Servicio De Análisis Y Recolección De Datos Basado En OSINT

Mauricio Andrés Martín
Departamento de Ingeniería de Sistemas y Computación
Universidad de los Andes
Bogotá, Colombia
ma.martinm1@uniandes.edu.co

Simon Guzman-Londoño
Departamento de Ingeniería de Sistemas y Computación
Universidad de los Andes
Bogotá, Colombia
s.guzman@uniandes.edu.co

***Abstract*—This project presents an OSINT-based cybersecurity solution designed specifically for small and medium enterprises (SME), addressing the critical gap between growing digital threats and limited security resources. The proposed platform integrates automated monitoring of exposed credentials, vulnerabilities, digital presence, and sector-specific threats into a unified, user-friendly interface. By transforming complex security data into actionable insights, the system enables non-technical teams to proactively identify and mitigate risks without requiring specialized expertise. Key innovations include sector-specific risk profiling, machine learning-powered predictive analysis, and automated reporting with prioritized recommendations. The solution stands out for its balanced approach combining enterprise-level capabilities with SME accessibility, offering continuous protection through cloud-based deployment while maintaining strict data privacy standards. This development not only enhances organizational security postures but also promotes cybersecurity awareness as a competitive advantage in the digital economy, particularly for resource-constrained businesses facing sophisticated threats.**

***Index Terms*—OSINT, AWS, Ciberseguridad, Cloud, SaaS**

I. Introducción

En el entorno digital actual, Open Source Intelligence (OSINT) se ha convertido en un pilar crítico para la ciberseguridad, pero también en un arma de doble filo. La red alberga un volumen exponencial de información pública, desde credenciales expuestas y fugas de datos hasta discusiones en foros y noticias sectorizadas. Si bien representa una mina de oro para la inteligencia estratégica, plantea desafíos abrumadores: ¿cómo distinguir señales de ruido? ¿Cómo validar la veracidad de los datos? Y, sobre todo, ¿cómo transformar esta información en acciones concretas que protejan a las organizaciones?

Las empresas, especialmente las pymes, enfrentan una falta de visibilidad sobre sus activos expuestos (correos, documentos sensibles, menciones en redes) y su potencial explotación por actores maliciosos. A esto se suma la sobrecarga de fuentes como blogs técnicos, bases de datos de vulnerabilidades, redes sociales y la escasez de recursos para procesarlas de manera efectiva.

II. Contexto

En el panorama actual de amenazas digitales, las pequeñas y medianas empresas se encuentran en una posición particularmente vulnerable. Mientras las grandes corporaciones cuentan con equipos especializados y presupuestos robustos para su seguridad digital, las PYMES enfrentan el desafío de protegerse en un entorno cada vez más hostil sin disponer de los mismos recursos. Esta disparidad ha llevado al surgimiento de iniciativas que buscan democratizar el acceso a herramientas de seguridad efectivas, diseñadas específicamente para organizaciones con capacidades técnicas y económicas limitadas.

En el ámbito corporativo, el OSINT se ha consolidado como un componente crítico de los programas de seguridad avanzados. Grandes empresas y firmas especializadas en *threat intelligence* emplean estas técnicas para monitorizar amenazas en tiempo real, analizar riesgos reputacionales y fortalecer sus posturas de seguridad. Un ejemplo claro es el uso de OSINT para identificar credenciales de empleados expuestas en filtraciones de datos, lo que permite prevenir ataques de *credential stuffing* antes de que ocurran. De igual forma, el monitoreo de redes sociales y foros especializados proporciona alertas tempranas sobre campañas de desinformación o intentos de suplantación de identidad.

El entorno de ciberseguridad actual se caracteriza por una escalada sin precedentes en la frecuencia, sofisticación e impacto de las amenazas digitales. Según el Informe Global de Amenazas 2024 de CrowdStrike, solo en 2023 se identificaron 34 nuevos grupos adversarios, lo que eleva a más de 230 los actores de amenaza monitoreados a nivel global [1]. Este crecimiento se ve acompañado por un aumento del 76% en incidentes de filtración de datos y un 75% en intrusiones a entornos en la nube, vectores particularmente críticos para organizaciones cuyas operaciones dependen de infraestructuras cloud.

Estas cifras no solo reflejan un panorama más hostil, sino que evidencian una realidad alarmante: la superficie de ataque se expande más rápido que la capacidad de muchas empresas para protegerla. En este contexto, Open Source Intelligence (OSINT) emerge como una herramienta indispensable para reducir la asimetría de información entre atacantes y defensores. Su valor trasciende la mera identificación de credenciales expuestas (como los 721.5 millones de accesos recuperados por SpyCloud en 2022) [2] y se extiende a capacidades estratégicas como:

- Detección temprana de activos comprometidos (servidores, dominios o dispositivos mal configurados).

- Monitoreo de menciones sospechosas en foros clandestinos o redes sociales que podrían indicar campañas de desinformación o suplantación de identidad.
- Identificación proactiva de vulnerabilidades asociadas a tecnologías utilizadas por la organización.
- Análisis de percepción pública para anticipar riesgos reputacionales.

La necesidad de estas capacidades queda demostrada en casos como el de Voova (Reino Unido), donde un exempleado accedió ilegalmente a sistemas críticos haciendo uso de credenciales y eliminó 23 servidores, causando pérdidas millonarias [3]. Incidentes de esta magnitud subrayan cómo solo se necesita un usurio con credenciales que puedan ser explotadas con consecuencias catastróficas. A esto se suman brechas masivas como la filtración de 267 millones de registros de usuarios de Facebook [4], que ilustran la escala de datos personales expuestos y su potencial uso malicioso.

El X-Force Threat Intelligence Index de IBM (2024) añade otra capa de urgencia: el 80% de los ataques modernos explotan identidades comprometidas [5], un vector que no requiere sofisticación técnica pero que puede mitigarse con OSINT efectivo. Para las PYMES, esta realidad es aún más crítica, ya que carecen de equipos dedicados para procesar manualmente el flujo constante de inteligencia de fuentes abiertas.

La paradoja actual es clara: mientras la cantidad de datos públicos útiles para la seguridad crece, también lo hace el ”ruido” que dificulta distinguir señales relevantes. Sin herramientas que automaticen la recolección, correlación y priorización de esta inteligencia, las organizaciones operan con visibilidad limitada, aumentando su exposición a riesgos evitables. Es aquí donde las soluciones Software as a Service (SaaS) basadas en OSINT, diseñadas para ser accesibles, escalables y accionables, se convierten en un componente esencial de cualquier estrategia de ciberseguridad moderna.

III. Objetivo

El propósito central de este proyecto es diseñar e implementar un sistema innovador basado en técnicas OSINT para apoyar las tareas de seguridad en una empresa PYME, con el objetivo de enriquecer su portafolio de servicios mediante una solución tecnológica que brinde a sus clientes la capacidad de identificar, analizar y gestionar de manera proactiva información sensible expuesta en internet. Esta herramienta está concebida para operar con un alto grado de automatización, eliminando la necesidad de personal especializado en ciberseguridad y democratizando el acceso a capacidades avanzadas de monitoreo digital.

La plataforma buscará generar valor tangible para las organizaciones clientes al proporcionar visibilidad sobre diversos riesgos digitales, incluyendo la exposición de credenciales en brechas de datos, menciones no autorizadas en redes sociales y medios digitales, vulnerabilidades críticas tecnológica, así como noticias sectoriales relevantes que puedan impactar su posición en el mercado. Más allá de la simple detección, el sistema estará diseñado para contextualizar esta información, priorizar amenazas según su nivel de criticidad y sugerir acciones concretas de mitigación.

IV. Objetivos específicos

El proyecto busca alcanzar los siguientes objetivos concretos para materializar la solución OSINT propuesta:

- Desarrollar un prototipo funcional que integre herramientas externas especializadas, permitiendo la automatización en la recopilación y análisis de información proveniente de fuentes abiertas.
- Diseñar e implementar un módulo especializado en la detección de credenciales expuestas, capaz de rastrear correos electrónicos, nombres de usuario, contraseñas y números de teléfono en bases de datos públicas, foros especializados y mercados clandestinos, proporcionando alertas tempranas sobre posibles brechas de seguridad.
- Crear un componente de monitoreo de noticias que analice automáticamente fuentes de información confiables para identificar amenazas y ataques relevantes según el sector de cada cliente, generando alertas contextualizadas y facilitando el acceso a las fuentes de información a los clientes
- Incorporar un sistema de identificación de vulnerabilidades en aplicaciones y tecnologías utilizadas comúnmente por los clientes, facilitando la detección oportuna de fallos críticos y apoyando la implementación de medidas preventivas antes de que sean explotadas.
- Establecer un mecanismo de priorización inteligente que clasifique los hallazgos según su nivel de criticidad, impacto potencial y relevancia para el negocio, simplificando la interpretación de riesgos y optimizando la toma de decisiones.
- Diseñar un generador de recomendaciones automatizadas que, basado en el tipo de hallazgo, sugiera acciones prácticas de mitigación accesibles incluso para usuarios sin conocimientos técnicos avanzados, cerrando así el ciclo de detección-acción.
- Garantizar una experiencia de usuario intuitiva mediante interfaces claras y visualizaciones comprensibles, asegurando que la plataforma sea accesible para pequeñas y medianas empresas, donde el conocimiento y experiencia en ciberseguridad suele ser limitada.
- Implementar un módulo de monitoreo de marca que rastree menciones no autorizadas, suplantaciones de identidad y fugas de información sensible en redes sociales y medios digitales, protegiendo la reputación corporativa de los clientes.
- Desarrollar un sistema de perfilamiento de empleados mediante técnicas de machine learning aplicadas a fuentes abiertas, identificando patrones de comportamiento que puedan representar riesgos internos o aportar inteligencia valiosa para estrategias de seguridad preventivas.

V. Propuesta de Valor

Con el fin de analizar la propuesta de valor, se van a analizar soluciones en el mercado para entender los pros y contras:

A. Herramientas disponibles en el Mercado

1) Herramientas OSINT: Shodan, SpiderFoot y Recon-ng son herramientas ampliamente utilizadas en el ámbito de la ciberseguridad, cada una con enfoques distintos. Shodan se especializa en escanear dispositivos y servicios expuestos en internet, identificando desde servidores vulnerables hasta sistemas IoT mal configurados, con costos que van desde los 59 dolares/mes para uso básico hasta planes empresariales personalizados. SpiderFoot, por su parte, es una suite de código abierto que automatiza la recolección de datos OSINT, ideal para análisis de amenazas y correlación de información, aunque requiere autoalojamiento y configuración manual. Recon-ng, una herramienta CLI(Interfaz de Línea de Comandos) preferida por pentesters, ofrece módulos para recopilar inteligencia de fuentes abiertas, pero su interfaz basada en comandos la hace inaccesible para usuarios no técnicos. A pesar de su potencia, estas herramientas comparten limitaciones importantes: requieren conocimientos avanzados para su configuración, no ofrecen guías claras de remediación y carecen de contextualización empresarial, dejando al usuario con datos crudos que debe interpretar por su cuenta.

Un problema recurrente con estas soluciones es que, si bien detectan riesgos (como credenciales filtradas o puertos abiertos), no explican cómo afectan específicamente al negocio del usuario ni proporcionan pasos concretos para mitigarlos. Por ejemplo, Shodan puede alertar sobre un servidor expuesto, pero no cómo protegerlo. SpiderFoot y Recon-ng, aunque flexibles, generan resultados que requieren análisis posterior por parte de expertos, algo que las PYMES rara vez tienen en su equipo. Además, ninguna de estas herramientas incluye monitoreo de amenazas sectoriales o noticias relevantes, dejando a las organizaciones sin visibilidad sobre riesgos emergentes en su industria.

2) Credenciales Expuestas: El mercado ofrece soluciones con un rango de precios que va desde gratuitas hasta aproximadamente $60 USD mensuales, destacándose por su facilidad de uso y variedad de parámetros de búsqueda. Sin embargo, la calidad de la información proporcionada varía significativamente: mientras algunas herramientas detallan el origen y fecha de las brechas (como filtraciones masivas en Twitter o Facebook), otras solo alertan sobre credenciales expuestas sin ofrecer contexto verificable, mostrando datos parcialmente censurados o sin evidencia concreta.

Un desafío crítico radica en cómo se presentan los resultados. La mayoría de las herramientas no generan reportes estructurados que permitan validar la fuente de los hallazgos, su criticidad o su distribución en distintas bases de datos, limitando la capacidad de los usuarios para priorizar acciones y mantener trazabilidad. Incluso se identificaron inconsistencias en los resultados, como herramientas que declaraban credenciales como seguras mientras otras las marcaban como comprometidas, lo que cuestiona su confiabilidad y utilidad práctica para la toma de decisiones informadas.

3) Análisis de Noticias: El ecosistema de noticias sobre ciberseguridad se divide en tres categorías principales, cada una con características distintivas. Los *medios generalistas* (como El Tiempo o El Espectador) cubren incidentes con lenguaje accesible, llegando a audiencias masivas, pero suelen carecer de profundidad técnica, usar titulares sensacionalistas y mezclar las noticias de seguridad en secciones amplias de tecnología. Si bien son útiles para alertar al público general, su falta de especialización limita su valor para la toma de decisiones técnicas.

Por otro lado, los *medios especializados* emplean terminología precisa y abordan temas como malware o vulnerabilidades con frecuencia y rigor. Aunque son ideales para profesionales, su enfoque técnico y la ausencia de contexto sectorial los hacen menos accesibles para empresas que buscan información aplicable a su industria específica.

Los *medios híbridos* (generalistas con secciones de seguridad) ofrecen un equilibrio, presentando amenazas digitales de manera comprensible sin sacrificar del todo el rigor. Sin embargo, su calidad es irregular: algunos artículos son sólidos, mientras otros son superficiales o alarmistas. Además, su cobertura es menos consistente que la de los especializados, con publicaciones que pueden espaciarse semanas entre sí. Para las PYMES, esta variabilidad representa un desafío, ya que dificulta la identificación de información fiable y relevante sin una curatoría experta.

4) Vulnerabilidades en Aplicaciones: El mercado ofrece soluciones avanzadas como Shodan, Censys e Intrigue, que proporcionan monitoreo continuo (24/7) de servicios expuestos, escaneo de puertos y correlación con vulnerabilidades conocidas. Estas plataformas, disponibles bajo suscripción, son ideales para empresas que requieren alertas automatizadas y capacidad de reacción proactiva. Sin embargo, su costo y complejidad las hacen inaccesibles para muchas PYMES.

Como alternativa, existen bases de datos gratuitas como Exploit Database, NIST NVD o CVE Details, que centralizan información sobre vulnerabilidades reportadas. Aunque son valiosas, su uso efectivo demanda conocimiento técnico para interpretar y contextualizar los datos, ya que no siempre es claro si una vulnerabilidad afecta a un entorno específico. Esto limita su utilidad para organizaciones sin equipos especializados.

Para las PYMES, esta brecha representa un riesgo crítico: la falta de herramientas accesibles que automaticen la detección, prioricen vulnerabilidades según su impacto real y guíen la remediación. Una solución efectiva debería combinar la exhaustividad de las bases de datos públicas con la usabilidad y contextualización que requieren las empresas con recursos limitados, cerrando así la brecha entre la identificación de riesgos y la acción práctica.

5) Análisis de Marca y Redes Sociales: El monitoreo de presencia digital se ha profesionalizado con herramientas como Mention, Awario o Mentionlytics, que ofrecen capacidades avanzadas: rastreo en tiempo real de menciones en redes sociales, foros y medios, generación de reportes automatizados y alertas personalizadas. Estas plataformas, con costos desde 30 dólares hasta 100 dólares mensuales, son ideales para empresas con operaciones globales que requieren escalabilidad

y análisis profundos. Sin embargo, su inversión las hace poco viables para PYMES con necesidades más acotadas.

Ante esto, muchas pequeñas empresas optan por métodos manuales: búsquedas directas en redes sociales y Google con palabras clave. Aunque evitan costos, esta aproximación consume recursos valiosos (tiempo del personal) y suele ser reactiva, incompleta o inconsistente. La falta de automatización expone a riesgos como respuestas tardías a crisis reputacionales, detección retrasada de suplantaciones de identidad o fugas de información en canales no monitorizados.

La brecha es clara: mientras las grandes empresas aprovechan inteligencia automatizada, las PYMES carecen de soluciones asequibles que equilibren cobertura básica, usabilidad y costos ajustados.

B. Valor Diferencial

En un mercado dominado por soluciones complejas y costosas diseñadas para grandes empresas, nuestro sistema surge como una alternativa que prioriza las necesidades reales de las pequeñas y medianas empresas. El valor diferencial radica en su capacidad para integrar inteligencia OSINT de múltiples fuentes, desde credenciales expuestas hasta amenazas reputacionales en redes sociales, en un solo sistema unificado, diseñado específicamente para usuarios sin formación técnica en ciberseguridad. A diferencia de herramientas fragmentadas que requieren configuración experta, nuestra solución automatiza los procesos más críticos: correlaciona datos dispersos, prioriza riesgos según el contexto específico de cada organización y genera recomendaciones claras y ejecutables.

Uno de los aspectos más innovadores es el enfoque en inteligencia accionable más que en simple recolección de datos. La plataforma no solo alerta sobre credenciales comprometidas o vulnerabilidades críticas, sino que guía paso a paso en el proceso de remediación, indicando exactamente qué acciones tomar y cómo implementarlas. El componente de machine learning añade una capa predictiva, analizando patrones de comportamiento digital para identificar riesgos potenciales antes de que materialicen. Esto es particularmente valioso para anticipar fugas de información o amenazas internas, aspectos tradicionalmente descuidados en soluciones para PYMES.

VI. Diseño de la solución

La solución propuesta se basa en una arquitectura desacoplada, organizada en capas independientes que facilitan la escalabilidad, mantenimiento, seguridad y extensibilidad futura del sistema. Esta arquitectura también está pensada para facilitar la integración de nuevos módulos.

A. Capas Principales

Capa de presentación (Frontend): Interfaz web a la que acceden los clientes para realizar consultas, visualizar resultados, reportes, alertas, modificar configuraciones y gestionar usuarios.

Capa de lógica de negocio: Implementa las reglas del sistema. Orquesta los microservicios, valida entradas, gestiona la lógica de análisis y determina los flujos según las reglas del sistema.

Capa de microservicios REST: Cada módulo opera como un servicio desacoplado expuesto a través de una API RESTFUL, lo que permite el mantenimiento independiente, reutilización y despliegue modular.

Capa de persistencia: Contiene las bases de datos y repositorios de almacenamiento de reportes, configuraciones, historiales, usuarios y cachés temporales.

B. Componentes del Sistema

Módulo de Autenticación y Gestión de Usuarios: Componente central garantiza acceso seguro mediante un sistema de autenticación robusto que incluye registro, inicio de sesión con múltiples factores, recuperación controlada de cuentas y cifrado avanzado de credenciales. Diseñado bajo principios de privacidad por defecto, almacena únicamente datos esenciales y registra toda actividad para auditoría, mientras asigna permisos mediante roles personalizables.

Módulo de Credenciales Expuestas: Este componente analiza posibles fugas de información mediante consultas automatizadas a múltiples fuentes OSINT, aplicando un sistema de verificación cruzada para validar hallazgos y minimizar falsos positivos. Cada resultado recibe un índice de confiabilidad que aumenta cuando aparece en varias bases de datos reconocidas, mientras se ajusta dinámicamente según el feedback histórico del cliente para refinar la precisión. El módulo procesa diversos datos (correos, dominios, teléfonos) y categoriza los riesgos por criticidad, generando un registro completo de consultas que permite monitorear evolutivamente las exposiciones. Su diseño combina automatización con aprendizaje continuo, priorizando resultados accionables y contextualizados para cada organización.

Módulo de Reportes: Este componente transforma datos técnicos en informes ejecutivos, priorizando hallazgos según su impacto real. Utiliza criterios de severidad como tipo de dato expuesto (contraseñas completas vs. parciales), antigüedad de la filtración y contexto de exposición para clasificar riesgos. Cada reporte incluye recomendaciones específicas de mitigación y se distribuye mediante la plataforma y correo electrónico, manteniendo un historial accesible que permite comparar evoluciones temporales. La presentación visual simplifica la interpretación para usuarios no técnicos, destacando únicamente información accionable.

Módulo de Noticias y Alertas Sectoriales: Diseñado para transformar datos complejos en inteligencia accionable, este módulo monitorea y filtra noticias de ciberseguridad según el perfil de cada organización. Utiliza técnicas avanzadas de crawling para analizar fuentes diversas (desde blogs técnicos hasta medios especializados), aplicando filtros de sector económico y ubicación geográfica para entregar solo información relevante. Las alertas se priorizan automáticamente, especialmente para industrias críticas como salud o finanzas, y se presentan en un dashboard intuitivo con resúmenes semanales. Su valor diferencial radica en traducir amenazas globales a riesgos específicos, permitiendo que equipos no

técnicos anticipen vulnerabilidades con base en tendencias sectoriales.

Módulo de Vulnerabilidades: Este componente identifica y prioriza vulnerabilidades en aplicaciones clave para cada cliente, adaptándose a sus patrones de uso reales para optimizar recursos. A diferencia de soluciones empresariales complejas, opera con un enfoque práctico: integra fuentes confiables como Exploit Database y NIST NVD, pero filtra la información según la tecnología utilizada por el usuario y su horario de operación, evitando alertas irrelevantes.

Las vulnerabilidades se clasifican automáticamente mediante criterios como puntuación CVSS, tipo de amenaza (ej: ejecución remota de código) y potencial impacto operacional. Los resultados se muestran en el dashboard con recomendaciones específicas (parches, configuraciones temporales), permitiendo incluso a equipos sin expertise técnico entender y actuar sobre riesgos críticos.

Módulo de Perfilamiento Predictivo: Este módulo utiliza Machine Learning para identificar patrones de riesgo específicos por sector, basándose en el análisis de credenciales expuestas y datos públicos. A diferencia de sistemas genéricos que monitorean actividades sospechosas, se enfoca en correlaciones sectoriales. Por ejemplo ”Sector financiero:Detecta mayor exposición en altos cargos” o ”Sector salud: Identifica empleados con +10 años en la organización como los más vulnerables a filtraciones”

Módulo de Monitoreo de Marca y Presencia Digital: Diseñado para proteger la reputación e integridad digital de las organizaciones, este módulo rastrea menciones de la marca, productos o ejecutivos en redes sociales (como facebook o X/twitter) y plataformas públicas. Utiliza palabras clave personalizadas para detectar desde intentos de suplantación hasta filtraciones de información sensible, siempre manejando los datos con confidencialidad absoluta diferenciando entre menciones rutinarias y aquellas que representan riesgos reales

Módulo de Interfaz Web para el Cliente: Este componente centraliza toda la funcionalidad del sistema en una plataforma web diseñada para democratizar el acceso a inteligencia de ciberseguridad. La interfaz prioriza la simplicidad sin sacrificar capacidades técnicas, permitiendo que usuarios sin formación especializada puedan monitorear riesgos y tomar decisiones informadas.

El dashboard principal ofrece una visión consolidada de hallazgos críticos, clasificados por nivel de urgencia (credentiales expuestas, vulnerabilidades activas, alertas reputacionales), con visualizaciones intuitivas que destacan tendencias y patrones. Cada módulo (desde perfilamiento predictivo hasta monitoreo de marca) mantiene una navegación unificada, con flujos guiados para acciones frecuentes.

VII. Arquitectura

El prototipo desarrollado tiene como objetivo principal la identificación automatizada de credenciales e información sensible expuesta en diversas fuentes abiertas. La arquitectura se diseñó bajo un enfoque modular, escalable y orientado a la eficiencia en la recolección y análisis de datos OSINT.

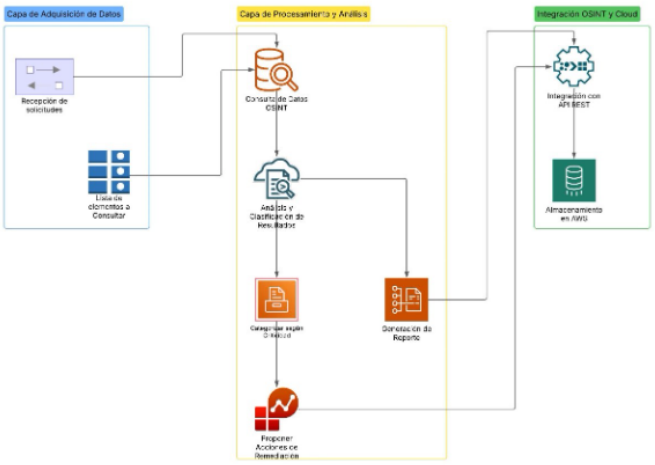

Fig. 1. Capas de la Arquitectura

En la primera fase del proceso (Fig1), el sistema recibe las solicitudes del usuario y puede trabajar tanto con entradas individuales como con listas de elementos a consultar. A partir de estas entradas, se inicia la capa de adquisición, la cual recopila información desde fuentes públicas como foros, bases de datos filtradas, portales especializados y otros repositorios accesibles sin autenticación.

Posteriormente, se ejecuta una consulta OSINT, seguida por una fase de análisis que permite clasificar los hallazgos obtenidos. Los resultados son categorizados con base en criterios de criticidad (alta, media o baja), utilizando reglas predefinidas o modelos de clasificación. Esta evaluación permite priorizar los riesgos detectados y generar recomendaciones de remediación adaptadas al contexto del hallazgo.

De forma paralela, el sistema genera un reporte estructurado con los hallazgos, el cual puede ser visualizado desde una interfaz web o exportado. Todo este proceso es gestionado a través de una API REST, que permite la integración con otros sistemas o flujos de trabajo externos. Esta capa facilita la consulta y administración de los datos de forma segura y estandarizada.

Finalmente, los datos son almacenados en una base de datos desplegada en infraestructura en la nube, específicamente sobre servicios de AWS. Este enfoque permite escalabilidad, alta disponibilidad y facilidad para futuras integraciones con sistemas de monitoreo, dashboards o motores de análisis más avanzados.

VIII. Alcance del proyecto

El presente proyecto tiene como alcance el diseño y desarrollo de un **prototipo funcional de un sistema OSINT**, orientado a la detección y análisis de información potencialmente expuesta en fuentes abiertas. Este sistema fue construido para facilitar a Resistance la identificación de credenciales comprometidas, vulnerabilidades asociadas a aplicaciones comunes, menciones en redes sociales y noticias relevantes sobre ciberseguridad.

El prototipo incluye una interfaz web interactiva, una API REST que permite la integración con herramientas externas, y un motor de clasificación de hallazgos basado en reglas simuladas, el cual evalúa la criticidad de los datos recolectados. Además, se implementaron funcionalidades para la generación automática de alertas y sugerencias de remediación, así como reportes estructurados por consulta.

El alcance se limita al desarrollo de un entorno simulado, utilizando fuentes representativas y técnicas básicas de análisis, sin integración directa con servicios en producción ni monitoreo en tiempo real. No se contempla, en esta etapa, la autenticación de usuarios, la implementación de un motor de IA entrenado en producción ni el almacenamiento persistente en bases de datos externas.

Esto responde a la necesidad de entregar un producto funcional en un periodo corto (febrero a mayo), priorizando la validación técnica del modelo y su aplicabilidad futura. Como parte del cierre del proyecto, se propone una hoja de ruta para la evolución del prototipo hacia una solución OSINT integral, escalable y automatizada.

IX. Implementación

En esta etapa se desarrollan componentes principales del backend y frontend del prototipo, se van integrando las funcionalidades necesarias para simular un sistema que pueda identificar credenciales expuestas en diferentes fuentes de internet. Dentro de las funcionalidades se encuentran también, una sección de noticias sobre actualidad de seguridad, categorizaciones de las vulnerabilidades encontradas y presencia en redes sociales.

El proceso se realiza de forma progresiva utilizando Python e integrando los componentes necesarios para el funcionamiento del prototipo OSINT. Estos módulos fueron desarrollados específicamente para la organización y contienen una lógica adaptable a sus necesidades, lo que incluye criterios de clasificación, métricas y visualización. A continuación, se presentan los componentes del sistema:

Componentes Desarrollados

Estos módulos fueron desarrollados desde cero con el objetivo de atender necesidades específicas de la organización. Su diseño se centro en la adaptabilidad, permitiendo realizar búsquedas dinámicas, evaluar el impacto de la información filtrada, clasificar los hallazgos según su nivel de criticidad y proponer acciones correctivas para mitigar posibles amenazas. Todo esto se integra dentro de un flujo de trabajo continuo.

Frontend (Formulario Web): Esta interfaz se diseña y se desarolla con HTML, Bootstrap y JavaScript. Permite al usuario ingresar hallazgos de forma accesible y rápida desde cualquier navegador. Cuenta con un entorno claro e intuitivo para el usuario, incluyendo diferentes ejemplos para realizar búsquedas; además de mostrar los resultados para un posterior análisis.

Menú Principal: Administra el control principal de la API, definiendo las rutas principales, maneja solicitudes, administra los módulos internos y coordina la generación del resultado.

Clasificador de criticidad: Simula IA para priorizar riesgos según el contenido recibido. Implementa la clasificación de criticidad usando un modelo de IA previamente entrenado (algoritmo de clasificación básica usando scikit-learn y joblib).

Módulo de Alertas: Es el soporte inicial de decisión. Genera una alerta textual en el front de la API, que está basada en el nivel de criticidad de acuerdo con la información que recibe.

Módulo de Recomendaciones: Funciona como una guía de respuesta. Una vez se identifica el tipo de información expuesta, esta sección de API, muestra diferentes sugerencias para remediar las vulnerabilidades.

Módulo de búsqueda expuesta: Realiza un proceso de simulación para la detección de credenciales expuestas en fuentes OSINT. Usa regex o APIs para identificar correos, usuarios, contraseñas, dominios filtrados.

Módulo de noticias/panorama: Consulta amenazas actuales o simula amenazas sectoriales desde fuentes tipo RSS.

Módulo de vulnerabilidades: Simula amenazas globales agrupandolas por tipo e impacto. Evalúa software común con CVEs críticos para alertar sobre vulnerabilidades actuales.

Módulo de monitoreo social: Funciona como un monitor de marca OSINT. Recolecta menciones de la organización en redes sociales o foros OSINT, lo que permite medir la reputación de la organización.

Modelo ML simulado: Emula decisiones inteligentes con base en características básicas (tipo de dato, fuente, longitud, etc.).

Logs locales: Permite trazabilidad, auditoría de entradas, futuras mejoras en calidad de resultados.

Uso de Componentes Externos

Las sigueintes herramientas permiten acelerar el desarrollo, mejorar el uso de código, asegurar la compatibilidad multiplataforma, permitir un futuro escalamiento o diferentes mejoras sobre las capacidades de los módulos generados, incluyendo el proceso de Inteligencia Artificial para el proceso de aprendizaje de la maquina:

API FastAPI: Actúa como núcleo de lógica, integrando todos los módulos OSINT mediante rutas claras y eficientes; es decir, los puntos finales (*endpoints*) optimizados para dar respuesta rápidamente a las solicitudes del usuario, gestionando adecuadamente los datos de entrada y salida, para garantizar un rendimiento ágil, incluso con múltiples peticiones simultáneas. FastAPI, además, permite el desarrollo de API RESTful usando Python, facilitando la integración con servicios externos mediante estándares como HTTPS y JSON.

Jinja2: El motor de plantillas para construir las vistas del frontend directamente desde FAastAPI. Permite generar documentos en HTML de forma rápida y sencilla.

Bootstrap5: Permite la creación de sitios web de forma fácil y sencilla. Utiliza un conjunto de componentes predefinidos, estilos y plantillas en HTML, CSS y JavaScript.

Chart.js: Librería de JavaScript de código abierto que facilita la creación de gráficos en aplicaciones web. Permite dibujar diferentes elementos gráficos como barras, líneas, áreas circulares, radar, etc.

Recurso	Descripción
Python / FastAPI	Backend modular y liviano para prototipo
HTML / Bootstrap 5	Interfaz simple y responsive
Podman / Docker	Contenedor para pruebas y despliegue local
Modelo ML simulado	Clasificación básica para hallazgos
Herramientas OSINT simuladas	Ejemplos de leaks, redes, menciones, etc.
Git / GitHub	Control de versiones
PC local / VSCode	Entorno de desarrollo
Fuentes públicas (simuladas)	Have I Been Pwned, NVD, Reddit, etc.

Fig. 2. Tabla de recursos para el desarrollo

Criterio de desempeño	Descripción	Valor observado	Unidad	Observación
Tiempo de respuesta promedio	Tiempo desde el envío hasta la respuesta	1.2	segundos	Rápido en entorno local, ideal para alertas rápidas
Precisión de clasificación IA	Porcentaje de hallazgos clasificados correctamente	93	%	Evaluado sobre datos simulados
Cobertura funcional	Módulos activos vs módulos definidos	100	%	Todos los módulos esperados fueron implementados
Carga de noticias al inicio	Noticias cargadas automáticamente	05-jun	unidades	Varía según fuente simulada
Rendimiento con múltiples consultas	Capacidad de procesar varias entradas seguidas	Sin errores	-	Pruebas de 10 iteraciones consecutivas exitosas
Integridad del historial	Registro correcto de cada entrada procesada	100	%	Visualización correcta en dashboard
Visualización de gráficas	Renderizado de torta y línea	Correcto	-	Se ejecutan desde frontend sin bloqueo

Fig. 3. Desempeño del modelo

Html2canvas: Al igual que Chart.js es una librería de JavaScript. Permite convertir el contenido de una página web HTML a un formato normalmente JPG o PNG. Esto se logra creando una representación del "Document Object Model" en un formato HTML5 (canvas) y luego exportarlo como una imagen.

Podman / Docker: Plataformas de código abierto, que permiten crear, ejecutar y gestionar aplicaciones dentro de contenedores. Facilitan el desarrollo y la ejecución de aplicaciones, haciéndolas más portables y eficientes.

Joblib + scikit-learn: es una biblioteca de código abierto en Python utilizada para el aprendizaje automático. Es una de las bibliotecas más populares para implementar modelos de machine learning y análisis de datos. La combinación de ambos permite el aprendizaje del modelo de inteligencia artifical que fue entrenado con un conjunto de ejemplos y casos simulados, representando distintos escenarios de filtración de datos y amenazas comunes, incluyendo correos electrónicos expuestos, contraseñas filtradas, mensajes de suplantación y referencias a campañas de phishing. Estos datos fueron estructurados en categorías según niveles de riesgo (alto, medio y bajo) para permitir que el modelo aprendiera a identificar patrones y priorizar amenazas automáticamente.

X. Desempeño del Modelo

Se realizaron pruebas funcionales, de integración y rendimiento sobre los principales módulos del sistema. Para cada uno se definieron entradas esperadas y se validaron las salidas, verificando que el comportamiento cumpliera los requisitos funcionales definidos. Se empleó el estándar **IEEE 829-2008** para estructurar los casos de prueba y la norma **ISO/IEC 25010** para justificar las propiedades evaluadas como fiabilidad, eficiencia, usabilidad y mantenibilidad. Además, se comprobaron los tiempos de respuesta del sistema en la clasificación de hallazgos y generación de reportes, garantizando la viabilidad de su uso en escenarios reales.

Durante la evaluación del prototipo, se midieron aspectos clave del desempeño como el tiempo de respuesta, la precisión de los módulos de clasificación, la capacidad de carga y la visualización de resultados (fig. 3). Los resultados indican que el sistema es capaz de responder de forma rápida y precisa, manteniendo la integridad del historial, generando visualizaciones efectivas e integrando todos los módulos definidos para esta versión inicial. La precisión del modelo de clasificación alcanzó un **93% en pruebas controladas**, lo cual es adecuado para el tipo de hallazgos esperados. Así mismo, el tiempo de respuesta medio por solicitud fue inferior a **1.5 segundos**, demostrando un rendimiento eficiente en el entorno actual de pruebas.

Estas métricas validan que el prototipo cumple satisfactoriamente los requisitos mínimos para una primera fase de despliegue, y sientan las bases para futuras optimizaciones en producción.

XI. Valores Esperados del Modelo

Además de evaluar el funcionamiento correcto y técnico del sistema, se evalúa su capacidad para cumplir con los propósitos estratégicos establecidos en su diseño: como la correlación de información dispersa, priorización de los riesgos según el contexto organizacional y generar las diferentes recomendaciones, que sean claras y accionables.

*1) **Correlación de Información Dispersa**:* El sistema integra múltiples fuentes OSINT (foros, filtraciones simuladas, fuentes de vulnerabilidades y redes sociales) y puede ser capaz de **consolidar hallazgos** en una vista unificada. Por ejemplo, ante una entrada que contiene una cuenta de correo expuesta y un dominio de la organización, el sistema puede detectar tanto el riesgo de suplantación como posibles menciones en foros, mostrando ambos resultados en el dashboard, asociados a un solo hallazgo.

*2) **Priorización Inteligente del Riesgo**:* Cada hallazgo sería evaluado con base en su contexto (tipo de dato, exposición, fuente) utilizando reglas internas y un modelo de clasificación basado en machine learning. El sistema puede ser capaz de asignar una *criticidad coherente con el impacto potencial*, incluso en casos donde las menciones se hacen sin contexto técnico, que serían clasificadas como riesgo bajo o informativo. Por ejemplo, sí se ingresa el "nombre de un directivo en grupo cerrado de Telegram", la criticidad que se asigna es media y la justificación sería "fuente pública no indexada" más el posible riesgo reputacional.

*3) **Recomendaciones Ejecutables**:* Cada resultado es acompañado de sugerencias para la remediación, alineadas con el tipo y nivel de amenaza detectada. Se valida que las recomendaciones sean **entendibles por usuarios no técnicos**, usando un lenguaje claro y acciones prácticas. El resultado ideal sería que el 100% de los hallazgos críticos incluyan

Componente Evaluado	Resultado alcanzado
Clasificación de criticidad	100% de los casos de prueba clasificados correctamente
Generación de sugerencias	Respuestas contextualizadas según nivel de criticidad
Detección de credenciales	Reconocimiento de correos, usuarios, teléfonos y dominios
Visualización de noticias	Noticias simuladas cargadas dinámicamente al inicio
Identificación de vulnerabilidades	CVEs actuales mostrados en un formato amigable
Panorama de amenazas (general y sectorial)	Integración clara por criticidad y tipo
Análisis en redes sociales	Resultados por plataforma y nivel de exposición
Generación de dashboard e historial	Historial funcional con navegación y visualización

Fig. 4. Evaluación de componentes

al menos 3 recomendaciones prácticas y el 87% de las recomendaciones se consideraren implementables.

*4) **Análisis Contextual y Sectorial**:* A través del módulo de noticias y panorama, el sistema puede ser capaz de entregar **inteligencia contextual** basada en amenazas del sector y noticias recientes. Este análisis permitiría que los hallazgos no sean interpretados de forma aislada, sino dentro de un entorno de amenazas actualizado, lo que refuerza la capacidad predictiva y estratégica del prototipo.

Ejemplo:

- Entrada: "filtración de credenciales en portal de salud"
- Contexto adicional: alerta sector salud por campañas activas de ransomware (proporcionada por módulo de noticias).
- Valor agregado: aumento de criticidad y refuerzo de recomendaciones de contención inmediata.

XII. Resultados

El sistema es capaz de simular una interacción real con entradas relacionadas con exposición de información, evaluarlas con modelos básicos de IA entrenados previamente, identificar patrones sensibles (como credenciales) y mostrar de forma ordenada el nivel de amenaza, sugerencias y contexto. Esto indica que el modelo funciona como una herramienta de detección temprana y monitoreo. Al ser un prototipo base, permite a futuro convertirse en un servicio más completo. Los diferentes modulos permiten la escalabilidad y que pueda ser integrado con APIs externas o SIEM en el futuro. (fig. 4)

Como resultado del desarrollo, se obtuvo un **prototipo funcional operativo en entorno local**, accesible vía navegador, que permite ingresar hallazgos manuales y clasificarlos según su nivel de criticidad. El sistema incluye una API REST para su integración, un clasificador básico simulado, y módulos complementarios que extraen información contextual, como noticias de ciberseguridad, vulnerabilidades activas y menciones en redes sociales.

Como parte del diseño e implementación del sistema OSINT desarrollado, se consideraron buenas prácticas de privacidad y tratamiento ético de la información, alineadas con los marcos de referencia **NIST Privacy Framework** e **ISO/IEC 27701**. **No se utilizan técnicas de intrusión, fuerza bruta ni ingeniería social**, y **limita su análisis únicamente a fuentes de información pública accesibles sin autenticación**. Por otro lado, se implementa un principio de **minimización de datos**, recolectando solo lo estrictamente necesario para el análisis. **Los datos sensibles identificados (como correos o credenciales)** se muestran anonimizados en el reporte de salida o acompañados de su fuente pública verificada. **Toda la actividad queda registrada** para propósitos de auditoría y trazabilidad, sin almacenar datos personales sin justificación técnica. Se prevé un proceso de **eliminación segura** de los datos almacenados, según políticas de retención y limpieza documentadas.

XIII. Pasos a futuro

Optimización y Escalabilidad

- Una vez implementado en AWS, monitorear el consumo de recursos (CPU, almacenamiento, ancho de banda) para identificar oportunidades de optimización sin comprometer el rendimiento.
- Ajustar la infraestructura según patrones de uso reales (ej.: escalamiento automático en horarios pico) para equilibrar costos y eficiencia.

Avances para ML

- Postergar el desarrollo avanzado de Machine Learning hasta contar con datos reales de clientes, que permitan entrenar modelos con casos concretos y patrones sectoriales.
- Iniciar con un modelo base simple y enriquecerlo progresivamente, usando métricas de precisión y retroalimentación directa para evitar sesgos o falsos positivos.

XIV. Conclusiones

Proyectos como este representan un paso crucial hacia la democratización de la ciberseguridad para PYMES, ofreciendo una solución que combina capacidades avanzadas de inteligencia OSINT con una usabilidad accesible para equipos no técnicos. Al integrar detección de credenciales expuestas, monitoreo de vulnerabilidades y análisis de riesgos reputacionales en una plataforma unificada, se proporciona a las organizaciones con recursos limitados la capacidad de anticipar y mitigar amenazas de manera proactiva. El verdadero valor diferencial radica en transformar datos técnicos complejos en recomendaciones claras y acciones concretas, cerrando la brecha entre la identificación de riesgos y la capacidad de respuesta efectiva.

A medida que el sistema evolucione con la implementación en AWS y el entrenamiento de modelos de machine learning con datos reales, su precisión y valor para los usuarios aumentará significativamente. Este desarrollo no solo mejorará la postura de seguridad de las PYMES, sino que también contribuirá a crear una cultura de prevención en un entorno digital cada vez más hostil. La plataforma está diseñada para crecer junto con las necesidades cambiantes del mercado, asegurando

su relevancia continua como herramienta estratégica para la protección del negocio en la era digital.

REFERENCES

[1] Crowdstrike. (2024). 2024 GLOBAL THREAT REPORT

[2] Prensa CambioDigital OnLine. (2023, 14 marzo). Más de 700 millones de credenciales expuestas y 22 millones de dispositivos infectados en 2022.

[3] Esage G, A. (2019, 26 marzo). Empleado de TI es despedido y elimina los servidores en la nube AWS de su empresa.

[4] Vanguardia. (2019, 20 diciembre). Facebook sufre una filtración de datos de 267 millones de usuarios expuestos en Internet.

[5] IBM Corporation. (2024a). X-Force Threat Intelligence Index 2024.

[6] NIST. (2020, 16 enero). NIST PRIVACY FRAMEWORK: A Tool For Improving Privacy Through Enterprise Risk Management.