descripción: Como tutor requiero que el sistema bloqueo la red social Facebook cuando se detecte un ataque y de esta forma evitar que el menor a mi cargo siga siendo victima del acosador.		
Responsable: Christian Oña		

HISTORIA DE USUARIO		
ID: US02-02	**Días estimados:** 1	**Prioridad:** 1
Título: Generador de ID		**Usuario:** Administrador
Descripción: Como administrador necesito que la maquinas generen un ID único para poder identificar a cada ordenador dentro de la red.		
Responsable: Christian Oña		

HISTORIA DE USUARIO		
ID: US04-01	**Días estimados:** 2	**Prioridad:** 3
Título: Configuración de estado aceptable		**Usuario:** Tutor
Descripción: Como tutor requiero que el sistema me permita configurar el nivel de Grooming que me parece poco apto y así tener un mayor control sobre las conversaciones que el menor a mi cargo está teniendo.		
Responsable: Jairo Proaño		

HISTORIA DE USUARIO

ID: US06-01	**Días estimados:** 1	**Prioridad:** 2
Título: Historial de incidentes		**Usuario:** Tutor
Descripción: Como tutor requiero un tablero donde pueda observar las incidencias presentadas sobre el ordenador del menor a mi cargo, y así poder las medidas necesarias.		
Responsable: Jairo Proaño		

HISTORIA DE USUARIO		
ID: US07-01	**Días estimados:** 1	**Prioridad:** 3
Título: Inicio de sesión de usuario		**Usuario:** Tutor
Descripción: Como tutor requiero que el sistema me permita iniciar sesión en una aplicación en el ordenador del menor a mi cargo y así asegurarme que solo yo pueda acceder al mismo.		
Responsable: Jairo Proaño		

HISTORIA DE USUARIO		
ID: US07-02	**Días estimados:** 1	**Prioridad:** 3
Título: Gestión de usuario y contraseña		**Usuario:** Tutor
Descripción: Como tutor necesito que el sistema me permita cambiar el usuario y contraseña de mi cuenta de usuario en el ordenador del menor y así asegurarme de que solo yo acceso al mismo.		
Responsable: Jairo Proaño		

7.6. Anexo VI: Historias de usuario de Sprint 4

HISTORIA DE USUARIO		
ID: US03-02	**Días estimados:** 4	**Prioridad:** 3
Título: Notificación de incidencia		**Usuario:** Tutor
Descripción: Como tutor requiero que el sistema me notifique a Telegram cuando una incidencia se presente y así poder tomar las medidas necesarias con el menor a mi cargo		
Responsable: Christian Oña		

HISTORIA DE USUARIO		
ID: US05-01	**Días estimados:** 1	**Prioridad:** 4
Título: Despliegue de sistema		**Usuario:** Administrador
Descripción: Como administrador necesito que el sistema sea levantado sobre el servidor y dominio que les entregue, y exista un método de instalación en el ordenador del menor y así esté disponible para ser usado.		
Responsable: Jairo Proaño		

HISTORIA DE USUARIO		
ID: US05-02	**Días estimados:** 1	**Prioridad:** 1
Título: Generación de certificador SSL validos		**Usuario:** Administrador
Descripción: Como administrador del sistema necesito que el sistema bajo un dominio cifrado para así proteger los mensajes que son transmitidos entre los componentes del sistema.		
Responsable: Jairo Proaño		

HISTORIA DE USUARIO		
ID: US08-01	**Días estimados:** 1	**Prioridad:** 3
Título: Desbloqueo de cuenta		**Usuario:** Tutor
Descripción: Como tutor requiero que el sistema me permita desbloquear la cuenta en el ordenador del menor a mi cargo cuando he decido que esta fuera de peligro y así él pueda ingresar nuevamente a la red social Facebook.		
Responsable: Jairo Proaño		

7.7. Anexo VII: Historias de usuario de Sprint 5

HISTORIA DE USUARIO		
ID: US02-03	**Días estimados:** 1	**Prioridad:** 1
Título: Bloqueo de la red social Facebook, a través de criterio de repetición de incidencias.		**Usuario:** Tutor
Descripción: Como tutor requiero que el sistema bloquee la red social Facebook después de un número de veces que se repite el mismo nivel de incidencia, y así darle al menor un poco más de libertad ya que actualmente es muy rígido.		
Responsable: Christian Oña		

HISTORIA DE USUARIO		
ID: US03-03	**Días estimados:** 1	**Prioridad:** 1
Título: Registro de clasificaciones en servidor centralizado.		**Usuario:** Administrador
Descripción: Como administrador requiero que los mensajes y clasificaciones se registren en una base de datos del servidor principal, para con esa información seguir entrenando al clasificador de Grooming		
Responsable: Jairo Proaño		

HISTORIA DE USUARIO		
ID: US08-02	**Días estimados:** 1	**Prioridad:** 1
Título: Estado de bloqueo del usuario.		**Usuario:** Tutor

Descripción: Como tutor requiero que se muestre en el tablero de la aplicación el estado del ordenador para así tener una mejor experiencia al navegar por el sistema.
Responsable: Christian Oña