UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR

POSGRADOS

MAESTRÍA EN
SEGURIDAD DE LA INFORMACIÓN
RPC-SO-28-NO.669-2021

OPCIÓN DE TITULACIÓN:
PROYECTO DE TITULACIÓN CON COMPONENTES DE INVESTIGACIÓN APLICADA Y/O DE DESARROLLO

TEMA:
DISEÑO Y EVALUACIÓN DE ESTRATEGIAS DE DEFENSA CIBERNÉTICA UTILIZANDO INTELIGENCIA ARTIFICIAL EN ENTORNOS DE LA INDUSTRIA 4.0.

AUTOR:

JOSÉ ANDRES TENEN QUIZHPI

DIRECTOR:

JUAN CARLOS DOMÍNGUEZ AYALA

CUENCA – ECUADOR
2025

Autor:

José Andres Tenen Quizhpi
Ingeniero en Sistemas.
Candidato a Magíster en Seguridad de la Información por la Universidad Politécnica Salesiana – Sede Cuenca.
jtenen@est.ups.edu.ec

Dirigido por:

Juan Carlos Domínguez Ayala
Ingeniero en Sistemas de Información.
Ingeniero en Telecomunicaciones.
Magíster en Seguridad de la Información.
jdominguez@ups.edu.ec

DEDICATORIA

Dedico este proyecto a mi familia, por estar siempre a mi lado en cada etapa de mi formación. A mis padres, gracias por confiar en mí incluso cuando yo dudaba, y por motivarme a seguir adelante sin rendirme. Deseo expresar mi gratitud a los profesores y mentores que, con su ejemplo y disposición para compartir lo que saben, me acompañaron en este camino y me hicieron crecer tanto en lo académico como en lo personal. Su pasión por enseñar fue siempre un motor de motivación para mí. También quiero agradecer a mis compañeros y amigos: por la complicidad, por las horas de trabajo compartidas y por estar ahí en los momentos más duros. Esta experiencia no solo nos formó en lo profesional, sino que nos dejó un lazo que va más allá de lo académico.

AGRADECIMIENTO

Antes que nada, le agradezco a Dios por darme la vida, la salud y la fuerza para alcanzar esta meta. Cada paso en este camino ha sido posible gracias a esa guía invisible pero constante. A mi familia, en particular a mis padres y hermanos, les debo más de lo que estas palabras pueden expresar. Su apoyo incondicional, sus consejos oportunos y su paciencia en los momentos más difíciles han sido clave para lograr finalizar esta etapa. Al Ingeniero Juan Carlos Domínguez, agradezco su presencia, su conocimiento y su paciencia para acompañarme en todo el transcurso de este proyecto. Además, deseo agradecer a los docentes de la Universidad Politécnica Salesiana, quienes durante el programa transmitieron no solo sus saberes, sino también su vocación y dedicación, contribuyendo a expandir mi perspectiva profesional. A mis compañeros y amigos de la maestría, agradezco la cooperación, los diálogos, los trabajos en equipo y el respaldo recíproco. Esta etapa no habría sido igual sin su apoyo, y los vínculos que formamos van más allá del aula. Finalmente, quiero expresar mi agradecimiento a la Universidad Politécnica Salesiana por brindarme esta oportunidad de educación. Voy con conocimientos que van más allá de lo académico, y con valores que me acompañarán tanto en lo personal como en lo laboral.

Tabla de Contenido

DISEÑO Y EVALUACIÓN DE ESTRATEGIAS DE DEFENSA CIBERNÉTICA UTILIZANDO INTELIGENCIA ARTIFICIAL EN ENTORNOS DE LA INDUSTRIA 4.0.

AUTOR(ES):

JOSÉ ANDRES TENEN QUIZHPI

Resumen

La presente tesis titulada "Diseño y Evaluación de Estrategias de Defensa Cibernética Utilizando Inteligencia Artificial en Entornos de la Industria 4.0" tiene como objetivo desarrollar un sistema de ciberdefensa automatizada basado en inteligencia artificial que permita identificar y responder ante amenazas cibernéticas en infraestructuras industriales. El estudio parte de una revisión detallada de las amenazas más comunes en entornos industriales interconectados y de la selección de técnicas avanzadas de IA como las redes neuronales. La metodología aplicada incluye la creación de un entorno virtualizado que simula una red de Industria 4.0, donde se ejecutaron ataques reales y se recolectó tráfico de red con Zeek para entrenar un modelo de aprendizaje profundo. Este modelo fue validado con métricas cuantitativas, obteniendo una precisión cercana al 100 % y una capacidad efectiva de respuesta automática. Los resultados demostraron la viabilidad de integrar IA en sistemas de seguridad industrial, mejorando la detección y mitigación de amenazas en tiempo real, cumpliendo con estándares internacionales y ofreciendo una solución adaptable y escalable para entornos críticos.

Palabras clave:

Industria 4.0, ciberseguridad industrial, inteligencia artificial, redes neuronales, tráfico de red, detección de amenazas, Zeek, respuesta automatizada.

Abstract

This thesis, entitled "Design and Evaluation of Cyber Defense Strategies Using Artificial Intelligence in Industry 4.0 Environments", aims to develop an automated cybersecurity system based on artificial intelligence to detect and respond to cyber threats within industrial infrastructures. The study begins with a thorough review of the most common threats in interconnected industrial environments and the selection of advanced AI techniques, particularly neural networks. The methodology involved the creation of a virtualized environment simulating an Industry 4.0 network, where real cyberattacks were executed and network traffic was captured using Zeek to train a deep learning model. This model was validated using quantitative metrics, achieving near-perfect accuracy and demonstrating effective automatic response capabilities. The results confirmed the feasibility of integrating AI into industrial security systems, improving real-time threat detection and mitigation, complying with international standards, and offering an adaptable and scalable solution for critical environments.

Key Words:

Industry 4.0, industrial cybersecurity, artificial intelligence, neural networks, network traffic, threat detection, Zeek, automated response.

1. Introducción

La Cuarta Revolución Industrial ha promovido un cambio digital profundo en los procesos de producción a través de tecnologías como el Internet de las cosas (IoT), el big data, la computación en la nube y el aprendizaje automático asi originando el concepto de Industria 4.0 o producción inteligente (García Ortega, 2021). Si bien estos cambios han mejorado la eficiencia y la productividad también han dejado a las empresas industriales vulnerables ante amenazas cibernéticas que cada vez son más avanzadas (Serror, 2020).

Varios informes como el Cost of a Data Breach Report 2024 de IBM Security, muestran que el 55% de las industrias 4.0 han experimentado incidentes severos de ciberseguridad, causando así pérdidas que superan los 4,88 millones de dólares y perjudicando directamente la continuidad de las operaciones (Security, 2024).

En estudios recientes sobre las amenazas que afectan a los entornos industriales se han identificado y examinado exhaustivamente numerosas formas de ciberataques, incluidos ransomware, malware, phishing, ataques DoS y robo de información privada. (Rahman, 2023). Su regularidad y nivel de sofisticación han aumentado, lo que ha demostrado cuán vulnerables son muchas infraestructuras esenciales a este tipo de amenazas.

Dada la situación actual, algunos expertos han señalado que es importante usar métodos de ciberseguridad más rápidos y que permitan anticiparse mejor a los problemas. También mencionan que no se debe proteger solo una parte, sino incluir tanto los sistemas de Tecnología Operacional (OT) como los de Tecnología de la Información (IT). Para lograrlo, una opción sería aplicar medidas que combinen varias capas de seguridad, el uso de cifrado en los datos y herramientas que ayuden a monitorear constantemente para detectar cualquier problema a tiempo. (Mosteiro-Sanchez, 2020).

En los últimos años, se ha empezado a usar inteligencia artificial y aprendizaje profundo dentro de las estrategias de ciberseguridad. Esto se debe a que estas tecnologías pueden analizar una enorme cantidad de datos y, gracias a eso, detectar comportamientos extraños o inesperados, algo que con los métodos tradicionales cuesta mucho más. Según Becher y Torka, los sistemas de seguridad actuales ya tienen incorporado estos algoritmos para dejar atrás los enfoques basados solo en reglas fijas, ya que así se puede reaccionar mejor frente a amenazas que cambian constantemente (Becher, 2024). Por otro lado, Goyal señala que este tipo de tecnologías representan una gran oportunidad para proteger mejor las infraestructuras industriales, sobre todo en ambientes donde todo está cada vez más conectado(Goyal, 2023).

Con base en eso, esta tesis propone diseñar y probar un sistema de ciberdefensa usando inteligencia artificial, pensado para detectar y responder ante amenazas en entornos industriales simulados. Para eso, se va a analizar primero qué tipos de ataques son los más comunes, luego se elegirán las técnicas de IA más apropiadas, se desarrollará una estructura modular y se harán pruebas en condiciones controladas. La idea es que este sistema sea adaptable, escalable y cumpla con los estándares internacionales, con el fin de reforzar la seguridad de las infraestructuras industriales más críticas.

2. Determinación del Problema

En este capítulo se examina detalladamente el problema que impulsa a esta investigación, tratando los riesgos cibernéticos más importantes que impactan a los entornos industriales interconectados además de la importancia de implementar nuevas tácticas de protección fundamentadas en inteligencia artificial.

La revolución industrial 4.0 ha impulsado la implementación a gran escala de tecnologías que están en auge como el Internet de las Cosas (IoT), la inteligencia artificial, la computación en la nube y los sistemas ciberfísicos. No obstante, este cambio ha incrementado significativamente la superficie de ataques cibernéticos, dejando a las organizaciones industriales bastante vulnerables a amenazas cada vez más avanzadas.

El Cisco Cybersecurity Readiness Index 2024 evidenció que solo el 3 % de las organizaciones están en un nivel de preparación madura para enfrentar ciberamenazas, mientras que el 71 % aún se encuentran en estados formativos o iniciales (Cisco, 2024). Este bajo nivel de preparación incrementa la vulnerabilidad frente a ataques que cada vez son más automatizados, especialmente aquellos que utilizan inteligencia artificial generativa para perfeccionar los ataques de phishing y malware.

Además, el State of Industrial Networking Report 2024 reveló que el 89% de las entidades industriales percibe la ciberseguridad como su mayor reto en el presente sin embargo únicamente una pequeña minoría cuenta con una coordinación eficaz entre los equipos de IT y OT, lo que causa una fragmentación en las estrategias de defensa (Cisco, 2024).

En cuanto a las amenazas internas, el 2024 Insider Threat Report de Gurucul mostró que el 48% de las entidades han experimentado un incremento en incidentes de amenazas internas y que el 29% de estas han tenido que lidiar con costos de remediación que superaron el millón de dólares (Cybersecurity Insiders & Gurucul,

2024). Estas amenazas son impulsadas principalmente por entornos de TI más complejos también por la falta de controles unificados y por la capacitación insuficiente del personal.

Por su parte, el X-Force Threat Intelligence Index 2025 confirma que la industria manufacturera es el sector más atacado por cuarto año consecutivo, representando el 26 % de los incidentes globales. Además, se ha identificado que aproximadamente un 30 % de los casos, los atacantes lograron ingresar utilizando credenciales legítimas que habían sido previamente comprometidas. Este tipo de acceso revela un nivel de sofisticación cada vez mayor en las tácticas empleadas para vulnerar sistemas, especialmente en la etapa inicial de los ataques, donde el sigilo y la eficacia resultan determinantes (IBM, 2025).

Por otro lado, el Data Breach Investigations Report 2025 destacó que el 44 % de los incidentes de seguridad estuvieron relacionados con ataques de ransomware. Además, remarcó que uno de los métodos de entrada más comunes sigue siendo el uso indebido de credenciales de acceso. (Verizon, 2025). Estos datos ponen en evidencia que, dentro de los entornos industriales actuales, los ataques más frecuentes y peligrosos suelen tener como punto de entrada el uso indebido de credenciales y la expansión de software malicioso tipo ransomware.

Finalmente, a nivel global, el The Global Risks Report 2024 detectó que la ciberseguridad es uno de los cinco riesgos globales más importantes para los próximos dos años junto con la desinformación y la polarización social, alertando que la ausencia de una gobernanza tecnológica coordinada puede incrementar los peligros de ciberataques a gran escala (World Economic Forum, 2024).

En respuesta a este panorama alarmante, se plantea la necesidad urgente de diseñar estrategias de ciberdefensa más proactivas, inteligentes y automatizadas, especialmente en entornos industriales donde la continuidad operativa depende directamente de la protección cibernética de sistemas críticos.

3. Marco teórico referencial

Para entender los desafíos que tienen hoy las industrias es importante saber bien qué es la Industria 4.0 cómo se cuida la seguridad digital y cómo ayuda la inteligencia artificial Este capítulo explica de manera simple y ordenada las ideas más importantes que sirven para proteger los sistemas modernos en un mundo cada vez más conectado.

Primero se cuenta qué hace diferente a la Industria 4.0 y cómo las tecnologías emergentes están cambiando la forma en que funcionan las fábricas y los procesos. También se mencionan algunas herramientas que permiten que todo esté más automatizado y conectado.

Después se habla de la seguridad digital en estos entornos se explican los peligros más comunes y lo que se puede hacer para evitarlos, las reglas que se siguen en todo el mundo y por qué es tan importante cuidar estos sistemas.

Posteriormente, se analiza cómo la inteligencia artificial contribuye a fortalecer la seguridad, mediante la detección de comportamientos inusuales en la red, el aprendizaje a partir de incidentes previos y la capacidad de actuar de forma automática frente a eventos sospechosos. Además, se abordan algunas de las técnicas más avanzadas que ya están siendo aplicadas en entornos industriales bajo el enfoque de la Industria 4.0.

Al final se repasan algunos casos donde estas ideas ya se están aplicando donde se habla de qué resultados se lograron qué estrategias se usaron y qué cosas todavía faltan por mejorar. Esta parte ayuda a ver cómo todo esto funciona en la vida real y qué caminos se pueden seguir para seguir cuidando las industrias en esta nueva etapa digital.

3.1 Industria 4.0: Concepto y características

La Industria 4.0, o la Cuarta Revolución Industrial, simboliza un cambio radical en los procesos de producción y manufactura caracterizada por la integración de tecnologías digitales avanzadas (Peralta-Abarca, Martínez-Bahena, & Enríquez-Urbano, 2020). La Industria 4.0 surgió en 2011, en el marco de la Feria de Hannover, en Alemania. Surgió como una estrategia para modernizar la industria, apostando por la digitalización, la automatización y el uso de sistemas ciberfísicos que vinculan el mundo tangible con el mundo virtual. (Meindl & Mendonça, 2021).

Las características principales de la Industria 4.0 incluyen:

- **Digitalización y automatización de procesos:** La Industria impulsa la informatización de la producción usando tecnologías digitales. Gracias a esto, se pueden automatizar y optimizar muchos procesos dentro de las fábricas, haciendo que todo funcione de forma más eficiente y conectada (Peralta-Abarca, Martínez-Bahena, & Enríquez-Urbano, 2020).
- **Conectividad y comunicación en tiempo real:** Con la ayuda de redes avanzadas y del del Internet de las Cosas Industriales (IIoT), las máquinas, los sistemas y las personas pueden estar conectados todo el tiempo, lo que permite una comunicación fluida y en tiempo real dentro del entorno industrial (Meindl & Mendonça, 2021).
- **Personalización de productos y servicios:** Los sistemas productivos en el sector de la Industria 4.0 son tan flexibles que pueden adaptarse fácilmente a lo que cada cliente necesita haciendo posible ofrecer productos personalizados a gran escala (Peralta-Abarca, Martínez-Bahena, & Enríquez-Urbano, 2020).
- **Optimización de la cadena de valor:** La digitalización de la cadena de suministro permite mejorar la planificación, producción y distribución de productos, dando lugar a procesos más eficientes, integrados y mejor sincronizados a lo largo de toda la operación. (Peralta-Abarca, Martínez-Bahena, & Enríquez-Urbano, 2020).

- **Sostenibilidad y eficiencia energética:** La Industria 4.0 impulsa el uso responsable de los recursos, promoviendo prácticas que reducen el impacto ambiental y mejoran la eficiencia energética en los procedimientos industriales (Meindl & Mendonça, 2021).

En resumen, la Industria 4.0 simboliza una evolución hacia sistemas de producción inteligentes, adaptables y eficientes, impulsados por la convergencia de tecnologías digitales y la interconexión masiva de dispositivos.

3.2 Ciberseguridad en entornos de Industria 4.0

La Industria 4.0 ha transformado por completo la estructura en que funcionan los procesos en las fábricas, gracias al uso de tecnologías emergentes digitales como el internet de las cosas en la industria, la inteligencia artificial y los sistemas que conectan el mundo físico con el digital. Pero con todos estos avances también han aparecido nuevos riesgos porque al estar más conectadas las industrias quedan más expuestas a ataques cibernéticos. Ahora que la tecnología que se usa para manejar la información y la que controla las máquinas están trabajando juntas eso ha traído retos nuevos para la seguridad digital. Por eso se necesitan formas especiales de proteger todo lo que está conectado y evitar que alguien pueda atacar o causar daños en sistemas que son clave para el funcionamiento de una empresa o de servicios importantes (Alqudhaibi, Albarrak, Jagtap, Williams, & Salonitis, 2025).

3.2.1 Riesgos y amenazas emergentes

La conexión entre equipos y sistemas en la Industria 4.0 ha ampliado significativamente las vulnerabilidades, permitiendo que actores maliciosos exploten debilidades en la cadena de suministro, en los sistemas de regulación industrial y en las redes de comunicación. Entre los riesgos más destacados se encuentran el ingreso no permitido a datos sensibles, la suspensión de procesos productivos, el sabotaje de infraestructuras críticas y la alteración de parámetros operativos mediante ciberataques sofisticados (Pedreira, Barros, & Pinto, 2021).

3.2.2 Estrategias de defensa y mejores prácticas

Para poder reducir estos riesgos se han propuesto varias formas de proteger los sistemas industriales, pero una de las más recomendadas es la defensa en profundidad que básicamente es poner varias capas de seguridad desde afuera hacia adentro para que en el caso de que algo falle, no todo esté expuesto y al atacante se le haga mucho más difícil avanzar.

También es clave usar cifrado de extremo a extremo eso ayuda a que los datos estén seguros tanto cuando se están enviando como cuando ya están guardados además es crucial mantener un control sobre quién ingresa y qué puede hacer dentro del sistema por eso se usa la gestión de identidades y accesos y, por último, pero no menos importante está el monitoreo constante que sirve para detectar cualquier actividad sospechosa y reaccionar de inmediato si algo pasa. (Mosteiro-Sanchez, Barcelo, Astorga, & Urbieta, 2020).

3.2.3 Normas internacionales en seguridad industrial

Hoy en día, con todo lo que implica la Industria 4.0, seguir ciertas normas internacionales se volvió algo necesario para cuidar los sistemas industriales, sobre todo porque ahora casi todo está conectado. Entre las más conocidas está la serie ISA/IEC 62443, que fue creada justamente para ayudar en temas de seguridad en sistemas automatizados. Lo interesante de estas guías es que no solo se usan para detectar posibles inconvenientes, sino que también explican qué aspectos técnicos se deben considerar en cada sección del sistema. Así, las empresas pueden aplicar medidas mejor organizadas y enfocadas en lo que realmente necesitan para mejorar su protección frente a los ataques. (ISA, s.f.).

3.2.4 Beneficios de tener una estrategia clara de ciberseguridad

Aplicar una buena estrategia de ciberseguridad en entornos industriales tiene muchas ventajas. Primero, porque ayuda a proteger los activos más importantes frente a cualquier ataque. Pero también es útil porque evita que los procesos se

frenen de golpe por algún problema digital, lo que permite seguir trabajando sin mayores interrupciones. Además, cuando una empresa cumple con ciertos estándares internacionales, no solo mejora internamente, sino que también transmite una imagen más seria y confiable hacia sus clientes o socios. En general, una estrategia bien pensada puede marcar la diferencia entre reaccionar a tiempo o sufrir daños graves. (Pedreira, Barros, & Pinto, 2021).

3.3 Riesgos Principales en la Industria 4.0

Con la aparición de la Industria 4.0 y el aumento de la interconexión entre dispositivos, los sistemas de control industrial (ICS) han evolucionado de ser fortalezas aisladas a transformarse en territorios vulnerables a potenciales ataques. Esta conectividad, que indudablemente incrementa la eficiencia y acelera los procesos, también revela más áreas vulnerables. Y es bien sabido que donde existe una grieta, alguien siempre entra. Las amenazas más comunes y problemáticas incluyen malware, phishing, ataques por fuerza bruta, inyecciones SQL y ataques de denegación de servicio. Todas comparten un mismo riesgo: afectan la estabilidad del entorno industrial. Sus consecuencias pueden ir desde dejar inactivos sistemas críticos, alterar datos importantes o incluso vulnerar la privacidad de la información. En la práctica, todo esto pone en peligro la continuidad de las operaciones.

3.3.1 Malware

El malware es uno de los riesgos más serios para los sistemas industriales, ya que puede infiltrarse en componentes clave, alterar su funcionamiento e incluso detener las operaciones. En los últimos años se ha demostrado que el uso de aprendizaje profundo (deep learning) resulta muy efectivo para detectar este tipo de amenazas en entornos industriales. Por ello, cada vez existe mayor interés en desarrollar herramientas que permitan identificarlas y contenerlas antes de que generen consecuencias más graves. (Kravchik, Biggio, & Shabtai, 2021).

3.3.2 Phishing

El phishing, que se basa en engañar a las personas a través de la manipulación psicológica, ha llegado a convertirse en una amenaza también para los entornos industriales. En estos escenarios, los atacantes se valen de la confianza de los trabajadores para abrirse paso hacia sistemas delicados. Investigaciones recientes advierten que, cuando este tipo de ataques se dirige a infraestructuras de control industrial, los daños pueden ser muy serios. Por eso, es fundamental invertir en la capacitación del personal y fortalecer una cultura de seguridad que permita detectar a tiempo estas trampas. (Al-Abassi, Karimipour, Dehghantanha, & Parizi, 2020).

3.3.3 Ataques de fuerza bruta

En los sistemas de control industrial (ICS), uno de los riesgos más comunes son los ataques de fuerza bruta. Este tipo de intentos se basa en probar contraseñas de manera repetida hasta dar con la correcta. Para hacerles frente no basta con tener claves complicadas; se requiere combinar métodos de autenticación sólidos con una supervisión constante del sistema. Solo así es posible detectar a tiempo cualquier intento de acceso indebido y proteger tanto la información como los recursos críticos. (Li, Ramanan, Gebraeel, & Paynabar, 2020).

3.3.4 Inyecciones SQL

Un ataque de inyección SQL se da cuando alguien se aprovecha de fallas en las aplicaciones que trabajan con bases de datos. Con esta técnica, los intrusos pueden colarse para ver información confidencial o, peor aún, alterarla. En un entorno industrial, las consecuencias pueden ser bastante serias: si se manipulan datos clave, las decisiones operativas pueden verse comprometidas y, con ello, el funcionamiento completo de los sistemas. (Kravchik, Biggio, & Shabtai, 2021).

3.3.5 Ataques DOS

El objetivo de los ataques de denegación de servicio es debilitar los recursos de un sistema o red, impidiendo que los clientes legítimos puedan hacer uso de los

servicios que usualmente emplean. En entornos industriales, pueden causar la interrupción de procesos automatizados, afectando la productividad y generando pérdidas económicas. La literatura especializada resalta la importancia de instaurar sistemas para identificar y minimizar las intrusiones para proteger las infraestructuras críticas contra este tipo de amenazas. (Al-Abassi, Karimipour, Dehghantanha, & Parizi, 2020).

3.4 Estrategias de defensa cibernética

Con todo lo que ha traído la Industria 4.0 los sistemas industriales ya no están aislados ahora todo está conectado y eso también hace que haya más formas de que los ataquen por eso se necesita sí o sí una defensa que aguante y que se adapte a lo que venga porque los riesgos no paran y no siempre son los mismos de antes.

De todo lo que se ha hablado en otros trabajos hay dos ideas que suenan fuerte una es la defensa en profundidad y la otra es usar marcos de ciberseguridad esas dos se usan bastante porque ayudan a tener control del sistema todo el tiempo a detectar cosas raras a responder sin esperar y a manejar los riesgos como se debe sobre todo en ambientes industriales donde cualquier fallo puede costar caro.

3.4.1 Defensa en profundidad y monitoreo continuo

La estrategia de defensa en profundidad establece múltiples capas de seguridad para proteger los sistemas industriales. Este enfoque garantiza que, en caso de que una capa de seguridad falle, aún existan otras barreras activas que sigan protegiendo el sistema. A esto se suma la importancia de mantener una vigilancia constante, que permita detectar comportamientos inusuales y reaccionar de inmediato ante cualquier amenaza. La combinación de ambos métodos es clave para asegurar la continuidad y confiabilidad de los procesos en entornos industriales (Alqudhaibi, Albarrak, Jagtap, Williams, & Salonitis, 2025).

3.4.2 Aplicación de marcos de ciberseguridad

Integrar marcos de ciberseguridad en los sistemas industriales permite organizar de manera clara y efectiva las acciones necesarias para gestionar riesgos y aplicar medidas de protección. Estas guías no solo reúnen buenas prácticas, sino que también están pensadas para responder a las particularidades de la Industria 4.0, ayudando a que las organizaciones estén mejor preparadas frente a posibles ciberataques (Ning & Jiang, 2022).

3.5 La función de la inteligencia artificial en la protección cibernética

En el entorno de la Industria 4.0, las amenazas digitales no solo han aumentado en cantidad, sino también en complejidad. Esta evolución ha dejado atrás la eficacia de muchos sistemas tradicionales de defensa, como aquellos que dependen exclusivamente de firmas o reglas predefinidas. En este nuevo escenario, la inteligencia artificial es un aliado significativo porque permite el desarrollo de sistemas de protección más versátiles y dinámicos que pueden adaptarse a riesgos que cambian continuamente. (Jada & Mayayise, 2024).

La inteligencia artificial tiene la capacidad de analizar grandes volúmenes de información en tiempo real, detectar patrones complejos y comportamientos que podrían estar relacionados con acciones dañinas debido a su capacidad de procesamiento. Los sistemas de inteligencia artificial tienen la ventaja de aprender y adaptarse continuamente, lo que les permite reaccionar mejor ante peligros emergentes o desconocidos. Esto es diferente a los métodos tradicionales, que dependen de normas establecidas y requieren actualizaciones manuales. (Salem, Azzam, Emam, & Abohany, 2024).

En estos últimos años, la inteligencia artificial ha dejado de ser una idea lejana para convertirse en algo muy concreto dentro del mundo de la ciberseguridad. Lo

interesante no es solo lo que puede hacer, sino cómo está cambiando la manera en que pensamos los riesgos digitales.

Una de sus contribuciones más significativas es la habilidad para identificar amenazas antes de que puedan provocar perjuicio. La Inteligencia Artificial no requiere una lista de lo que es correcto y lo que es incorrecto, solo necesita observar patrones y identificar lo que va más allá de lo convencional. Y cuando algo no encaja,

lanza una señal. No es infalible, claro, pero sí es sorprendentemente eficaz a la hora de adelantarse a muchos ataques que, de otro modo, pasarían desapercibidos según lo señalan estos autores en sus investigaciones recientes. (Salem, Azzam, Emam, & Abohany, 2024).

También ha demostrado ser útil para reducir esas molestas falsas alarmas que saturan a los equipos de seguridad. A fuerza de entrenamiento, los sistemas basados en IA aprenden a diferenciar el ruido del peligro real. Esto permite que los especialistas se concentren en lo urgente, sin desgastarse con advertencias que no llevan a ningún lado, como bien explican Jada y Mayayise. (Jada & Mayayise, 2024).

Hay otra ventaja que se está volviendo clave en contextos donde el tiempo es oro: la posibilidad de reaccionar de forma automática. La IA puede actuar sin esperar órdenes humanas, lo que significa que puede aislar una red comprometida o frenar una amenaza justo en el momento en que aparece. Esa capacidad de respuesta inmediata, casi instintiva, acorta los tiempos críticos y minimiza el daño. Salem y su equipo han documentado este tipo de intervenciones con resultados contundentes.

Y encima de eso, estos sistemas no se limitan a lo que ya conocen. La inteligencia artificial se va volviendo más resistente con el tiempo: aprende de cada ataque, incluso de aquellos que no tienen nada que ver con los previos. Precisamente esa capacidad de adaptación es lo que la convierte en un recurso tan valioso en un mundo donde las amenazas se modifican con una frecuencia casi igual a la del clima.

Por lo tanto, no es inesperado que su uso esté creciendo, sobre todo en situaciones industriales. La inteligencia artificial proporciona una defensa ininterrumpida en ese lugar, donde la interrupción de una operación puede tener consecuencias millonarias y que, a diferencia de los humanos, no se distrae ni un momento.

3.6 Métodos básicos de Inteligencia Artificial en la ciberseguridad industrial

En el campo de la ciberseguridad industrial, la inteligencia artificial se basa en varios métodos que permiten manejar elevados volúmenes de información, detectar patrones complejos y reaccionar rápidamente a riesgos potenciales. La automatización del aprendizaje, las redes neuronales artificiales y el aprendizaje profundo son algunas de las herramientas más empleadas en la actualidad para mejorar la protección de sistemas esenciales. (Halbouni, y otros, 2022).

Aprendizaje automático: Este procedimiento permite que los sistemas comprendan directamente los datos, sin la necesidad de instrucciones particulares para cada circunstancia. El aprendizaje automático se emplea en el campo de la ciberseguridad para detectar comportamientos inusuales, reconocer patrones que indiquen un ataque y anticipar posibles amenazas. Se ha convertido en una herramienta esencial para proteger las infraestructuras industriales debido a su habilidad para ajustarse ante estrategias recientes empleadas por los asaltantes. (Halbouni, y otros, 2022).

Aprendizaje profundo: El aprendizaje profundo, que es parte del aprendizaje autónomo, emplea redes neuronales de varias capas para examinar datos con diferentes niveles de complejidad. Esta técnica resulta especialmente útil para detectar amenazas complejas, ya que tiene la capacidad de reconocer patrones muy sutiles en grandes volúmenes de información, los cuales pueden no ser detectados por métodos más tradicionales. (Halbouni, y otros, 2022).

Redes neuronales: Basándose en el desempeño cerebral humano, las redes neuronales se conforman de nodos interconectados que gestionan los datos de manera similar a la que hacen las neuronas biológicas. En el ámbito de la ciberseguridad industrial, este tipo de redes ha demostrado ser clave para construir sistemas capaces de identificar y reaccionar ante amenazas en tiempo real, lo que refuerza significativamente la capacidad de resolución frente a ataques. (Halbouni, y otros, 2022).

Estas tecnologías han conseguido potenciar significativamente los sistemas de detección de intrusiones, además de salvaguardar la información contra conductas malintencionadas. Su integración en contextos industriales ha proporcionado un apoyo firme a las tácticas de defensa, mejorando su efectividad y confiabilidad.

3.7 Identificación de anomalías, análisis de comportamiento y respuesta automática

En los entornos industriales, proteger los sistemas críticos va más allá de aplicar medidas tradicionales. Actualmente, técnicas como la detección de anomalías, el análisis del comportamiento de la operación y las respuestas automatizadas se han convertido en elementos clave. Gracias a estas herramientas es posible reconocer amenazas en tiempo real y reaccionar de inmediato, lo que refuerza la resiliencia de la infraestructura y ayuda a mantener la continuidad de las operaciones frente a cualquier incidente.

3.7.1 Identificación de irregularidades

Por decirlo simple, este método trata de conocer cómo se comporta normalmente un sistema para después detectar cualquier cosa rara que no encaje. Lo que se busca es reconocer esas señales que podrían indicar algo sospechoso. Hoy en día, la inteligencia artificial permite revisar enormes cantidades de datos casi al momento y sacar a la luz patrones extraños que antes pasarían desapercibidos. En un entorno industrial, esto ayuda a reaccionar rápido y a contener el problema

antes de que se convierta en un dolor de cabeza mayor, evitando pausas en la operación y manteniendo los procesos lo más estables posible. (Muheidat, Mallouh, Al-Saleh, Al-Khasawneh, & Tawalbeh, 2024).

3.7.2 Análisis de comportamiento

El análisis de comportamiento, en pocas palabras, trata de observar cómo se usan los sistemas en el día a día y cómo responden ante distintas acciones de los usuarios. La idea es poner atención a lo que se sale de lo común, aunque parezca un detalle mínimo. En este sentido, la inteligencia artificial es una aliada importante: no solo procesa información, también aprende de cada nuevo caso que se le presenta. Esto permite identificar comportamientos extraños que pueden significar intentos de entrar sin permiso o un manejo inadecuado de los recursos. Además, este enfoque resulta especialmente útil para descubrir amenazas internas o ataques más complejos que, de otra forma, suelen pasar desapercibidos con las medidas de seguridad tradicionales. (Salem, Azzam, Emam, & Abohany, Advancing cybersecurity: A comprehensive review of AI-driven detection techniques, 2024).

3.7.3 Respuesta automática

Una capacidad esencial en los sistemas de ciberseguridad industrial es la respuesta automática, porque permite que se tomen medidas inmediatamente después de identificar una amenaza, sin requerir intervención individual humana. Entre las medidas que se pueden llevar a cabo de forma autónoma están el aislamiento de equipos sensibles, la eliminación de accesos potencialmente peligrosos y activación de protocolos predefinidos para emergencias. Esta capacidad de respuesta en tiempo real permite reducir significativamente el impacto de los incidentes en las operaciones, lo que mejora los períodos de mitigación y contención frente a posibles ataques. (Ali, Shah, & ElAffendi, 2025).

3.8 Investigaciones previas y modelos desarrollados

Uno de los estudios más destacados en este ámbito es la Detección de Anomaly Basada en Inteligencia Artificial para la Ciberseguridad en tiempo real. Lo fascinante de este estudio es que propone una opción distinta a los procedimientos tradicionales de ciberseguridad, que generalmente se apoyan en bases de datos con firmas previamente reconocidas. En lugar de eso, el sistema propuesto utiliza

inteligencia artificial —concretamente, algoritmos como redes neuronales y máquinas de vectores de soporte— para detectar comportamientos inusuales dentro de una red, incluso si no coinciden con amenazas registradas anteriormente. Esta capacidad le permite identificar ataques del tipo “día cero”, es decir, aquellos que explotan vulnerabilidades desconocidas, así como posibles amenazas internas que los sistemas tradicionales suelen pasar por alto (Goswami, 2024).

Un aporte significativo lo constituye el trabajo titulado Multi-step Attack Detection in Industrial Networks Using a Hybrid Deep Learning Architecture, en el cual se propone una arquitectura híbrida basada en la combinación de redes neuronales convolucionales (CNN) y redes de creencia profunda (DBN). Este enfoque fue desarrollado con el objetivo de identificar ataques complejos que se producen en varias etapas dentro de redes industriales. Al ser evaluado utilizando el conjunto de datos MSCAD, el modelo obtuvo resultados destacados en términos de precisión para la detección de intrusiones avanzadas, superando el rendimiento de múltiples soluciones previamente documentadas en estudios similares. (Jamal, y otros, 2023).

En el estudio titulado An Ensemble Deep Learning-based Cyber-Attack Detection in Industrial Control System, se desarrolló un modelo orientado a la detección de ciberataques en sistemas de control industrial (ICS), a partir del uso de técnicas de aprendizaje profundo. La propuesta integró redes neuronales profundas con algoritmos basados en árboles de decisión, con el propósito de mejorar la precisión en la identificación de amenazas. Los resultados obtenidos durante su evaluación

evidenciaron un rendimiento superior en comparación con clasificadores tradicionales como Random Forest y AdaBoost, tanto en términos de exactitud como en la disminución de falsos positivos. (Al-Abassi, Karimipour, Dehghantanha, & Parizi, 2020).

A pesar de los avances presentados por diversas investigaciones, aún queda un largo camino por recorrer. Si bien es cierto que la inteligencia artificial ha demostrado un potencial enorme para reforzar la seguridad en entornos industriales, la mayoría de los trabajos revisados se enfocan en escenarios demasiado acotados o controlados, muchas veces alejados de la complejidad y variabilidad del entorno real. Este enfoque limitado deja en evidencia la necesidad urgente de pensar soluciones más amplias, que combinen distintos métodos de detección y respuesta. Solo con enfoques más integrales será posible enfrentar el amplio espectro de amenazas que hoy afectan a sistemas industriales interconectados, dinámicos y en constante evolución.

4. Materiales y metodología

Este capítulo detalla qué recursos se emplearon y la manera en que se trabajó para alcanzar las metas de la investigación. Se creó una red virtual controlada que simula una infraestructura de Industria 4.0 donde se recolectaron datos del tráfico usando Zeek para capturar y analizar la información y con eso se desarrolló un prototipo de inteligencia artificial que detecta y responde de forma automática a ataques cibernéticos.

La metodología se basó en crear un entorno de prueba que se pueda repetir simular ataques para obtener datos reales procesar y limpiar esa información entrenar una red neuronal para detectar amenazas y luego integrar un sistema que bloquee las fuentes maliciosas sin intervención humana.

El trabajo se organizó en seis partes que son describir el entorno configurar el sistema para capturar el tráfico elegir los ataques que se van a probar, recolectar y procesar los datos, entrenar el modelo de IA y por último hacer pruebas para ver si todo funciona bien.

4.1 Revisión de amenazas y ataques cibernéticos en la Industria

Para poder simular ataques realistas en el entorno virtual de pruebas, primero fue necesario entender cuáles son las amenazas más habituales que enfrentan los sistemas industriales bajo el enfoque de la Industria 4.0. Esta sección se enfoca en identificar, analizar y clasificar esas amenazas usando fuentes académicas e institucionales confiables, con el fin de establecer una base firme para el diseño de las pruebas y el desarrollo del modelo de detección.

4.1.1 Estrategia de búsqueda y selección de fuentes

Para identificar las amenazas cibernéticas más relevantes que afectan a los sistemas industriales dentro del contexto de la Industria 4.0, Se llevó a cabo un repaso bibliográfico en bases de datos académicas prestigiosas por su calidad y enfoque en temas de ciberseguridad. Las principales plataformas utilizadas fueron, IEEE Xplore, Scopus, Web of Science y Google Académico.

Se definieron ciertos criterios para seleccionar los documentos más útiles. Entre esos era que los artículos estuvieran publicados entre 2020 y 2025, disponibles en inglés o español y con acceso completo. También debían enfocarse claramente en amenazas relacionadas con sistemas de control industrial (ICS), entornos OT o tecnologías propias de la Industria 4.0.

Por otro lado, se excluyeron documentos sin revisión por pares, trabajos no indexados como tesis de grado y artículos duplicados o que trataran sobre sectores fuera del ámbito industrial.

Las búsquedas se hicieron utilizando combinaciones de palabras clave como “Industrial Cybersecurity”, “Cyber threats in Industry 4.0”, “ICS malware”, “Phishing OT”, “Denial of Service ICS” y “SQL injection industrial systems”. También se aplicaron filtros por fecha y tipo de publicación para garantizar la actualidad y importancia de los resultados.

4.1.2 Análisis de fuentes institucionales especializadas

Para complementar lo encontrado en la revisión académica también se analizaron informes técnicos y reportes publicados por entidades reconocidas en el área de la ciberseguridad. Estas fuentes vienen tanto del sector privado como de organizaciones internacionales y ayudaron a validar patrones de ataque detectar tendencias nuevas y fortalecer la clasificación de amenazas ya establecida.

Entre las fuentes más importantes analizadas está IBM Security con su informe X Force Threat Intelligence Index 2025 que presenta un análisis detallado de las

amenazas más comunes en entornos industriales con estadísticas sobre técnicas usadas por atacantes y los sectores más afectados. Este documento se consultó directamente en el sitio oficial de IBM.

Cisco Systems aportó dos informes relevantes que son el Cisco Cybersecurity Readiness Index 2024 y el State of Industrial Networking Report 2024 donde se describen las vulnerabilidades más frecuentes en redes industriales el nivel de preparación de las empresas ante ataques y cómo la unión entre sistemas IT y OT afecta la seguridad.

Cybersecurity Insiders junto con Gurucul publicaron el 2024 Insider Threat Report que se enfoca en amenazas internas dentro de organizaciones complejas incluyendo entornos industriales donde destacan prácticas como el uso indebido de accesos y la suplantación de identidad.

El Foro Económico Mundial también fue considerado a través del Global Risks Report 2024 disponible en su sitio web donde se reconoce la ciberseguridad industrial como una de las amenazas más serias a nivel global por el aumento en la dependencia de sistemas conectados.

Otro informe clave fue el Cost of a Data Breach Report 2024 hecho por IBM y Ponemon Institute que muestra datos económicos actualizados sobre el impacto financiero de los ciberataques especialmente en sectores con infraestructura crítica como la manufactura.

Además, se incluyeron estudios científicos encontrados en bases como IEEE Xplore Scopus y Google Scholar con autores como Al-Abassi Jamal y Kravchik que ofrecen datos prácticos sobre cómo detectar y responder a ataques en sistemas de control industrial.

Gracias a todas estas fuentes se lograron identificar patrones en común entre lo que dice la investigación académica y lo que muestran los reportes técnicos lo que dio más solidez al análisis general. Toda la información fue ordenada en una clasificación de amenazas basada en el tipo de ataque el impacto que puede causar

y qué tan difíciles son de detectar lo cual ayudó a decidir qué riesgos eran más importantes.

4.1.3 Clasificación y priorización de amenazas

Con los datos recogidos en la revisión bibliográfica académica y en los informes técnicos se hizo una clasificación de las amenazas cibernéticas que afectan a entornos de Industria 4.0. Esta clasificación se armó con base en cinco criterios clave que ayudan a entender mejor el tipo de riesgo y su importancia. Estos criterios son:

- Tipo de amenaza que señala la categoría técnica o el modelo de ataque.
- Vector de ataque que muestra por dónde entra el atacante al sistema.
- Impacto potencial que señala las consecuencias que puede causar en las operaciones o la seguridad.
- Frecuencia observada que se refiere a qué tan común es según estudios y reportes.
- Dificultad de detección que muestra qué tan fácil o difícil es identificar el ataque con los mecanismos de defensa actuales.

La siguiente tabla resume los resultados obtenidos de la investigación de todas las fuentes consultadas:

Tabla 1

Clasificación de amenazas cibernéticas relevantes en entornos de Industria 4.0

Nota. Elaboración propia

Tipo de amenaza	**Vector de ataque**	**Impacto potencial**	**Frecuencia observada**	**Dificultad de detección**
Malware	Archivos adjuntos maliciosos, descargas web, USB infectados	Control del sistema, sabotaje, robo de información	Alta	Media
Phishing	Correos electrónicos fraudulentos,	Robo de credenciales,	Alta	Alta

	formularios falsos	accesos no autorizados		
Fuerza bruta	Interfaces de autenticación expuestas (SSH, FTP, login web)	Acceso forzado a sistemas críticos	Media	Baja
Inyección SQL	Formularios web vulnerables, consultas mal sanitizadas	Manipulación o robo de datos, alteración de procesos	Media	
DoS	Peticiones HTTP/UDP masivas, tráfico malicioso	Interrupción de servicios, caída de sistemas	Alta	Alta

4.2 Evaluación y elección de métodos de inteligencia artificial para defensa cibernética

Para identificar qué métodos de inteligencia artificial son más apropiadas en temas de ciberseguridad dentro de entornos de Industria 4.0 se realizó una investigación dividida en cuatro etapas que se complementan entre sí. Estas fueron el levantamiento de información a través de encuestas a expertos la revisión de literatura técnica el análisis comparativo de herramientas y un análisis multicriterio. Este enfoque permitió evaluar si las técnicas son viables desde lo técnico y si realmente se pueden aplicar en contextos industriales.

4.2.1 Recolección de perspectivas mediante encuestas a expertos

Como primer paso en la evaluación se aplicó una encuesta a 12 expertos en temas clave como inteligencia artificial, ciberseguridad industrial, redes, programación y automatización. El objetivo fue identificar cuáles son los criterios técnicos más importantes al momento de elegir técnicas de IA en entornos OT.

Los encuestados calificaron la importancia de 12 criterios técnicos en una medida Likert del 1 al 5. Los hallazgos revelaron que las técnicas más valoradas son las que ofrecen buena precisión, alta velocidad de detección, capacidad de respuesta automática y que además funcionen bien en ambientes industriales. También se tuvo en cuenta la escalabilidad y la reducción de falsos positivos.

Tabla 2

Promedio de importancia de criterios técnicos según encuestas a expertos

Nota. Elaboración propia

Criterio evaluado	**Promedio (1-5)**
Precisión del modelo	5.0
Velocidad de detección	5.0
Facilidad de implementación	4.0
Escalabilidad	4.0
Compatibilidad con entornos industriales	5.0
Automatización de respuesta	5.0
Adaptabilidad a nuevas amenazas	5.0
Facilidad de mantenimiento	3.0
Nivel de interpretabilidad	4.0
Requerimientos computacionales	3.0
Tasa de falsos positivos	5.0
Incorporación con sistemas ya establecidos	4.0

4.2.2 Revisión de literatura y benchmarking de técnicas de IA aplicadas a ciberseguridad

Mientras se realizaban las encuestas también se hizo una verificacion sistemática de literatura usando bases académicas como IEEE Xplore Scopus y Web of Science junto con algunos artículos técnicos de Google Scholar. Esta revisión sirvió para identificar qué técnicas de inteligencia artificial se están usando actualmente para detectar y responder a ciberataques en entornos industriales.

Las técnicas más mencionadas fueron:

- Redes neuronales sintéticas
- Máquinas de vectores de soporte
- Árboles de decisiones
- Aprendizaje profundo
- Redes convolucionales
- Modelos de clustering no supervisado
- Algoritmos evolutivos y métodos de optimización bioinspirada

Estas técnicas fueron comparadas a partir de estudios relevantes como los de Al-Abassi Kravchik y Halbouni donde se analizó su rendimiento en tareas como detección de intrusiones clasificación de tráfico y predicción de comportamientos maliciosos.

4.2.3 Evaluación por análisis multicriterio

Las técnicas de inteligencia artificial que se identificaron fueron comparadas usando un análisis multicriterio basado en los criterios definidos en las encuestas a expertos. La siguiente tabla muestra un resumen con los resultados de esa evaluación cualitativa:

Tabla 3

Comparación de técnicas de IA mediante análisis multicriterio

Nota. Elaboración propia

Técnica evaluada	Precisión	Velocidad	Escalabilidad	Interpretabilidad	Automatización	Puntaje total
Árboles de decisión	Media	Alta	Alta	Alta	Media	3.5
SVM	Alta	Media	Media	Media	Media	3.7
Redes neuronales artificiales (RNA)	Alta	Alta	Alta	Media	Alta	4.5
Aprendizaje profundo (DL)	Muy alta	Alta	Alta	Baja	Alta	4.4
Bosques aleatorios	Alta	Alta	Media	Media	Media	4.0

Según los resultados del análisis se concluyó que las redes neuronales artificiales y el aprendizaje profundo son las técnicas con mayor potencial para entornos industriales por su capacidad de adaptarse a nuevos patrones de ataque, trabajar en tiempo real y automatizar respuestas.

4.2.4 Justificación de la técnica seleccionada

La elección de las redes neuronales artificiales como base del sistema desarrollado en esta investigación se justificó por su buen rendimiento en estudios anteriores su facilidad para integrarse con sistemas que operan en tiempo real y su capacidad de mejorar con nuevos datos a través de entrenamiento incremental. Estas cualidades son especialmente útiles en entornos de Industria 4.0 donde los procesos cambian con frecuencia y los ataques son cada vez más complejos por lo que se necesitan modelos que sean adaptables precisos y autónomos.

4.3 Desarrollo práctico del marco de defensa cibernética

Luego de identificar las amenazas más comunes en entornos de Industria 4.0 y de seleccionar los métodos de inteligencia artificial más adecuadas en especial las redes neuronales se pasaron a la parte práctica del sistema de defensa cibernética propuesto.

Para organizar mejor el trabajo se usó una metodología basada en el marco Scrum. Al mismo tiempo se creó un entorno experimental virtual que simula una infraestructura industrial vulnerable donde se generó tráfico de red con actividades normales y también con ataques cibernéticos controlados.

Este entorno sirvió para capturar datos usando la herramienta Zeek, procesarlos con scripts hechos en Python y entrenar un modelo de red neuronal que es capaz de organizar el tráfico de forma automática.

En las proximas secciones se explica cada parte del sistema desarrollado incluyendo cómo se gestionó con Scrum, cómo se armó el entorno virtual, cómo se simularon las amenazas, cómo se entrenó el modelo de IA, cómo se implementó la respuesta automática y cómo se validó todo frente a los objetivos definidos.

4.3.1 Metodología ágil aplicada: Scrum

Para organizar bien las etapas del desarrollo del sistema de defensa cibernética basado en inteligencia artificial se usó una metodología ágil con base en el marco Scrum. Este enfoque ayudó a dividir el trabajo en sprints iterativos para seguir el avance y entregar componentes funcionales de forma progresiva según los objetivos definidos.

Scrum se adaptó a las necesidades del proyecto tanto en lo académico como en lo técnico lo que permitió una gestión flexible enfocada en resultados. Las tareas se organizaron en ciclos cortos con metas claras lo que facilitó integrar cada parte del sistema poco a poco y detectar errores o desviaciones a tiempo.

El control de actividades se hizo con una herramienta de gestión llamada Jira usando tableros tipo Kanban con las columnas “Por hacer”, “En proceso” y “Completado”. Esto permitió asignar tareas, ver el progreso y llevar un registro claro de cada entrega dentro de los sprints, lo que mejoró la coordinación y el control del proyecto.

Los sprints definidos fueron los siguientes:

- **Sprint 1 – Diseño y configuración del entorno experimental**
    - Instalación de máquinas virtuales
    - Integración de herramientas como Zeek DVWA y scripts para simular ataques
- **Sprint 2 – Captura y procesamiento de datos de red**
    - Recolección de tráfico con Zeek
    - Limpieza transformación y organización de los datos en formato Excel
- **Sprint 3 – Desarrollo del modelo de inteligencia artificial**
    - Codificación y entrenamiento de la red neuronal
    - Validación del modelo con métricas de rendimiento
- **Sprint 4 – Implementación de predicción y respuesta automatizada**
    - Integración del sistema de detección en tiempo real
    - Automatización del bloqueo de IPs maliciosas
    - Validación del sistema completo en funcionamiento

El uso de este enfoque ágil junto con herramientas modernas ayudó a organizar el trabajo técnico integrar las funciones más importantes de forma ordenada y cumplir con los objetivos del proyecto de manera clara y controlada.

4.3.2 Entorno experimental virtualizado

Como parte clave del desarrollo del sistema de defensa cibernética basado en inteligencia artificial se construyó un entorno de pruebas controlado usando virtualización. Este entorno permitió simular una red industrial con varios dispositivos y vulnerabilidades lo que ayudó a generar tanto datos normales como maliciosos que sirvieron para entrenar y validar el modelo de inteligencia artificial.

Para esto se usó la herramienta Vagrant junto con VirtualBox como hipervisor lo que permitió crear y gestionar de forma automática cinco máquinas virtuales conectadas entre sí por una red privada. Esta configuración representa una red típica de Industria 4.0 con servidores web y de base de datos estaciones de trabajo nodos de monitoreo y actores maliciosos.

Cada máquina virtual fue configurada según su función específica y todas se conectaron a la subred privada 192.168.56.0/24 lo que garantizó un entorno seguro y aislado sin conexión con redes externas. Las configuraciones incluyeron memoria RAM cantidad de CPU resolución de pantalla y tipo de teclado ajustado a las necesidades de cada equipo.

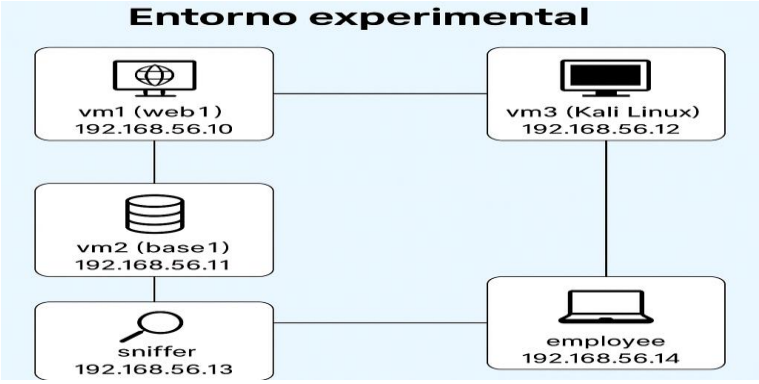


Figura 1

Diagrama del entorno experimental utilizado para la captura de tráfico de red

Nota. Elaboración propia

A continuación, se describe la función y configuración de cada máquina virtual:

- **web1 (192.168.56.10):** Servidor web con AlmaLinux 8 que utiliza Apache y PHP. En este equipo se instaló la aplicación Damn Vulnerable Web Application (DVWA) que fue configurada para simular vulnerabilidades reales de seguridad web. Se modificaron parámetros como allow_url_include safe_mode y magic_quotes_gpc y se activaron los permisos de escritura necesarios para que la aplicación funcione correctamente.

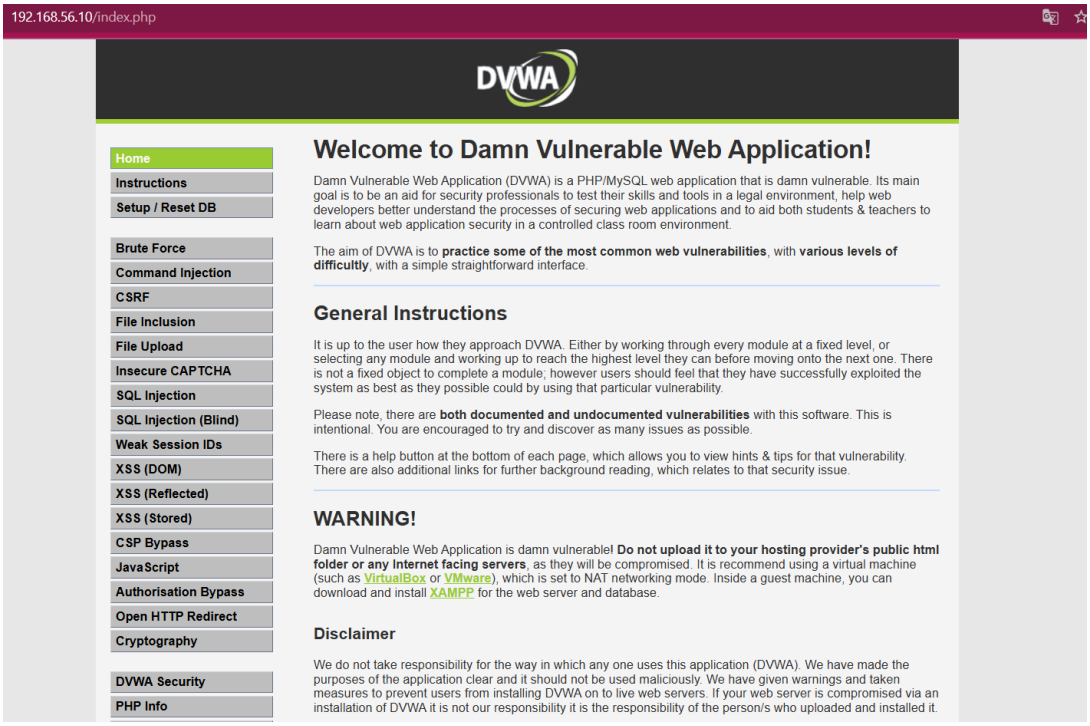


Figura 2

Aplicación DVWA desplegada en el servidor web.

Nota. Elaboración propia

- **base1 (192.168.56.11):** Servidor de base de datos MySQL con AlmaLinux 8 configurado para funcionar junto con la aplicación DVWA. Se creó una base de datos específica con un usuario y permisos adecuados y se habilitó la conexión entre la base de datos y el servidor web usando la red privada

Figura 3

Conexión de DVWA a la base de datos MySQL.

Nota. Elaboración propia

- **Kali Linux (192.168.56.12):** Máquina destinada a simular ataques, equipada con herramientas como Hydra, sqlmap, hping3, curl y dig. Se utiliza para ejecutar pruebas controladas de ciberataques, incluyendo fuerza bruta, inyecciones SQL, ataques de denegación de servicio (DoS), distribución de malware y campañas de phishing.

Figura 4

Simulación de ataques desde la máquina atacante

Nota. Elaboración propia

- **sniffer (192.168.56.13):** Nodo de monitoreo con AlmaLinux 8 donde se instaló Zeek como herramienta de análisis de tráfico. Se configuró con scripts personalizados para registrar eventos como ataques Dos, inyecciónes SQL, malware o fuerza bruta en un archivo de log estructurado llamado custom_conn_log.log.

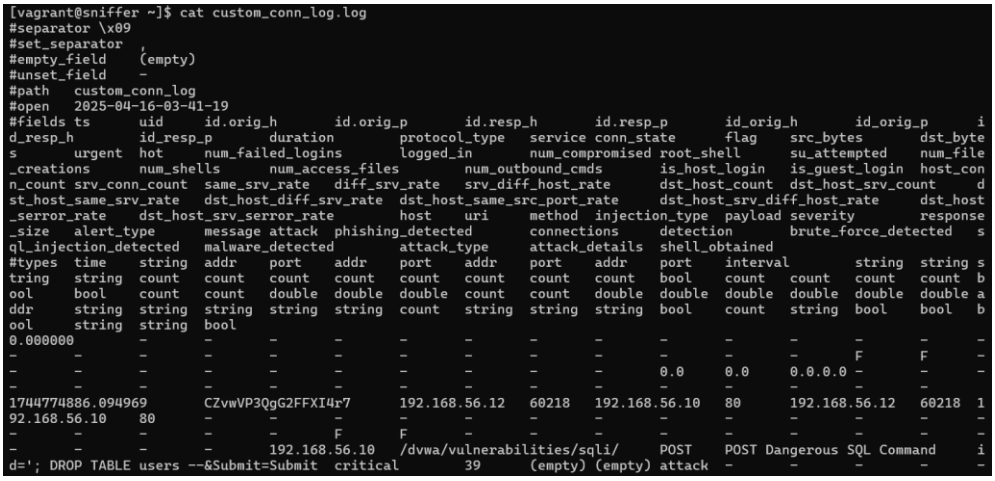


Figura 5

Análisis de tráfico malicioso mediante Zeek.

Nota. Elaboración propia

Además, esta máquina contó con Python y las bibliotecas necesarias, como pandas, numpy y sklearn, junto con un script de predicción llamado predecir_custom_log.py. Dicho programa se encarga de analizar los registros generados por el sistema y, mediante una red neuronal desarrollada con Keras, aplica bloqueos automáticos a direcciones IP sospechosas utilizando iptables.

- **employee (192.168.56.14):** Estación de trabajo simulada basada en Ubuntu Bionic con entorno gráfico configurada para realizar acciones normales como navegar en DVWA, para consultar datos y generar tráfico legítimo. Su función fue ayudar a balancear el dataset y representar el comportamiento típico de usuarios en un entorno industrial.

Figura 6

Generación de tráfico normal desde la máquina employee

Nota. Elaboración propia

Este entorno experimental permitió no solo simular ataques en condiciones controladas sino también asegurar que los datos capturados para el modelo de inteligencia artificial representaran escenarios reales con tráfico normal y malicioso. La separación de funciones la configuración segura de la red y el uso de herramientas automatizadas cumplieron con las prácticas de sandboxing y validación en entornos de prueba seguros tal como se planteó en el objetivo del sistema.

4.3.3 Simulación de ataques y generación de datos

Para entrenar y validar el sistema de inteligencia artificial se necesitaba un conjunto de datos que incluyera tanto tráfico normal como patrones maliciosos que representaran amenazas reales en entornos industriales. Por eso se diseñaron y se ejecutaron ataques cibernéticos controlados desde la máquina atacante Kali Linux hacia otros equipos del entorno usando herramientas y técnicas comunes entre los atacantes reales.

Los ataques se eligieron en función de su relevancia dentro de la clasificación de amenazas en sistemas de control industrial y por su capacidad de generar

comportamientos anómalos que un modelo de inteligencia artificial pueda detectar. Estos ataques quedaron registrados como eventos en los logs de Zeek con distintos vectores y niveles de impacto lo que permitió trabajar con datos útiles para detección automatizada y análisis de comportamiento.

A continuación, se describen los ataques simulados:

- **Ataques de denegación de servicio:** Se creó un script llamado dos.sh que combinaba tráfico TCP, tipo SYN flood, tráfico UDP y múltiples solicitudes HTTP usando hping3 y curl. El objetivo fue saturar el servidor web simulando un intento de denegación de servicio.

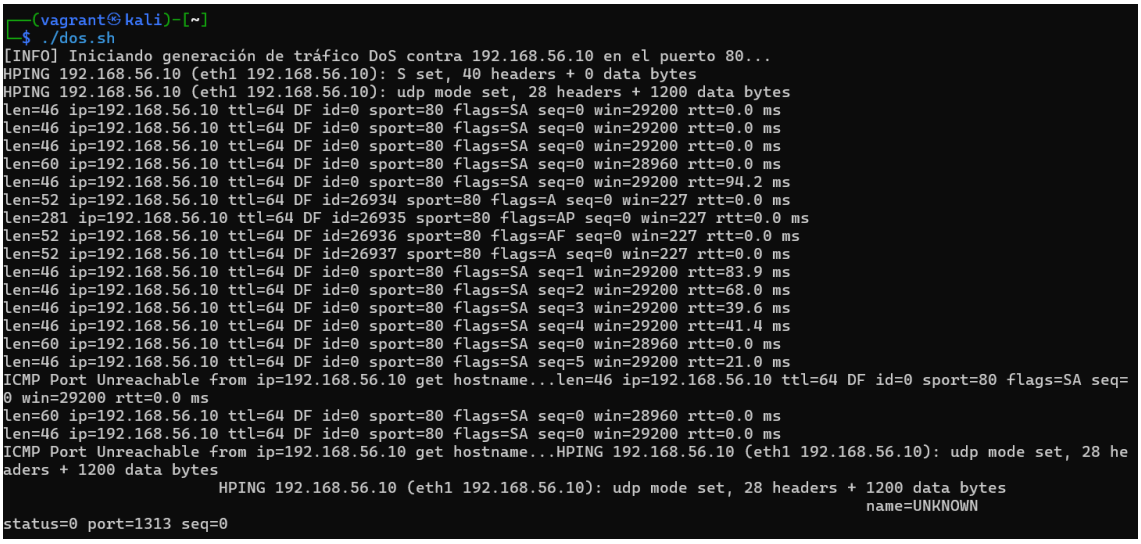


Figura 7

Ejecución del ataque de denegación de servicio

Nota. Elaboración propia

- **Ataques de fuerza bruta:** Se desarrolló un script llamado fuerza_bruta.sh que realizaba intentos fallidos y exitosos de inicio de sesión en los servicios SSH y FTP del servidor. Para esto se usaron varias combinaciones de credenciales incorrectas con el fin de simular un ataque clásico de fuerza bruta.

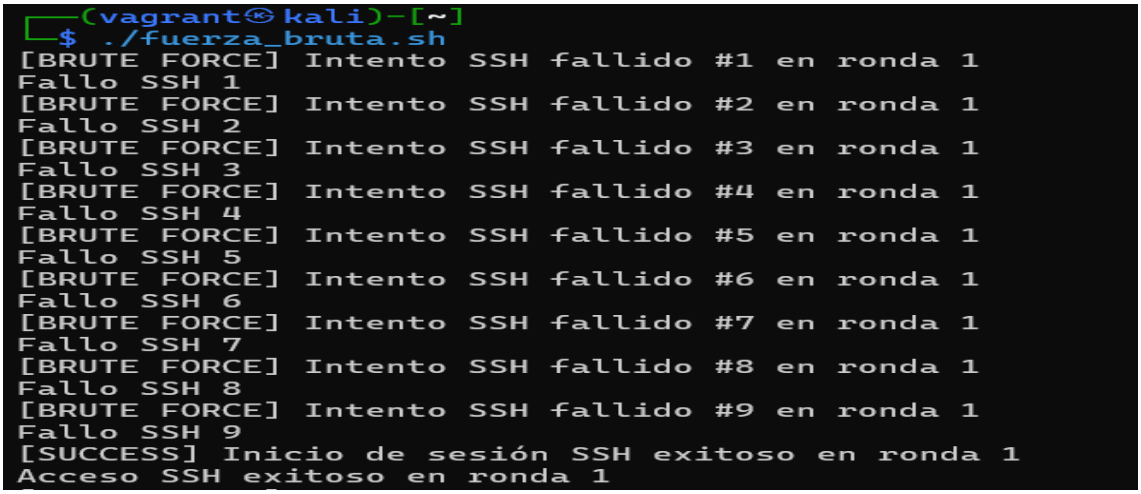


Figura 8

Ejecución del ataque de fuerza bruta

Nota. Elaboración propia

- **Ataques de inyección SQL:** Se usó un script llamado sql.sh para enviar varios payloads de inyección SQL a formularios vulnerables dentro de la aplicación DVWA. Se aplicaron técnicas como UNION SELECT OR 1=1 SLEEP y DROP TABLE para probar si el sistema tenía la habilidad de identificar este tipo de ataques.

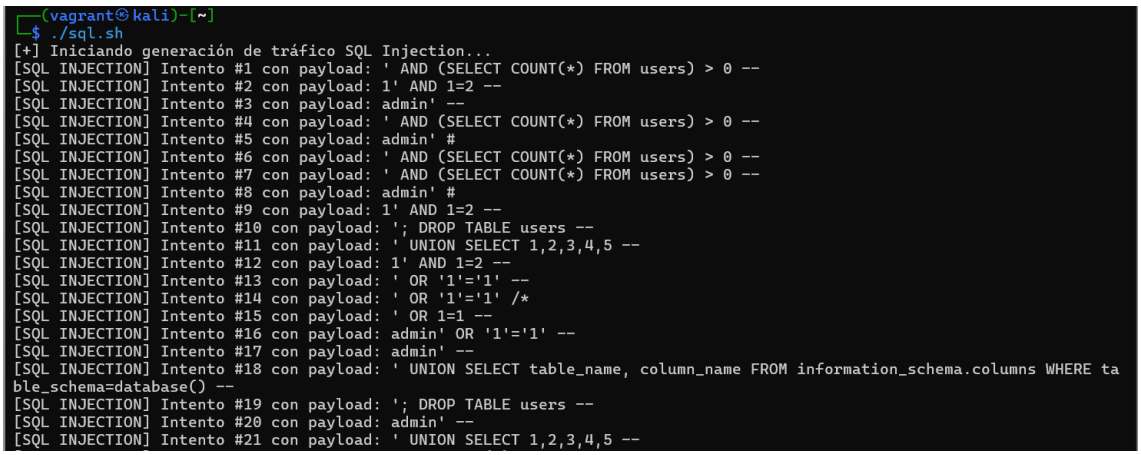


Figura 9

Ejecución del ataque de inyección SQL

Nota. Elaboración propia

- **Ataques de malware y exfiltración de datos:** Con el script malware.sh se generó tráfico que imitaba descargas de archivos maliciosos consultas DNS

sospechosas y exfiltración de datos sensibles usando cargas HTTP o tráfico masivo hacia destinos no autorizados. Para simular este comportamiento se usaron herramientas como curl y dig.

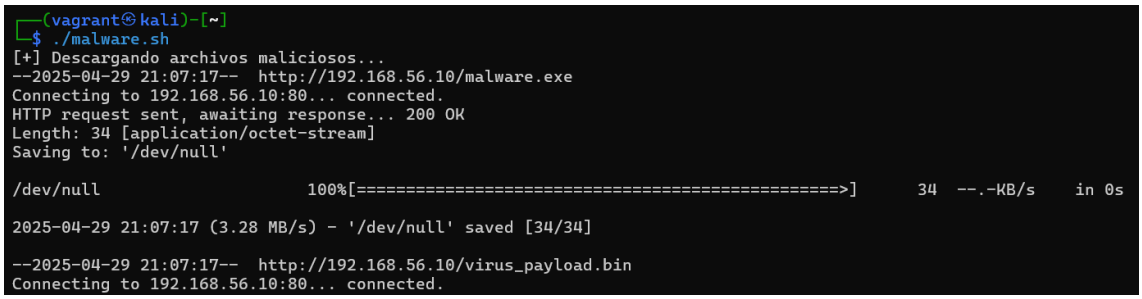


Figura 10

Simulación de descargas de malware y exfiltración

Nota. Elaboración propia

- **Ataques de phishing y suplantación de identidad:** Con el script phishing.sh se simuló tráfico típico de phishing incluyendo el envío de formularios maliciosos, redirecciones sospechosas, headers modificados, descargas de archivos falsos, envíos de correos por SMTP y consultas DNS a dominios maliciosos.

Figura 11

Simulación de ataques de phishing

Nota. Elaboración propia

Cada uno de estos ataques se ejecutó en intervalos controlados y bajo condiciones que se podían repetir dentro del entorno virtual. Esta forma de trabajo permitió capturar registros detallados del comportamiento de la red ante eventos maliciosos que luego se usaron para entrenar el modelo de inteligencia artificial.

Gracias a esta simulación se generaron datos reales que muestran anomalías en el comportamiento de red lo que es clave para que el modelo aprenda a diferenciar entre tráfico normal y ataques. Además, se logró una distribución equilibrada entre clases que incluyó varios tipos de amenazas lo que mejora la capacidad del sistema para analizar el comportamiento y detectar patrones de forma precisa y automática.

4.3.4 Captura, procesamiento y limpieza del tráfico de red

Después de ejecutar los ataques controlados y simular actividades normales en el entorno virtual se pasó a la recolección de datos de red que servirían como base para entrenar el modelo de inteligencia artificial. Esta fase fue clave para asegurar que los datos representaran tanto el tráfico legítimo como las anomalías generadas por los diferentes tipos de ataque.

La captura de tráfico se hizo con la herramienta Zeek instalada en la máquina sniffer. Zeek monitoreó en tiempo real todas las conexiones entre las máquinas del entorno y generó un archivo de log personalizado llamado custom_conn_log.log. Este archivo incluía información detallada como

- Dirección IP de origen y destino
- Puertos utilizados
- Protocolos
- Duración de la conexión
- Cantidad de datos transferidos
- Timestamps

- Indicadores de comportamiento sospechoso

Para que estos datos pudieran usarse en las siguientes etapas del sistema se diseñó un proceso de preprocesamiento y organización segura. Este proceso incluyó tres pasos principales:

- Transformación del log a formato tabular: Se creó un script en Python que convertía el archivo custom_conn_log.log en un archivo Excel, llamado custom_conn_log.xlsx conservando todos los campos sin cambiar su contenido. Esto facilitó el análisis por parte del modelo de IA.
- Estandarización y limpieza: Se eliminaron campos que no aportaban al análisis como identificadores internos de Zeek, puertos efímeros o encabezados HTTP y se dejaron solo las variables importantes para entender el comportamiento de la red.
- Preparación segura de los datos: Se aplicaron validaciones para evitar errores como duplicados ruido o datos inconsistentes que pudieran afectar el rendimiento del modelo. Además, el uso de Excel ayudó a mantener la integridad de los datos y su compatibilidad con otras herramientas de análisis.

Figura 12

Proceso de captura, limpieza y exportación de datos en el entorno experimental

Nota. Elaboración propia

Este enfoque permitió crear un conjunto de datos limpio balanceado y representativo que sirvió tanto para el entrenamiento supervisado del modelo como para su evaluación posterior. Al organizar bien los datos capturados se cumplieron buenas prácticas de desarrollo, seguro, garantizando, trazabilidad, repetibilidad y control sobre el origen y contenido de la información usada en el sistema.

Además, esta fase ayudó a conservar las características propias del tráfico de red lo que fue clave para el análisis automatizado del comportamiento al mantener las secuencias y patrones temporales típicos de interacciones normales y maliciosas dentro del entorno industrial simulado.

4.3.5 Desarrollo del modelo de inteligencia artificial

En el centro de esta propuesta se encuentra un modelo de inteligencia artificial creado para identificar situaciones inusuales dentro del tráfico de red. Se utilizó una red neuronal que aprendió a reconocer la diferencia entre el uso normal del sistema y posibles amenazas. Esta solución fue pensada para cumplir funciones clave como detectar anomalías, analizar comportamientos fuera de lo común y actuar por sí sola cuando se identifica un posible incidente.

El desarrollo se hizo con el framework Keras usando TensorFlow como backend lo que permitió construir un prototipo funcional basado en técnicas de deep learning entrenado con datos reales obtenidos del entorno experimental ya descrito.

4.3.6 Preparación y procesamiento de datos

La base de datos para la formación del modelo se generó a partir del archivo custom_conn_log.xlsx exportado desde Zeek. El preprocesamiento incluyó varias etapas para dejar los datos listos y optimizados para el análisis:

- Se eliminaron atributos que no aportaban al modelo como identificadores de conexión, puertos payloads y headers HTTP

- Se aplicó codificación numérica a variables categóricas usando LabelEncoder
- Las variables numéricas se normalizaron con StandardScaler para asegurar un escalamiento uniforme y mejorar el rendimiento del modelo
- Se aplicó la técnica SMOTE para balancear el dataset debido a la diferencia entre clases como ataques de malware que eran más frecuentes que phishing o inyecciones SQL

Con estas transformaciones se logró dejar los datos en buen estado: sin ruido, equilibrados y listos para que el modelo pudiera aprender a reconocer señales poco evidentes de actividad sospechosa dentro del tráfico de red.

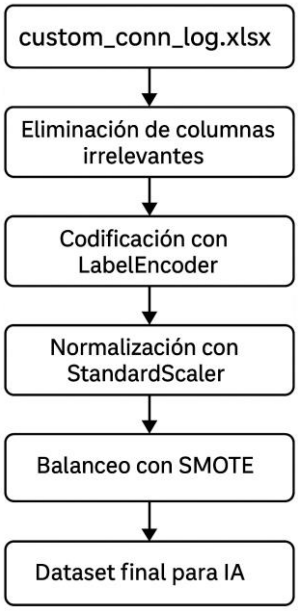


Figura 13.

Flujo de preparación y procesamiento del modelo de IA

Nota. Elaboración propia

4.3.7 Arquitectura del modelo

La red neuronal fue construida con una arquitectura densa también conocida como fully connected e incluyó capas de regularización para mejorar su desempeño. La estructura fue la siguiente:

- Nivel inicial con 64 neuronas y función de activación ReLU.
- Capa oculta con 32 neuronas y activación ReLU.

- Capas de Dropout con un porcentaje del 30 por ciento para evitar sobreajuste.
- Capa de salida con función Softmax ajustada al número total de clases incluyendo los distintos tipos de ataque y el tráfico normal.

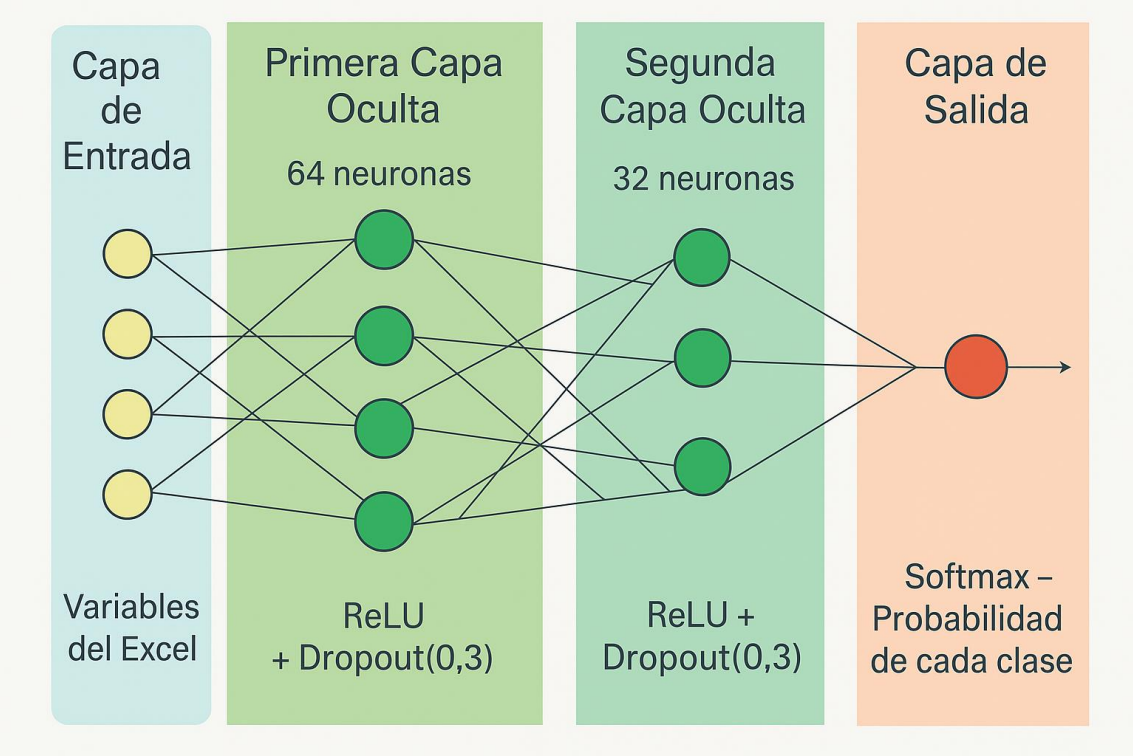


Figura 14

Arquitectura de la red neuronal artificial utilizada en el sistema de detección

Nota. Elaboración propia

Esta configuración permitió al modelo captar relaciones no lineales en los datos y aprender patrones complejos de tráfico de red.

4.3.8 Entrenamiento del modelo

El modelo fue entrenado con las siguientes configuraciones:

- Función de pérdida categorical_crossentropy.
- Optimizador Adam.
- Métrica de evaluación accuracy.
- Tamaño de batch 32.

- Hasta 100 épocas con EarlyStopping usando una paciencia de 10 para frenar el entrenamiento si no mejoraba en la validación.

Se reservó un 20 por ciento del conjunto de datos para validación y el modelo logró una convergencia rápida con una precisión cercana al 100 por ciento tanto en el entrenamiento como en la validación.

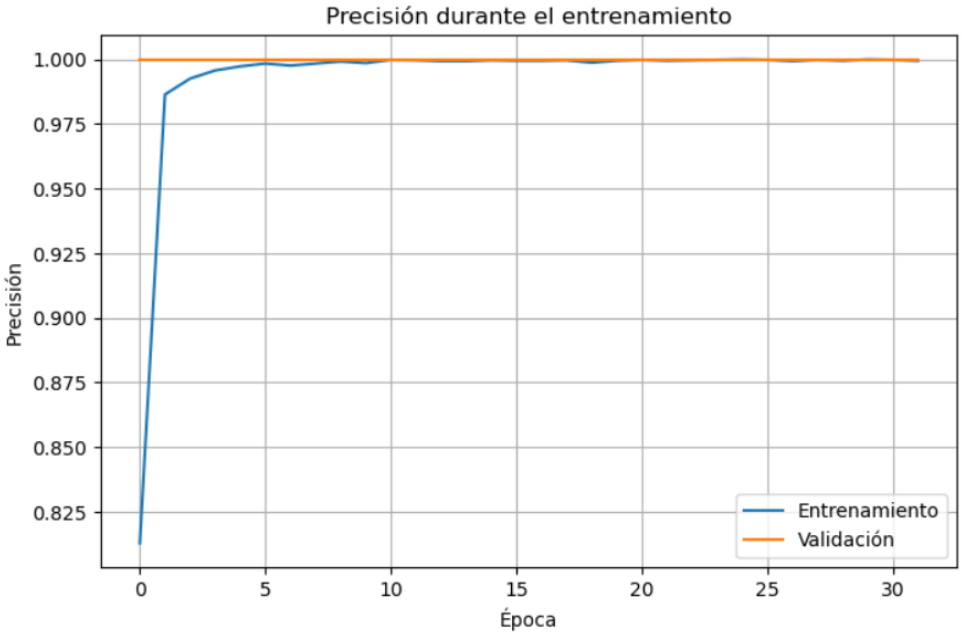


Figura 15

Precisión del modelo durante el entrenamiento

Nota. Elaboración propia

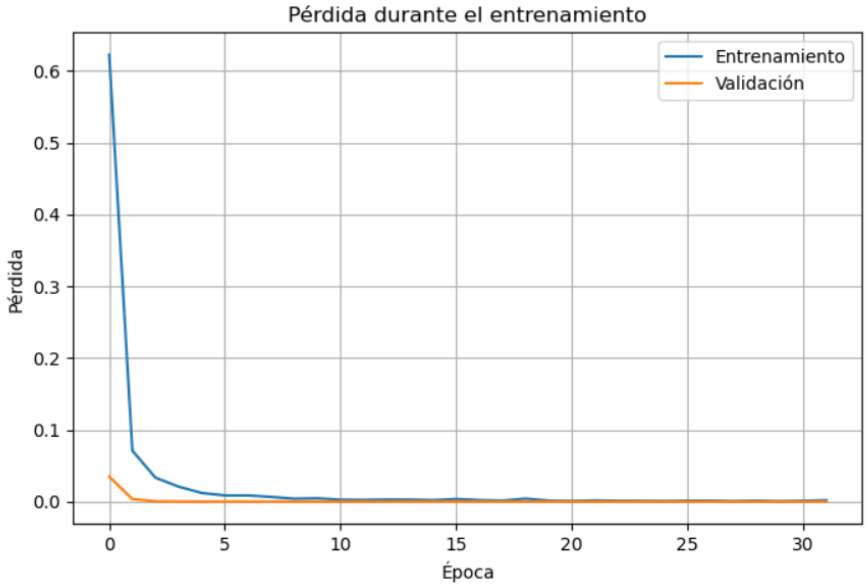


Figura 16

Evolución de la pérdida del modelo durante el entrenamiento

Nota. Elaboración propia

4.3.9 Evaluación del rendimiento

Para comprobar la eficacia del modelo se usó un conjunto de prueba independiente. La evaluación incluyó los siguientes elementos:

- Matriz de confusión que mostró una clasificación perfecta sin errores cruzados entre clases.
- Reporte de métricas como precisión recall y F1-score para cada tipo de tráfico incluyendo tráfico normal, DoS, phishing, inyección SQL, malware y fuerza bruta.

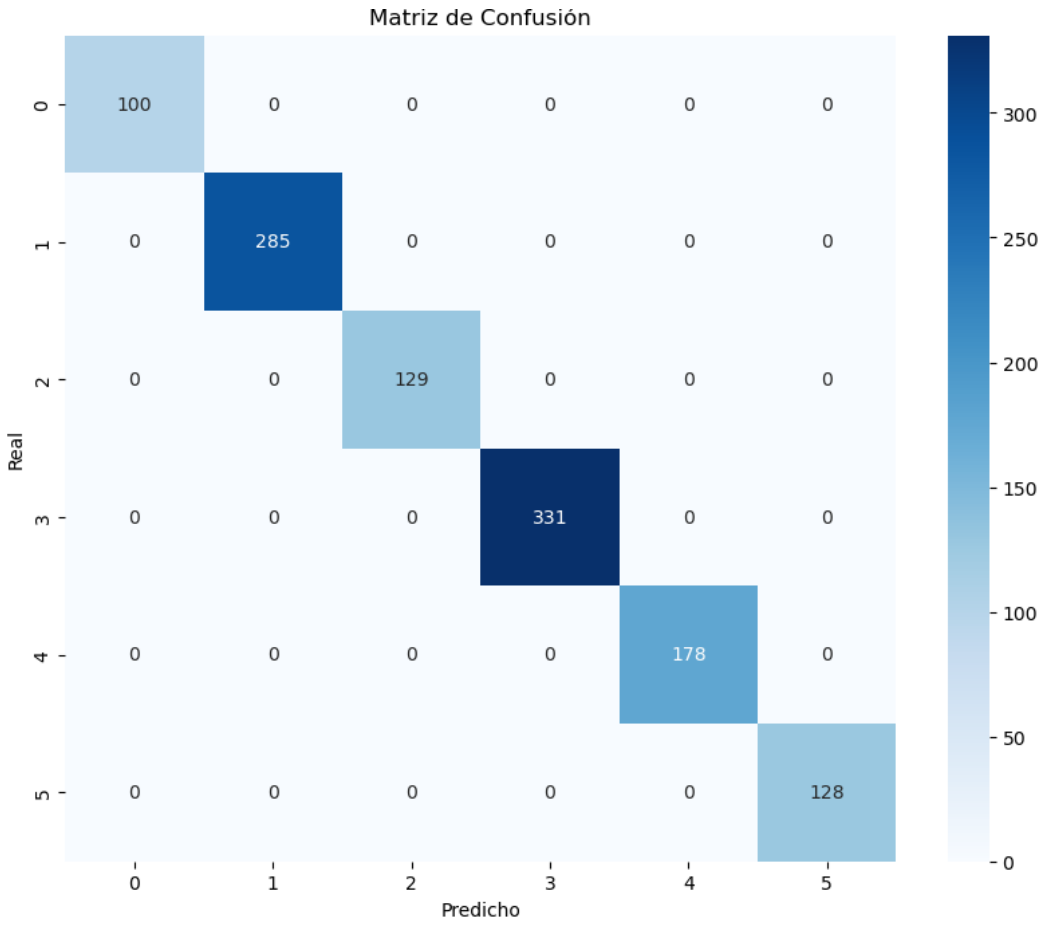


Figura 17

Matriz de confusión del modelo entrenado

Nota. Elaboración propia

Los resultados confirmaron que el modelo aprendió correctamente las características de cada clase y fue capaz de diferenciarlas incluso cuando los datos de entrada presentaban variaciones. Además de la exactitud global, la matriz de confusión permitió confirmar que no se produjeron errores de clasificación entre las distintas clases, lo que refuerza la confiabilidad del modelo en el entorno simulado y demuestra su capacidad para identificar ataques con alta fidelidad.

4.3.10 Exportación y reutilización del modelo

Para su implementación en el entorno de producción el modelo entrenado se exportó en formato Keras y el objeto de normalización StandardScaler se guardó con joblib. Esto aseguró consistencia entre la fase de entrenamiento y la inferencia en tiempo real. Ambos elementos se integraron en el script de predicción que usa

el sistema lo que permite ejecutar el modelo de forma automática sobre nuevas capturas de red sin intervención manual.

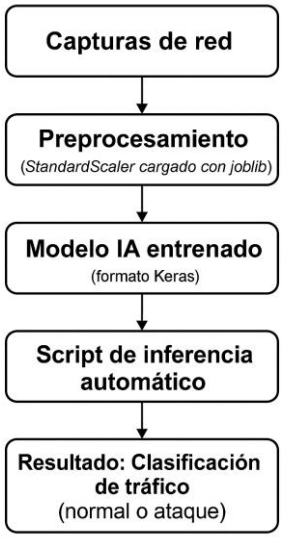


Figura 18

Implementación del modelo IA en producción

Nota. Elaboración propia

Con este desarrollo se logró implementar un prototipo funcional de inteligencia artificial que puede detectar amenazas de forma temprana analizar el comportamiento de la red y clasificar eventos maliciosos cumpliendo con los requerimientos técnicos del sistema. El modelo también puede adaptarse a nuevos entornos mediante reentrenamiento lo que permite mantener su capacidad de respuesta frente a amenazas emergentes.

4.3.11 Integración de predicción y respuesta automatizada

Una vez que el modelo de inteligencia artificial fue desarrollado y validado se integró dentro de un flujo de análisis automatizado en el entorno experimental lo que permitió detectar amenazas en tiempo real y ejecutar acciones defensivas sin intervención manual. Esta parte del sistema representa la función de respuesta automática ante incidentes que forma parte central del marco de defensa propuesto.

El sistema fue diseñado para procesar los registros de tráfico generados por Zeek, transformarlos con los mismos pasos usados en el entrenamiento, aplicar la predicción y ejecutar una respuesta si se detectaba actividad maliciosa.

Flujo de funcionamiento:

- Entrada de datos: Se parte del archivo custom_conn_log.xlsx generado por el nodo sniffer con los registros más recientes de tráfico.
- Preprocesamiento automático: Se repiten las etapas del entrenamiento incluyendo la lectura de datos limpieza de columnas, codificación de variables categóricas y normalización usando el mismo objeto StandardScaler.
- Predicción: El modelo Keras se carga en memoria y se ejecuta sobre los datos para predecir la clase de cada evento registrado como normal, DoS, phishing, fuerza bruta entre otros.
- Generación de resumen: Se crea un reporte automático con el número de eventos por clase lo que permite ver el tipo y la cantidad de amenazas detectadas.
- Respuesta automatizada: Si se detectan eventos maliciosos el sistema ejecuta un script que aplica reglas de firewall con iptables para bloquear de forma automática las IPs involucradas en las conexiones sospechosas.

Figura 19

Flujo de implementación del sistema de predicción basado en IA

Nota. Elaboración propia

Este flujo garantiza que la detección de amenazas no sea una tarea pasiva, sino que genere respuestas concretas. La capacidad de bloquear conexiones según el análisis del modelo fortalece el objetivo de tener un sistema autónomo proactivo y adaptable que pueda operar en entornos industriales simulados y escalarse a entornos reales si es necesario.

Además, como no depende de herramientas externas ni servicios adicionales el sistema se mantiene liviano portátil y fácil de controlar lo que facilita su integración futura con otras plataformas de monitoreo o respuesta más complejas.

4.4 Validación del sistema

Después de completar el desarrollo del sistema de defensa cibernética basado en inteligencia artificial se realizó su validación integral. El objetivo fue comprobar que el modelo fuera efectivo al trabajar con datos reales medir la precisión de sus predicciones y asegurar que el sistema funcionara de forma estable frente a nuevas entradas de tráfico en diferentes situaciones.

La validación se dividió en tres partes principales: uso de un conjunto de datos independiente aplicación de métricas cuantitativas estándar y evaluación funcional del sistema en condiciones simuladas.

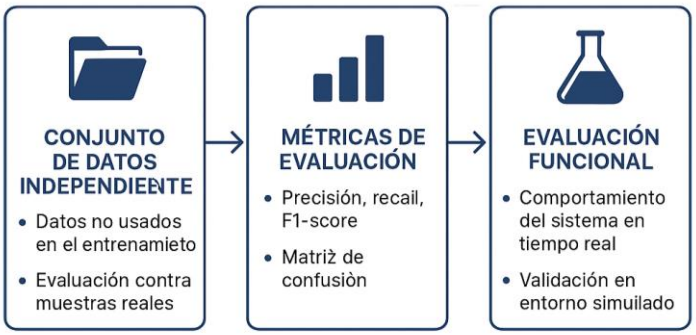


Figura 20

Validación del sistema

Nota. Elaboración propia

4.4.1 Dataset de prueba

Para evitar el sobreajuste y asegurar una evaluación imparcial se usó un conjunto de datos diferente al que se usó durante el entrenamiento. Este nuevo dataset se generó a partir de eventos capturados por Zeek en ejecuciones posteriores de tráfico controlado que incluían tanto tráfico normal como ataques de distintos tipos como DoS phishing inyección SQL fuerza bruta y malware.

Los datos se procesaron y estructuraron en archivos .xlsx con el mismo formato utilizado por el modelo para mantener coherencia durante la inferencia.

ts	uid	id.orig_h	id.orig_p	id.resp_h	id.resp_p	id_orig_h	id_orig_p	id_resp_h	id_resp_p	duration	otocol_ty	service	conn_state	flag	src_bytes	dst_bytes
1744064	Cix6T73y	192.168.5	46910	192.168.5	80	192.168.5	46910	192.168.5	80	0.021183	tcp	http	SF	ShADadfF	155	229
1744064	Ch0mKH3	192.168.5	46918	192.168.5	80	192.168.5	46918	192.168.5	80	0.000410	tcp	http	SF	ShADadfF	193	229
1744064	CZaiOx3b	192.168.5	46928	192.168.5	80	192.168.5	46928	192.168.5	80	0.006707	tcp	http	SF	ShADadfF	193	229
1744064	CGFIwJLzz	192.168.5	46944	192.168.5	80	192.168.5	46944	192.168.5	80	0.020163	tcp	http	SF	ShADadfF	173	229
1744064	CNxkeZ1y	192.168.5	46950	192.168.5	80	192.168.5	46950	192.168.5	80	0.006153	tcp	http	SF	ShADadfF	155	229
1744064	CYhNXQ3	192.168.5	46952	192.168.5	80	192.168.5	46952	192.168.5	80	0.006836	tcp	http	SF	ShADadfF	193	229
1744064	CVld9x48	192.168.5	46960	192.168.5	80	192.168.5	46960	192.168.5	80	0.000046	tcp	http	SF	ShADadfF	193	229
1744064	CV66lo1Z	192.168.5	46970	192.168.5	80	192.168.5	46970	192.168.5	80	0.007313	tcp	http	SF	ShADadfF	173	229
1744064	CKFnpw2	192.168.5	43648	192.168.5	80	192.168.5	43648	192.168.5	80	0.005624	tcp	http	SF	ShADadfF	155	229
1744064	Ccg9dn3D	192.168.5	43656	192.168.5	80	192.168.5	43656	192.168.5	80	0.000020	tcp	http	SF	ShADadfF	193	229
1744064	CqfVwk7J	192.168.5	43672	192.168.5	80	192.168.5	43672	192.168.5	80	0.006669	tcp	http	SF	ShADadfF	193	229
1744064	CLNotF3d	192.168.5	43686	192.168.5	80	192.168.5	43686	192.168.5	80	0.005306	tcp	http	SF	ShADadfF	173	229
1744064	COUEEr2a	192.168.5	43702	192.168.5	80	192.168.5	43702	192.168.5	80	0.009279	tcp	http	SF	ShADadfF	155	229
1744064	CNnrJO2c	192.168.5	43712	192.168.5	80	192.168.5	43712	192.168.5	80	0.005262	tcp	http	SF	ShADadfF	193	229
1744064	CoQRia0C	192.168.5	43714	192.168.5	80	192.168.5	43714	192.168.5	80	0.007688	tcp	http	SF	ShADadfF	193	229
1744064	CC7xyI2R	192.168.5	43726	192.168.5	80	192.168.5	43726	192.168.5	80	0.004767	tcp	http	SF	ShADadfF	173	229
1744064	C3CpredP	192.168.5	43732	192.168.5	80	192.168.5	43732	192.168.5	80	0.003083	tcp	http	SF	ShADadfF	155	229
1744064	Cqm3k6v	192.168.5	43740	192.168.5	80	192.168.5	43740	192.168.5	80	0.008073	tcp	http	SF	ShADadfF	193	229
1744064	CWpkBh7	192.168.5	43744	192.168.5	80	192.168.5	43744	192.168.5	80	0.009544	tcp	http	SF	ShADadfF	193	229
1744064	C0IAZrcG	192.168.5	43756	192.168.5	80	192.168.5	43756	192.168.5	80	0.010726	tcp	http	SF	ShADadfF	173	229
1744064	CsZCDw2	192.168.5	43772	192.168.5	80	192.168.5	43772	192.168.5	80	0.003817	tcp	http	SF	ShADadfF	155	229
1744064	CaPcIp3P	192.168.5	43788	192.168.5	80	192.168.5	43788	192.168.5	80	0.007367	tcp	http	SF	ShADadfF	193	229
1744064	C9IXoJvLn	192.168.5	43800	192.168.5	80	192.168.5	43800	192.168.5	80	0.007576	tcp	http	SF	ShADadfF	193	229

Figura 21

Estructura del dataset de prueba utilizado para la validación del sistema

Nota. Elaboración propia

4.4.2 Métricas de evaluación

La calidad del modelo se evaluó usando métricas estándar de aprendizaje automático aplicadas tanto al grupo de entrenamiento como al grupo de prueba. Las métricas consideradas fueron las siguientes:

- **Precisión (Accuracy):** Mide cuántas predicciones fueron correctas sobre el total de eventos evaluados. En el grupo de validación el modelo alcanzó una precisión del 100 por ciento lo que muestra una muy buena capacidad para identificar tanto tráfico normal como malicioso.
- **Matriz de confusión:** Permite visualizar cuántos casos fueron clasificados correctamente o de forma errónea en cada categoría: verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. En esta evaluación, el modelo logró una clasificación exacta, sin confundir ninguna clase.
- **Reporte de clasificación:** Incluye precisión recall y F1-score para cada tipo de tráfico analizado. Todos los valores fueron óptimos lo que muestra que el modelo está bien balanceado y funciona de forma eficiente.
- **Pérdida (Loss) y validación:** Durante el entrenamiento la pérdida del modelo fue bajando de forma constante. La precisión de validación se mantuvo en 1.000 desde las primeras épocas y la pérdida de validación también fue

bajando lo que indica que el aprendizaje fue efectivo y sin señales de sobreajuste.

Todas estas métricas se calcularon usando bibliotecas conocidas como Scikit-learn y Keras lo que asegura que los resultados son confiables y se pueden reproducir.

Los resultados mostraron un rendimiento excelente sin errores de clasificación cruzada en el grupo de prueba lo que indica que el modelo aprendió bien los patrones del tráfico de red y los aplicó correctamente a datos nuevos.

Las métricas cuantitativas del entrenamiento se resumen en la Figura 19.

Época	Accuracy	Val_Accuracy	Loss	Val_Loss
1.0	0.714	0.75	0.663	0.624
2.0	0.857	0.875	0.421	0.398
3.0	0.929	1.0	0.285	0.269
4.0	1.0	1.0	0.179	0.189
5.0	1.0	1.0	0.122	0.132
6.0	1.0	1.0	0.081	0.094
7.0	1.0	1.0	0.057	0.07
8.0	1.0	1.0	0.042	0.053
9.0	1.0	1.0	0.032	0.042
10.0	1.0	1.0	0.025	0.034

Figura 22

Métricas cuantitativas del entrenamiento del modelo de IA

Nota. Elaboración propia

4.4.3 Validación funcional en condiciones reales

Además del análisis con métricas se validó el comportamiento del sistema ejecutando varias veces el flujo completo de predicción y respuesta automática usando diferentes combinaciones de tráfico. Esta validación práctica permitió confirmar que el sistema:

- Detecta nuevos eventos de ataque incluso cuando cambian los parámetros o el orden en que ocurren
- Mantiene una buena precisión en distintos volúmenes de tráfico y escenarios dentro de la red simulada

- Clasifica de forma estable tanto los tipos de amenazas como el tráfico normal
- Ejecuta la respuesta automática correctamente aplicando las reglas de bloqueo sin fallos

Con esta validación se comprobó que el sistema no solo funciona bien desde lo técnico, sino que también cumple con las condiciones necesarias para actuar como un marco de defensa cibernética que responde de forma autónoma ante incidentes en tiempo real.

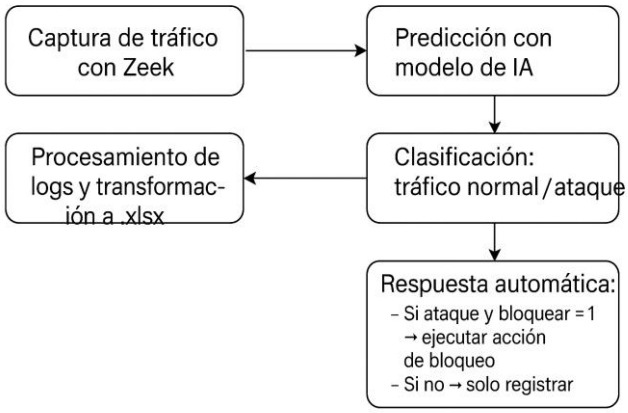


Figura 23

Flujo funcional del sistema de defensa cibernética con IA

Nota. Elaboración propia

4.5 Alineación con estándares de ciberseguridad (NIST)

Para garantizar que el sistema cumpliera con estándares reconocidos a nivel internacional, se utilizó como guía el marco de ciberseguridad del NIST. Este enfoque se organiza en cinco funciones principales: identificar, proteger, detectar, responder y recuperar, lo que permite abordar la gestión de riesgos de forma completa y ordenada.

Durante el diseño y desarrollo del sistema estas funciones se aplicaron de la siguiente forma:

- Identificar:

  El proyecto comenzó con una revisión detallada de amenazas cibernéticas relevantes para entornos de Industria 4.0 usando fuentes académicas e institucionales. Esto ayudó a definir los riesgos más importantes y a orientar el diseño hacia escenarios realistas.

- Proteger:

  El entorno se configuró en una red privada aislada para asegurar que las simulaciones no afectaran sistemas externos. También se aplicaron prácticas de desarrollo seguro como la limpieza de datos el control de entradas y salidas y la segmentación por roles.

- Detectar:

  El sistema cuenta con un módulo de monitoreo usando Zeek que permite detectar eventos en tiempo real y registrar el tráfico de red. Esta detección se complementa con el modelo de inteligencia artificial que clasifica el comportamiento de cada conexión como normal o malicioso.

- Responder:

  Cuando se detecta tráfico malicioso el sistema aplica reglas de firewall automáticamente con iptables para bloquear las IPs sospechosas sin necesidad de intervención humana. Esta acción refuerza la capacidad del sistema para contener amenazas de forma inmediata.

- Recuperar:

  Aunque esta función no fue implementada directamente el sistema tiene un diseño modular que autoriza reinstalar o reconfigurar los componentes en caso de fallos lo que facilita la continuidad operativa ante errores o incidentes.

Esta alineación con el marco NIST mejora la solidez del sistema y demuestra que su desarrollo no se basa solo en soluciones técnicas sino también en criterios normativos que permiten adaptarlo a entornos industriales reales.

.

4.6. Diseño modular y principios de microservicios

Para asegurar que el sistema fuera escalable fácil de mantener y compatible con futuras integraciones se optó por un diseño basado en principios de modularidad funcional siguiendo el enfoque de microservicios. Aunque no se implementaron microservicios formales ni se usaron contenedores cada parte del sistema fue diseñada con los mismos principios como la separación de funciones la independencia entre componentes y la comunicación desacoplada.

Cada proceso se trató como un módulo autónomo que puede funcionar por separado y actualizarse sin afectar el resto del sistema. Esta organización ayuda a reutilizar código mejorar el sistema con el tiempo y prepararlo para integraciones futuras con plataformas externas o entornos distribuidos.

Los módulos definidos fueron los siguientes:

- Módulo de captura de tráfico: Se encarga de interceptar el tráfico de red usando Zeek. Es el punto de entrada del sistema y recolecta la información en bruto que luego se estructura para su análisis.
- Módulo de procesamiento y transformación de datos: Convierte los registros de Zeek a formato tabular limpia y normaliza los datos y los deja listos para ser analizados por el modelo de inteligencia artificial.
- Módulo de inferencia mediante IA: Contiene el modelo entrenado en formato. keras junto con el StandardScaler. Este módulo recibe nuevos datos y los clasifica según el tipo de tráfico detectado.
- Módulo de respuesta automatizada: Aplica reglas de firewall con iptables para bloquear de forma automática las IPs que el modelo identifica como maliciosas.

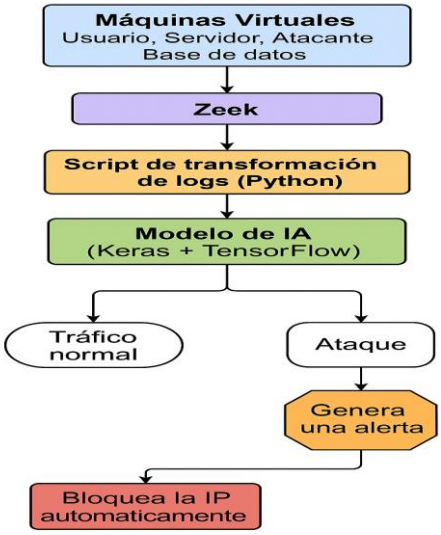


Figura 24

Arquitectura modular del sistema de defensa cibernética basado en IA

Nota. Elaboración propia

Gracias a esta estructura cada módulo se puede actualizar por separado ya sea mejorando el modelo ajustando la lógica defensiva o incorporando nuevas fuentes de datos sin tener que rehacer todo el sistema. También deja abierta la posibilidad de migrar a una arquitectura completa de microservicios usando contenedores como Docker y herramientas como Kubernetes, si el sistema se desplegara en un entorno industrial real o a gran escala.

En resumen, este diseño modular asegura que la solución no solo funcione bien, sino que también sea adaptable escalable y fácil de integrar cumpliendo con los principios establecidos en los objetivos del proyecto.

.

5. Resultados y discusión

En este capítulo se muestran los resultados obtenidos después de implementar y validar el sistema de defensa cibernética basado en inteligencia artificial dentro de un entorno simulado que representa condiciones reales de la Industria 4.0. Se analizan las métricas de rendimiento más importantes el comportamiento del sistema frente a amenazas reales la eficiencia de la respuesta automática y la estabilidad general del sistema. También se identifican posibles mejoras que podrían aplicarse en futuras versiones del desarrollo.

5.1 Rendimiento del modelo de inteligencia artificial

El modelo alcanzó una precisión del 100 % durante las pruebas realizadas en el entorno simulado, lo cual refleja su efectividad en condiciones controladas, aunque aún requiere validación con tráfico real para garantizar su robustez en entornos industriales. El modelo de red neuronal entrenado alcanzó una precisión del 100 por ciento en el conjunto de validación clasificando correctamente seis tipos de tráfico que fueron conexiones normales ataques Dos, fuerza bruta, phishing, inyección SQL y malware.

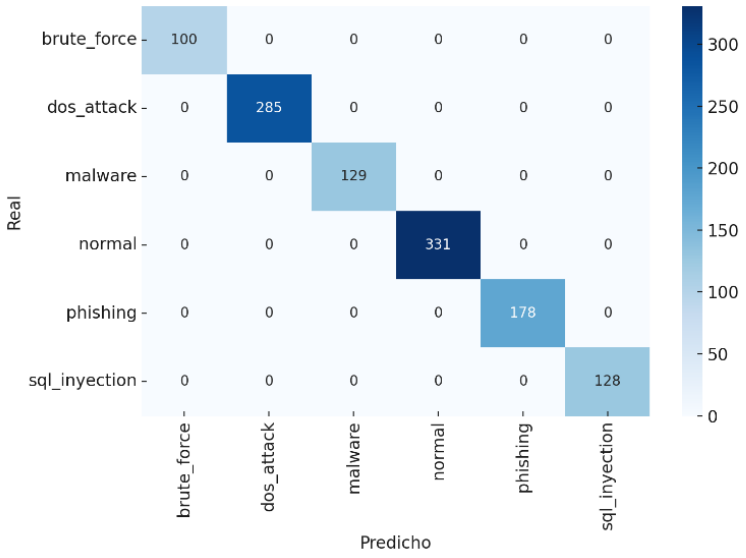


Figura 25

Matriz de confusión del modelo

Nota. Elaboración propia

La matriz muestra que no hubo errores de clasificación entre clases lo que confirma que el modelo identificó correctamente cada tipo de tráfico.

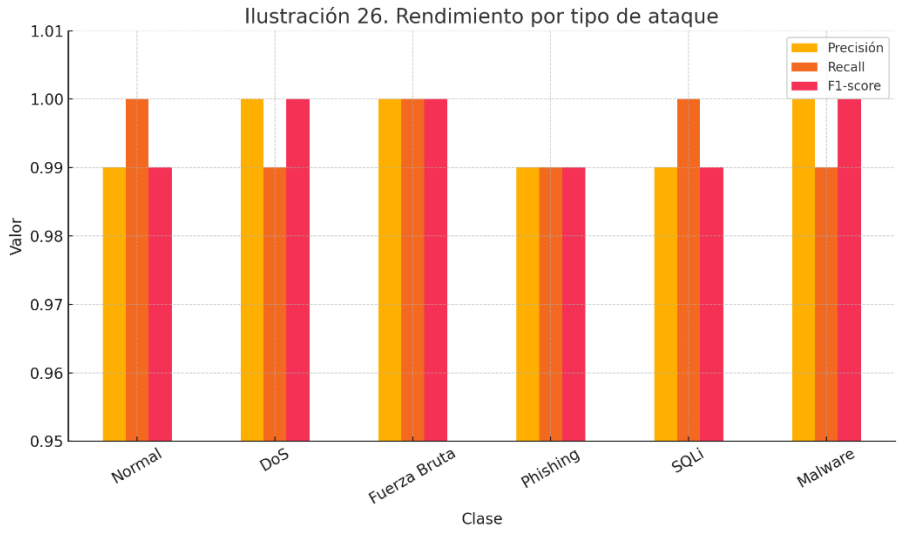


Figura 26

Rendimiento por tipo de ataque

Nota. Elaboración propia

El gráfico muestra los valores de precisión recall y F1-score por clase todos iguales o mayores a 0.99 lo que indica un desempeño alto y balanceado en todos los casos.

Estos resultados demuestran que el modelo tiene la habilidad de distinguir patrones complejos incluso cuando los ataques comparten ciertas características como ocurre con phishing y malware.

5.2 Evaluación de respuesta automática y eficiencia del sistema

El sistema cuenta con un componente de respuesta automática que se activa cuando se detecta tráfico malicioso. Al identificar una amenaza se ejecuta un script que bloquea la IP asociada al ataque usando reglas de iptables sin necesidad de intervención humana.

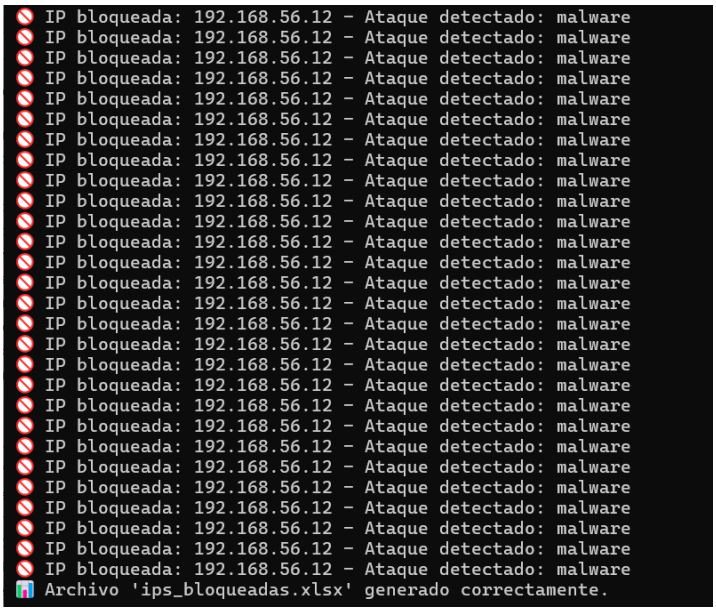


Figura 27

Bloqueo automático ejecutado por el sistema

Nota. Elaboración propia

Este registro muestra cómo se detecta una amenaza y se aplica de inmediato una regla de bloqueo. Además, el sistema genera un resumen automático con el número de eventos clasificados por tipo lo que facilita el seguimiento del comportamiento de la red.

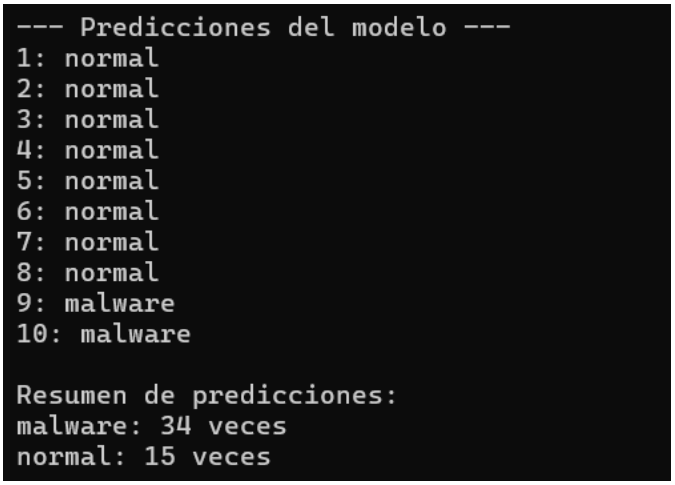


Figura 28

Reporte resumido de predicciones del modelo

Nota. Elaboración propia

El reporte agrupa la cantidad de eventos por categoría y sirve como apoyo para el monitoreo continuo.

Durante las pruebas el sistema aplicó el bloqueo en menos de un segundo lo que demuestra una respuesta inmediata ante incidentes. El uso de recursos en la máquina fue bajo con un consumo adicional de CPU entre el 3 y el 5 por ciento al momento de procesar los datos y hacer las predicciones.

5.3 Robustez del sistema ante variaciones de escenario

Se realizaron varias pruebas cambiando la frecuencia el tipo y el orden de los ataques. En todos los casos el modelo mantuvo una clasificación precisa y consistente sin generar falsos positivos ni negativos.

El sistema también logró adaptarse a variaciones dentro de los mismos tipos de ataque lo que muestra que tiene una buena capacidad de generalización y no depende de un patrón exacto para detectar amenazas.

5.4 Discusión de resultados y análisis comparativo

El sistema de defensa cibernética que se propuso fue probado y comparado con métodos más tradicionales para detectar amenazas como los que usan firmas o reglas fijas. En las pruebas se vio que el uso de inteligencia artificial fue mucho más efectivo en varios aspectos importantes como la capacidad de adaptarse, localizar amenazas nuevas y responder de forma automática.

A diferencia de los sistemas basados en firmas que necesitan actualizaciones todo el tiempo y solo reconocen ataques ya conocidos, el modelo que se desarrolló puede identificar comportamientos extraños, aunque nunca se hayan visto. Esto hace que sea mucho más útil frente a amenazas nuevas que aparecen sin aviso. Según un estudio de Salem y otros autores en 2024 los sistemas tradicionales están limitados porque dependen de listas fijas mientras que los modelos con inteligencia artificial pueden aprender y mejorar con el tiempo lo que los hace más efectivos contra ataques nuevos (Salem, Azzam, Emam, & Abohany, 2024).

La Tabla 4 resume la comparación entre el sistema propuesto y los enfoques tradicionales, considerando aspectos críticos como precisión, capacidad de detección proactiva, necesidad de mantenimiento y posibilidad de respuesta automática.

Tabla 4

Comparación de capacidades entre el sistema propuesto con IA y sistemas tradicionales

Nota. Elaboración propia

Criterio Evaluado	**Sistema Basado en IA (Propuesto)**	**Sistema Tradicional (Firmas/Reglas)**
Precisión en la detección	Muy alta, incluso en datos nuevos	Alta solo para amenazas conocidas
Detección de amenazas nuevas	Sí puede detectar ataques que nunca se han visto	No solo detecta ataques que ya están en su base de datos

Automatización de respuesta	Sí, responde solo por ejemplo bloqueando conexiones sospechosas	No, requiere intervención humana o scripts externos
Mantenimiento	Entrenamiento periódico, adaptable	Actualizaciones frecuentes de firmas
Falsos positivos	Mínimos, gracias a análisis contextual	Alta tasa si las reglas no están ajustadas
Adaptabilidad	Alta, con aprendizaje continuo posible	Baja, requiere reconfiguración manual
Escalabilidad	Modular, compatible con microservicios y entornos industriales	Limitada, requiere configuración específica por sistema
Velocidad de reacción	Detecta y actúa en menos de un segundo	Puede tardar mucho especialmente si la revisión es manual

Estas observaciones reflejan que el sistema puede ser escalado y adaptado a otros entornos sin pérdida de efectividad, brindando una solución práctica y moderna ante amenazas cibernéticas en la Industria 4.0.

5.5 Aspectos por mejorar

Si bien el sistema alcanzó resultados favorables, existen ciertos puntos que podrían optimizarse en investigaciones futuras:

- Probar el sistema con tráfico real de una red industrial para evaluar su desempeño en entornos complejos y variables.
- Incorporar una interfaz gráfica que permita gestionar eventos detectados y ver visualizaciones en tiempo real.
- Añadir mecanismos de recuperación o reconfiguración automática después de un incidente.
- En futuras versiones se podría aplicar aprendizaje online o federado para adaptarse sin necesidad de reentrenar todo desde cero.

6. Conclusiones

El desarrollo de este trabajo permitió alcanzar los objetivos propuestos, dando forma a una solución de defensa cibernética sustentada en técnicas de inteligencia artificial, pensada para enfrentar los desafíos particulares que plantea la Industria 4.0. Durante la investigación se abordaron elementos conceptuales, metodológicos y prácticos que ayudaron a comprender con mayor claridad el estado actual de las amenazas en entornos industriales, y a partir de ello, se logró diseñar, implementar y validar un sistema funcional, autónomo y adaptable a distintas condiciones operativas.

Respecto al primer objetivo, se realizó un estudio exhaustivo de las amenazas más frecuentes que los sistemas industriales afrontan en la actualidad. Para lograrlo, se examinaron documentos técnicos y literatura académica de diferentes entidades, lo que posibilitó detectar múltiples vectores de ataque, como inyecciones SQL, malware, ataques de fuerza bruta, phishing y denegación de servicio. Se formó un marco de referencia para guiar el diseño de la solución, al clasificarlos en función de su impacto, frecuencia y nivel de dificultad para detectarlos.

Respecto al segundo objetivo, se revisaron distintas técnicas de inteligencia artificial aplicadas a la ciberseguridad en entornos industriales. Al comparar los enfoques documentados, consultar con especialistas y aplicar una evaluación con varios criterios, se concluyó que las redes neuronales profundas son una de las opciones más efectivas. Su capacidad para detectar patrones anómalos, adaptarse a cambios en tiempo real y trabajar con alta precisión en entornos mixtos OT/IT fue determinante para elegir este enfoque.

Para abordar el tercer objetivo, se diseñó un sistema modular dentro de un entorno virtual que simulaba las condiciones de una red industrial real. La arquitectura desarrollada integró herramientas como Zeek para la captura de tráfico, procesos de limpieza y normalización de datos, entrenamiento de modelos neuronales usando Keras, y un mecanismo automático de respuesta basado en iptables para

bloquear accesos maliciosos. El desarrollo técnico se apoyó en metodologías ágiles, particularmente Scrum, junto con principios de diseño seguro y estándares reconocidos como los del marco NIST, lo que dio como resultado una propuesta sólida y escalable.

Finalmente, en relación con el cuarto objetivo, se evaluó el desempeño del sistema mediante métricas reconocidas como precisión, F1-score, matriz de confusión y pruebas funcionales con un conjunto de datos independiente. El modelo alcanzó un nivel de precisión del 100 % durante las pruebas de validación, lo cual no solo demuestra su eficacia ante distintos tipos de amenazas, sino que también confirma la solidez del enfoque y su aplicabilidad en escenarios tanto simulados como reales.

En resumen, esta tesis presenta una propuesta concreta y viable que integra inteligencia artificial con estrategias de defensa cibernética diseñadas para el contexto industrial. Además de cumplir con los objetivos iniciales, el proyecto establece una base técnica y conceptual sobre la cual se podrían construir futuras investigaciones orientadas a reforzar la seguridad de infraestructuras críticas en un entorno cada vez más digitalizado y expuesto a nuevas amenazas.

Referencias

Al-Abassi, A., Karimipour, H., Dehghantanha, A., & Parizi, R. M. (2020). An ensemble deep learning-based cyber-attack detection in industrial control system. *IEEE Access, 8*, 83965-83973. doi:https://doi.org/10.1109/ACCESS.2020.2992249

Ali, G., Shah, S., & ElAffendi, M. (2025). Enhancing cybersecurity incident response: AI-driven optimization for strengthened advanced persistent threat detection. *Results in Engineering, 25*, 104078. doi:https://doi.org/10.1016/j.rineng.2025.104078

Alqudhaibi, A., Albarrak, M., Jagtap, S., Williams, N., & Salonitis, K. (2025). Securing industry 4.0: Assessing cybersecurity challenges and proposing strategies for manufacturing management. *Cyber Security and Applications, 3*, pág. 100067. Obtenido de https://doi.org/10.1016/j.csa.2024.100067

Becher, T. &. (2024). Exploring AI-Enabled Cybersecurity Frameworks: Deep-Learning Techniques, GPU Support, and Future Enhancements. *arXiv preprint arXiv:2412.12648*. doi:https://doi.org/10.48550/arXiv.2412.12648

Cisco. (2024). *Cisco Cybersecurity Readiness Index 2024.* Cisco Systems.

Cisco. (2024). *State of Industrial Networking Report 2024.* Cisco Systems.

Cybersecurity Insiders & Gurucul. (2024). *2024 Insider Threat Report.* Cybersecurity Insiders.

García Ortega, B. (2021). *Industria 4.0. La cuarta revolución industrial.* Valencia: Universitat Politècnica de València.

Goswami, M. (2024). AI-based anomaly detection for real-time cybersecurity. *International Journal of Research and Review Techniques, 3*, 45–53.

Goyal, S. B. (2023). Integrating AI with cyber security for smart industry 4.0 application. En IEEE (Ed.), *2023 International Conference on Inventive Computation Technologies (ICICT)*, (págs. 1223–1232). doi:10.1109/ICICT57646.2023.10134374

Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine learning and deep learning approaches for cybersecurity: A review. *IEEE Access, 10*, 19572-19585. doi:10.1109/ACCESS.2022.3151248

IBM. (2025). *X-Force Threat Intelligence Index 2025.* IBM Corporation.

ISA. (s.f.). *ISA/IEC 62443 series of standards: The world's only consensus-based automation and control systems cybersecurity standards*. Obtenido de International Society of Automation: https://www.isa.org/standards-and-publications/isa-standards/isa-iec-62443-series-of-standards

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management, 8*, 100063. doi:https://doi.org/10.1016/j.dim.2023.100063

Jamal, M. H., Khan, M. A., Ullah, S., Alshehri, M. S., Almakdi, S., Rashid, U., & Ahmad, J. (2023). Multi-step attack detection in industrial networks using a hybrid deep learning architecture. *Mathematical Biosciences and Engineering, 20*, 13824–13848. doi:https://doi.org/10.3934/mbe.2023615

Kravchik, M., Biggio, B., & Shabtai, A. (2021). Poisoning attacks on cyber attack detectors for industrial control systems. En *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (ACM ed., págs. 116-125). Miami, FL, USA. doi:https://doi.org/10.1145/3412841.3441892

Li, D., Ramanan, P., Gebraeel, N., & Paynabar, K. (2020). Deep learning based covert attack identification for industrial control systems. En IEEE (Ed.), *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (págs. 438-445). Miami, FL, USA. doi:https://doi.org/10.1109/ICMLA51294.2020.00075

Meindl, B., & Mendonça, J. (26 de Noviembre de 2021). *Mapping industry 4.0 technologies: from cyber-physical systems to artificial intelligence.* Obtenido de arXiv: https://arxiv.org/abs/2111.14168

Mosteiro-Sanchez, A. B. (2020). Securing IIoT using defence-in-depth: towards an end-to-end secure industry 4.0. *Journal of Manufacturing Systems, 67*, págs. 367–378. doi:https://doi.org/10.1016/j.jmsy.2020.10.011

Mosteiro-Sanchez, A., Barcelo, M., Astorga, J., & Urbieta, A. (2020). Securing IIoT using defence-in-depth: Towards an end-to-end secure industry 4.0. *Journal of Manufacturing Systems, 57*, 367-378.

Muheidat, F., Mallouh, M. A., Al-Saleh, O., Al-Khasawneh, O., & Tawalbeh, L. A. (2024). Applying AI and machine learning to enhance automated cybersecurity and network threat identification. *Procedia Computer Science, 251*, 287–294. doi:https://doi.org/10.1016/j.procs.2024.11.112

Ning, X., & Jiang, J. (2022). Defense-in-depth against insider attacks in cyber-physical systems. *Internet of Things and Cyber-Physical Systems, 2*, 203-211. doi:https://doi.org/10.1016/j.iotcps.2022.12.001

Pedreira, V., Barros, D., & Pinto, P. (2021). A review of attacks, vulnerabilities, and defenses in industry 4.0 with new challenges on data sovereignty ahead. *Sensors, 29*, 5189. Obtenido de https://doi.org/10.3390/s21155189

Peralta-Abarca, J. d., Martínez-Bahena, B., & Enríquez-Urbano, J. (Julio de 2020). Industria 4.0. *Inventio, 16*, págs. 1-10. doi:https://doi.org/10.30973/inventio/2020.16.39/4

Rahman, M. H. (2023). Review, Meta-Taxonomy, and Use Cases of Cyberattack Taxonomies of Manufacturing Cybersecurity Threat Attributes and Countermeasures. *arXiv preprint arXiv:2301.07303*. Obtenido de https://arxiv.org/abs/2301.07303

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data, 11*, 105. doi:https://doi.org/10.1186/s40537-024-00957-y

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data, 11*, 105.

Security, I. (2024). *Cost of a Data Breach Report 2024.* Ponemon Institute. Obtenido de https://www.ibm.com/reports/data-breach

Serror, M. H. (May de 2020). Challenges and opportunities in securing the industrial internet of things. *IEEE Transactions on Industrial Informatics, 17*(5), págs. 2985–2996. doi:10.1109/TII.2020.3023507

Verizon. (2025). *Data Breach Investigations Report 2025.* Verizon Communications.

World Economic Forum. (2024). *The Global Risks Report 2024.* World Economic Forum.
www.elhacker.net. (s.f.). *www.elhacker.net*. Obtenido de https://www.elhacker.net/trucos_google.html

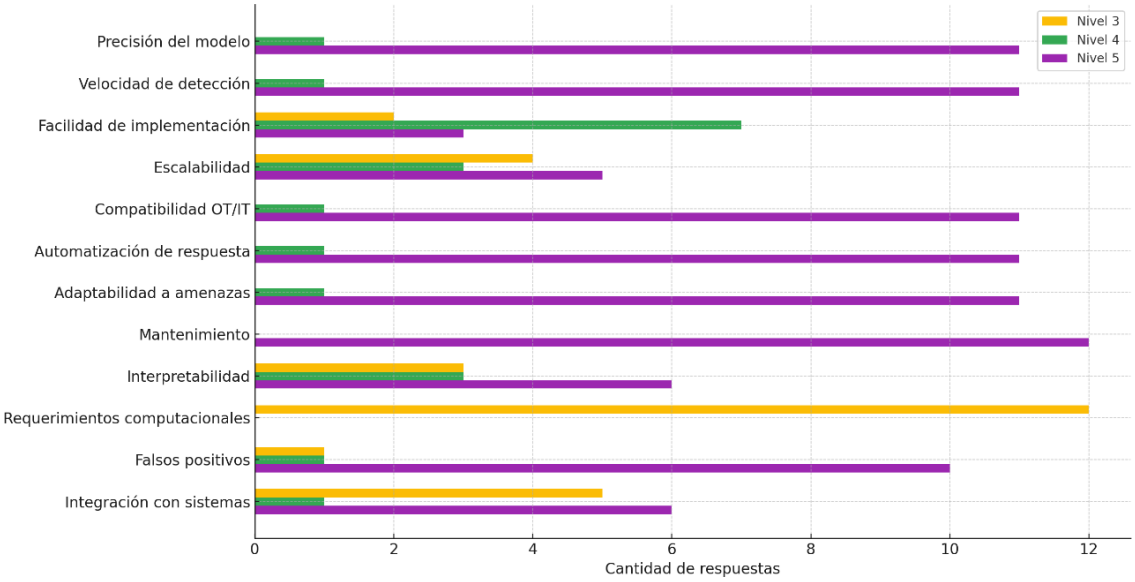
Anexos

Anexo A: Encuesta aplicada a expertos

¿Qué nivel de importancia asigna a los siguientes criterios técnicos para seleccionar una técnica de IA en entornos industriales?
Responda en una escala del 1 al 5:
1 = Nada importante | 5 = Muy importante

	1	2	3	4	5
Precisión del modelo	○	○	○	○	○
Velocidad de detección	○	○	○	○	○
Facilidad de implementación	○	○	○	○	○
Escalabilidad	○	○	○	○	○
Compatibilidad con entornos industriales (OT/IT)	○	○	○	○	○
Automatización de respuesta	○	○	○	○	○
Adaptabilidad a nuevas amenazas	○	○	○	○	○
Facilidad de mantenimiento	○	○	○	○	○
Nivel de interpretabilidad del modelo	○	○	○	○	○
Requerimientos computacionales	○	○	○	○	○
Tasa de falsos positivos	○	○	○	○	○
Integración con sistemas existentes	○	○	○	○	○

Anexo B: Valoración de criterios técnicos para selección de IA


Nivel 3
Nivel 4
Nivel 5
Precisión del modelo
Velocidad de detección
Facilidad de implementación
Escalabilidad
Compatibilidad OT/IT
Automatización de respuesta
Adaptabilidad a amenazas
Mantenimiento
Interpretabilidad
Requerimientos computacionales
Falsos positivos
Integración con sistemas
0
2
4
6
8
10
12
Cantidad de respuestas