UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

TÍTULO:

ANÁLISIS DE IDS CON IA PARA PROTEGER INFRAESTRUCTURA CRÍTICA EN LA CLÍNICA GRANADOS, SANTA ELENA.

AUTOR

CIRSTOIU ROJAS KATY ALINA

TRABAJO DE TITULACIÓN

Previo a la obtención del grado académico en
MAGÍSTER EN CIBERSEGURIDAD

TUTOR

LSI. OSCAR OMAR APOLINARIO ARZUBE, Ph.D.

Santa Elena, Ecuador

Año 2024

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

TRIBUNAL DE SUSTENTACIÓN

Firmado electrónicamente por:
ALICIA GERMANIA ANDRADE VERA

ING. ALICIA ANDRADE VERA, MGTR.
COORDINADORA DEL PROGRAMA

Firmado electrónicamente por:
OSCAR OMAR APOLINARIO ARZUBE

LSI. OSCAR APOLINARIO ARZUBE, Ph.D.
DOCENTE TUTOR

Firmado electrónicamente por:
JAIME BENJAMIN OROZCO IGUASNIA

ING. JAIME OROZCO IGUASNIA, MGTR.
DOCENTE ESPECIALISTA

Firmado electrónicamente por:
DANIEL IVAN QUIRUMBAY YAGUAL

LSI. DANIEL QUIRUMBAY YAGUAL, MGTR.
DOCENTE ESPECIALISTA

Firmado electrónicamente por:
MARIA MARGARITA RIVERA GONZALEZ

ABG. MARÍA RIVERA GONZÁLEZ. MGTR.
SECRETARIA GENERAL
UPSE

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por CIRSTOIU ROJAS KATY ALINA, como requerimiento para la obtención del título de Magister en Ciberseguridad.

Santa Elena, 25 de septiembre de 2024

TUTOR

Firmado electrónicamente por:
OSCAR OMAR APOLINARIO ARZUBE

LSI. OSCAR OMAR APOLINARIO ARZUBE, Ph.D.

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

DECLARACIÓN DE RESPONSABILIDAD

Yo, **KATY ALINA CIRSTOIU ROJAS**

DECLARO QUE:

El trabajo de Titulación, Análisis de IDS con IA para proteger infraestructura crítica en la Clínica Granados, Santa Elena, previo a la obtención del título en Magíster en Ciberseguridad, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Santa Elena, 25 de septiembre de 2024

EL AUTOR

Firmado electrónicamente por:
KATY ALINA CIRSTOIU ROJAS

KATY ALINA CIRSTOIU ROJAS

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE CIENCIAS DE LA INGENIERÍA
INSTITUTO DE POSTGRADO

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado Análisis de IDS con IA para proteger infraestructura crítica en la Clínica Granados, Santa Elena., presentado por el estudiante, CIRSTOIU ROJAS KATY ALINA fue enviado al Sistema Antiplagio COMPILATIO MAGISTER, presentando un porcentaje de similitud correspondiente al 1% y una herramienta adicional turnitin del 3%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.


CERTIFICADO DE ANÁLISIS
magister

2024_Cirstoiu_Rojas_Katy_Alina_ANALISIS_DE_IDS

< 1% Textos sospechosos

< 1% Similitudes
0% similitudes entre comillas
0% entre las fuentes mencionadas
7% Idiomas no reconocidos (ignorado)

Nombre del documento: 2024_Cirstoiu_Rojas_Katy_Alina_ANALISIS_DE_IDS.docx
ID del documento: 3cd65f930420ef4809896ae3dc1ea9bfda755398
Tamaño del documento original: 1,09 MB
Autores: []

Depositante: OSCAR OMAR APOLINARIO ARZUBE
Fecha de depósito: 11/9/2024
Tipo de carga: interface
fecha de fin de análisis: 11/9/2024

Número de palabras: 15.299
Número de caracteres: 104.782

Ubicación de las similitudes en el documento:

Santa Elena, 25 de septiembre de 2024

TUTOR

Firmado electrónicamente por:
OSCAR OMAR APOLINARIO ARZUBE

LSI. OSCAR OMAR APOLINARIO ARZUBE, Ph.D.

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

AUTORIZACIÓN

Yo, **KATY ALINA CIRSTOIU ROJAS**

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales de mi trabajo de examen de carácter complexivo con fines de difusión pública, además apruebo la reproducción de este trabajo de examen de carácter complexivo dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor.

Santa Elena, 25 de septiembre de 2024

EL AUTOR

Firmado electrónicamente por:
KATY ALINA CIRSTOIU ROJAS

KATY ALINA CIRSTOIU ROJAS

AGRADECIMIENTO

Expreso mi sincera gratitud a mi tutor, cuya guía y dedicación han sido fundamentales en la realización de este trabajo. Su experiencia y apoyo constante no solo enriquecieron esta investigación, sino que también contribuyeron significativamente a mi crecimiento profesional.

Agradezco también a mis docentes, quienes a lo largo de mi vida universitaria se convirtieron en más que educadores, en verdaderos amigos y mentores. Su pasión por la enseñanza, su disposición para compartir conocimientos más allá del aula y su apoyo incondicional han sido pilares cruciales en mi formación académica y personal. Este logro es en gran parte resultado de su invaluable contribución.

Katy Alina Cirstoiu Rojas

DEDICATORIA

Dedicado con profunda gratitud a mi familia, amigos y docentes, pilares fundamentales en este viaje académico. Su apoyo inquebrantable, sabiduría compartida y constante aliento fueron la fuerza impulsora detrás de cada logro. Este trabajo es el fruto de su confianza en mí y un testimonio de lo que se puede alcanzar con el respaldo de quienes creen en nuestros sueños.

Adicional este trabajo está dedicado a todos los profesionales de la salud y de la tecnología que trabajan incansablemente para proteger la integridad y confidencialidad de los datos médicos. Su dedicación inspira la búsqueda de soluciones innovadoras en ciberseguridad, contribuyendo así a un sistema de salud más seguro y eficiente para todos.

Katy Alina Cirstoiu Rojas

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

Este estudio investiga la aplicación de sistemas de detección de intrusiones (IDS) basados en inteligencia artificial (IA) para proteger la infraestructura crítica en la Clínica Granados en Santa Elena. Dado el aumento de amenazas cibernéticas en el sector salud, la seguridad de la infraestructura crítica de la clínica es vital para garantizar la continuidad de los servicios médicos y la protección de datos sensibles. El objetivo es examinar la viabilidad y eficacia de esta tecnología en un entorno de atención médica. La investigación emplea un enfoque mixto, combinando métodos cuantitativos y cualitativos, análisis de datos y simulaciones. Se recopilarán datos propios de la clínica, fuentes externas y datos sintéticos para evaluar el rendimiento del IDS-IA. Los resultados esperados incluyen una evaluación detallada de la efectividad del sistema en la detección de amenazas, su impacto en la eficiencia operativa y los desafíos de implementación. Las conclusiones proporcionarán una base para determinar si el IDS-IA puede ofrecer una protección robusta para infraestructura crítica en entornos de salud.

Palabras claves: Detección de Intrusiones, Inteligencia Artificial, Infraestructura Crítica, Seguridad en Salud, Clínica Granados.

ABSTRACT

This study investigates the application of artificial intelligence (AI) based intrusion detection systems (IDS) to protect critical infrastructure at Clínica Granados in Santa Elena. Given the increase in cyber threats in the healthcare sector, the security of the clinic's critical infrastructure is vital to ensure the continuity of medical services and the protection of sensitive data. The objective is to examine the feasibility and effectiveness of this technology in a healthcare environment. The research employs a mixed approach, combining quantitative and qualitative methods, data analysis, and simulations. Data will be collected from the clinic's own sources, external sources, and synthetic data to evaluate the performance of the AI-IDS. Expected results include a detailed assessment of the system's effectiveness in threat detection, its impact on operational efficiency, and implementation challenges. The conclusions will provide a basis for determining whether AI-IDS can offer robust protection for critical infrastructure in healthcare settings.

Keywords: Intrusion Detection, Artificial Intelligence, Critical Infrastructure, Health Security, Clínica Granados.

INTRODUCCIÓN

En la era digital actual, la seguridad cibernética se ha convertido en un desafío crítico para el sector salud. La Clínica Granados en Santa Elena, como muchas instituciones médicas, maneja una gran cantidad de datos sensibles y depende de infraestructuras tecnológicas complejas para proporcionar atención de calidad. Sin embargo, la creciente sofisticación de las amenazas cibernéticas pone en riesgo la integridad y confidencialidad de esta información vital. Según un estudio reciente, el sector salud experimentó un aumento del 55% en los ciberataques en 2020, siendo uno de los sectores más afectados (Jalali et al., 2020).

En este contexto, los sistemas de detección de intrusiones (IDS) tradicionales han indicado ser insuficientes para lograr hacer frente a la evolución compleja y rápida de las amenazas cibernéticas actuales. Como lo hace referencia (García-Teodoro et al., 2009), "Los IDS convencionales a menudo fallan en la detección de ataques nuevos o variantes de los conocidos, lo que resalta la necesidad de enfoques más avanzados y adaptables". Esta serie de limitaciones han guiado a la investigación de soluciones basadas en inteligencia artificial (IA) como una alternativa novedosa y prometedora para lograr la mejorara de la seguridad en los entornos críticos tales como hospitales y clínicas.

Una representación de innovación significativa en el campo de la ciberseguridad es lo que propone la implementación de IDS basados en IA. Al lograr aprovechar técnicas de aprendizaje automático y análisis de datos avanzados para la detección de patrones de comportamiento anómalos que podrían indicar una intrusión. "Los IDS basados en IA han demostrado una mayor capacidad para adaptarse a nuevas amenazas y reducir la tasa de falsos positivos en comparación con los sistemas tradicionales" mencionado por (Ferrag et al., 2020). La gran capacidad de adaptación que presentan estos sistemas es particularmente crucial en el sector salud, donde las consecuencias de una brecha de seguridad pueden ser devastadoras tanto para los pacientes como para la institución.

Sin embargo, se debe tener en cuenta que el sector de la salud no está exento de desafíos al momento de adaptar las tecnologías de IA en la seguridad cibernética. Temas como los datos que se manejan y su nivel de privacidad, la interpretabilidad de los modelos de IA y la integración con los sistemas existentes deben ser cuidadosamente analizados. Como se indica en el estudio realizado por (Lopez-Martin et al., 2020) en el que enfatiza que "la

implementación exitosa de IDS basados en IA en entornos de salud requiere no solo de avances tecnológicos, sino también de una consideración cuidadosa de los aspectos éticos y regulatorios".

En este contexto, mediante este trabajo de titulación se propone evaluar la eficacia de un sistema de detección de intrusiones basado en IA para mejorar la protección de la infraestructura crítica en la Clínica Granados, en el que se lo compara directamente con los sistemas tradicionales en los puntos importantes como eficiencia, precisión y capacidad de respuesta ante amenazas cibernéticas. Mediante esta investigación se prevé contribuir a la institución mencionada y de la misma manera al conocimiento sobre la aplicación de IA en la seguridad cibernética del sector de la salud en general

1.1 PROBLEMA DE INVESTIGACIÓN

La Clínica Granados en Santa Elena enfrenta un desafío crítico en el ámbito de la ciberseguridad. El problema principal radica en la capacidad limitada de los sistemas de detección de intrusiones (IDS) tradicionales para detectar y responder a amenazas cibernéticas avanzadas en tiempo real, lo que pone en riesgo su infraestructura crítica. A medida que las amenazas evolucionan y se vuelven más sofisticadas, los sistemas tradicionales resultan cada vez más insuficientes para abordar ataques que pueden afectar significativamente la operativa y la seguridad de la clínica.

La rápida evolución de la visualización de las amenazas cibernéticas actuales es una desventaja del problema en cuestión. De acuerdo con el informe de (*Cisco Annual Internet Report - Cisco Annual Internet Report (2018–2023) White Paper - Cisco*, n.d.), el sector salud ha experimentado un aumento del 11% en los ataques de malware y un 6% en los ataques de ransomware durante el último año. Los IDS tradicionales que normalmente están basados en reglas estáticas, tienen conflictos para mantenerse al día con estas amenazas en constante cambio. De la misma manera, la digitalización rápida en el sector salud ha llevado a la implementación de sistemas cada vez más complejos e interconectados. En el estudio de (Kruse et al., 2017) indica que esta complejidad acrecienta la superficie de ataque y dificulta la detección de anomalías por parte de los IDS tradicionales.

Otro factor que aumenta la problemática se refiere al gran volumen de datos y tráficos de red que manejan las diferentes instituciones de la salud como la Clínica Granados. Los

IDS tradicionales normalmente se ven colapsados por la magnitud de datos que deben procesar en tiempo real. Al generarse esta sobrecarga puede dar como resultado en una alta tasa de falsos positivos, identificando erróneamente actividades normales como amenazas, y de falsos negativos, en el que pasan por alto las amenazas reales. Además, en el sector salud es de vital importancia la capacidad de detectar y responder a las amenazas en tiempo real, debido a que la disponibilidad de los sistemas puede ser en varios casos crítico para la vida de los pacientes, debido a que manejan todo su historial médico. Sin embargo, los IDS tradicionales a menudo carecen de esta capacidad de respuesta inmediata (Aldawood & Skinner, n.d.).

A través del análisis y revisión del informe de (Cybersecurity Workforce Study, n.d.), se logra observar la escasez de profesionales especialidad en ciberseguridad contribuyendo así a la problemática actual, dando a conocer que existen más de 3 millones de puestos sin ser cubiertos correctamente en este ámbito. Es mucho mas notoria esta ausencia en el sector de la salud, presentando la dificultad de la gestión apropiada de los IDS tanto tradicionales en los cuales se requiere una supervisión constante. Además, las instituciones de salud deben cumplir con regulaciones estrictas como HIPAA en Estados Unidos o el RGPD en Europa. Se indica que los IDS tradicionales a menudo no están diseñados para abordar estos requisitos específicos del sector salud(Martin et al., 2017).

En esta circunstancia, se presenta la necesidad de investigar y explorar soluciones más avanzadas como los IDS basados en inteligencia artificial, los cuales puedan superar las limitaciones de los sistemas tradicionales. La IA tiene el potencial de mejorar significativamente la detección de amenazas, reducir los falsos positivos y negativos, y proporcionar una respuesta más rápida y precisa a los incidentes de seguridad (Buczak & Guven, 2016). Sin embargo, la implementación de IDS basados en IA en el sector salud plantea sus propios desafíos, incluyendo preocupaciones sobre la privacidad de los datos, la interpretabilidad de los modelos de IA y la integración con los sistemas existentes (Sahi et al., 2016).

Por lo consiguiente, es importante lograr evaluar diligentemente la eficacia y viabilidad de estos sistemas en el entorno específico como es el de la Clínica Granados antes de su implementación. Mediante esta investigación se plantea analizar esta problemática, evaluando y comparando adecuadamente la eficacia de un IDS basado en IA frente a los

sistemas tradicionales en el entorno de la institución. El objetivo es proporcionar insights valiosos para la toma de decisiones sobre la implementación de estas nuevas tecnologías avanzadas en el sector salud, apoyando así a mejorar la protección de la infraestructura crítica y los datos sensibles de los pacientes.

1.2 OBJETIVOS

1.2.1 OBJETIVO GENERAL:

Evaluar la eficacia de un sistema de detección de intrusiones (IDS) basado en inteligencia artificial (IA) para mejorar la protección de la infraestructura crítica en la Clínica Granados, Santa Elena, comparándolo con los sistemas tradicionales en términos de precisión, eficiencia y capacidad de respuesta ante amenazas cibernéticas.

1.2.2 OBJETIVOS ESPECÍFICOS:

1. Evaluar la eficacia de un IDS basado en IA en la detección de intrusiones dentro de la infraestructura crítica de la Clínica Granados.
2. Comparar el rendimiento del IDS basado en IA con los sistemas tradicionales de detección de intrusiones en términos de precisión y eficiencia.
3. Identificar las ventajas y limitaciones de la integración de IA en los sistemas de seguridad para entornos críticos de salud.

1.3 PREGUNTAS DE INVESTIGACIÓN

1. ¿Cómo se desempeña el IDS basado en IA en comparación con los sistemas tradicionales en la detección de intrusiones en la Clínica Granados?
2. ¿Qué mejoras en la detección de amenazas cibernéticas se observan al utilizar IA en el IDS?
3. ¿Cuáles son las principales ventajas y limitaciones de aplicar IA en la protección de infraestructura crítica en el sector salud?

CAPITULO II

2.1 REVISIÓN DE LITERATURA

2.1.1 SÍNTESIS DE ESTUDIOS PREVIOS

La revisión de la literatura indica que los IDS tradicionales, aunque efectivos en algunos contextos, tienen limitaciones en la detección de amenazas emergentes debido a su dependencia de firmas y reglas predefinidas. Investigaciones recientes han demostrado que la inteligencia artificial, particularmente el aprendizaje automático y el aprendizaje

profundo, ofrece mejoras significativas en la detección de patrones anómalos y en la adaptación a nuevas amenazas.

El concepto de Sistemas de Detección de Intrusiones tiene sus inicios en la década de 1980. (Lunt, 1993) señala que el término "detección de intrusiones" fue recalcado por James P. Anderson en su prestigioso artículo técnico de 1980 para la Fuerza Aérea de los Estados Unidos. Anderson propuso un modelo de auditoría de seguridad que sentó las bases apropiadas para los futuros IDS. Este trabajo inicial logro establecer los fundamentos conceptuales para el adecuado desarrollo de sistemas automatizados que serían capaces de identificar y responder a actividades maliciosas en entornos computacionales.

Durante el desarrollo de los primeros IDS que lograron ser operativos, se produjo entre las décadas de 1980 en sus finales y a principios de 1990. Uno de los primeros modelos de IDS en tiempo real fue presentado por (Denning, 1987), este lograba monitorear los registros de auditoría del sistema para así detectar las anomalías. El desarrollo de este trabajo fue un pilar fundamental para lograr establecer los principios básicos de la detección de intrusiones y a partir de ahí ser la base para las futuras investigaciones. El modelo de Denning incluyó la idea de perfiles de comportamiento normal y la detección de desviaciones de estos como indicadores de posibles intrusiones.

La evolución de los IDS puede dividirse en varias etapas significativas. La primera generación de IDS, que surgió en la década de 1990, se basaba principalmente en sistemas de detección de firmas. Estos sistemas utilizaban una base de datos de patrones de ataques conocidos para identificar amenazas. (Axelsson, 2000) en su artículo menciona estos sistemas, aunque efectivos contra ataques conocidos, eran inherentemente ineficaces contra nuevas formas de amenazas no registradas en su base de datos. Esta limitación llevó a la búsqueda de enfoques más flexibles y adaptables.

La segunda generación de los IDS surge entre los finales de la década de 1990 y principios de 2000, en el cual su característica principal estaba basada en anomalías. En el artículo científico explicado por (García-Teodoro et al., 2009), buscaban desviaciones del comportamiento normal del sistema para identificar posibles intrusiones. Este enfoque ofrecía la ventaja de poder detectar amenazas desconocidas, pero sufría de altas tasas de

falsos positivos debido a la dificultad de definir con precisión el comportamiento "normal" de un sistema.

En la década de los 2000 surge la necesidad de superar las limitaciones presentadas anteriormente de los enfoques basados en firmas y anomalías dando como resultado el desarrollo de IDS híbridos. Según indica (Liao et al., 2013), estos sistemas combinaban técnicas de detección basadas en firmas y anomalías para mejorar la precisión y reducir los falsos positivos. Los IDS híbridos representaron un avance significativo, pero aún enfrentaban desafíos en términos de adaptabilidad a la rápida evolución de las amenazas cibernéticas.

La inteligencia artificial en la integración de los IDS es el comienzo de la cuarta generación de estos sistemas, en lo que va de la década del 2010 hasta la actualidad. (Buczak & Guven, 2016) proporcionan una revisión exhaustiva de los métodos de minería de datos y aprendizaje automático aplicados a la detección de intrusiones. Su investigación destaca cómo los IDS basados en IA ofrecen ventajas significativas sobre los sistemas tradicionales, incluyendo una mejor detección de amenazas desconocidas, adaptabilidad a nuevos tipos de ataques, reducción de falsos positivos y la capacidad de manejar grandes volúmenes de datos en tiempo real.

Basándose en una síntesis de múltiples estudios científicos recientes, se han elaborado dos tablas comparativas que ilustran las diferencias clave entre los IDS tradicionales y los basados en IA. Esta información se deriva de un análisis exhaustivo de artículos publicados en revistas especializadas y conferencias de ciberseguridad entre 2018 y 2023.

A continuación, se presenta una comparación estadística del rendimiento en la ilustración 1:

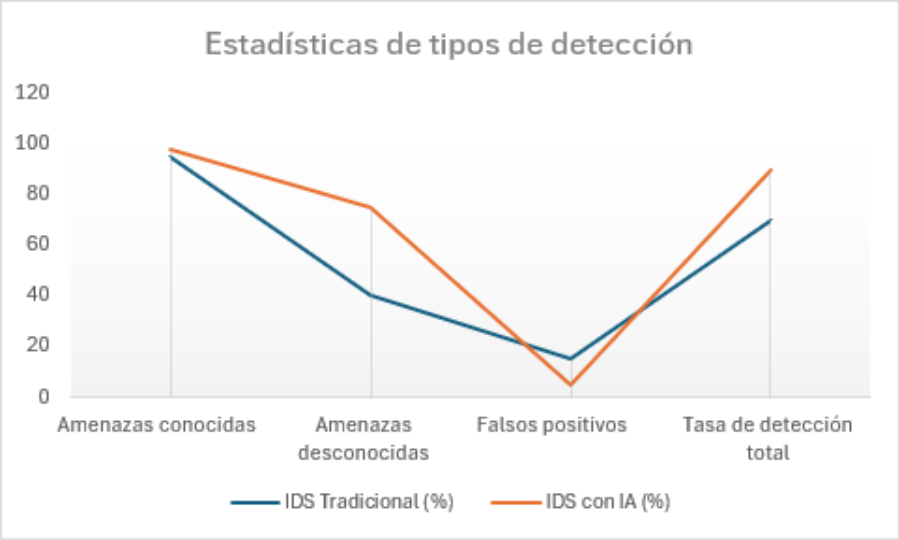

Ilustración 1: Estadísticas de tipos de detección

Este gráfico expone los descubrimientos de varios estudios realizados a lo largo del tiempo, como son las investigaciones de (Buczak & Guven, 2016; Khraisat et al., 2019; Latif et al., 2020). Los datos muestran que tanto los IDS tradicionales y los basados en IA tienen un rendimiento similar en la detección de amenazas conocidas, pero se ve una ventaja significativa de los sistemas con IA al reconocer amenazas desconocidas y su reducción de falsos positivos.

La tasa de detección de amenazas conocidas es alta para ambos sistemas (95% para IDS tradicionales y 98% para IDS con IA), lo que refleja la eficacia de ambos enfoques en escenarios familiares. Sin embargo, la diferencia más notable se observa en la detección de amenazas desconocidas, donde los IDS con IA muestran una tasa de éxito del 75% frente al 40% de los sistemas tradicionales. Esta mejora se atribuye a la capacidad de aprendizaje y adaptación de los algoritmos de IA. Además, la reducción en la tasa de falsos positivos (del 15% en sistemas tradicionales al 5% en sistemas con IA) es un avance significativo, ya que disminuye la carga de trabajo del personal de ciberseguridad y permite una respuesta más eficiente a las amenazas reales.

Considerando tanto las amenazas conocidas como las desconocidas, se presenta a través de una tasa de detección total, donde se visualiza una ventaja considerable para los sistemas basados en IA, obteniendo un 90% frente a un 70% de IDS tradicionales. El notable incremento del 20% en la efectividad total logra representar una mejora valiosa en la capacidad de protección de las infraestructuras consideras críticas.

Mediante estos datos estadísticos se pretende respaldar la propensión hacia la implementación de IDS basados en IA en los entornos considerados críticos como lo es el sector de la salud, donde una detección oportuna, rápida y precisa de las amenazas es importante para la protección de datos sensibles y la continuidad de los diferentes servicios médicos que ofertan diferentes instituciones de salud.

Para identificar mejor las diferencias entre los enfoques tradicionales y los basados en IA, se presenta a continuación la tabla 1 que contiene comparativa de diferentes técnicas de IDS, en el que se resume las ventajas y desventajas de cada enfoque.

ASPECTO	IDS TRADICIONAL		IDS CON IA	
	VENTAJA	DESVENTAJA	VENTAJA	DESVENTAJA
Detección de amenazas conocidas	Alta precisión para patrones conocidos	Limitado a amenazas predefinidas	*Alta precisión con capacidad de generalización * Puede detectar variaciones de amenazas conocidas	N/A
Detección de amenazas desconocidas	N/A	*Capacidad limitada * Requiere actualización manual de firmas	*Mejor Capacidad para detectar anomalías nuevas *Puede identificar patrones inusuales	N/A
Falsos positivos	N/A	*Tasa relativamente alta *Puede generar alertas innecesarias	Tasa reducida gracias al aprendizaje continuo	Aún puede generar algunos falsos positivos
Adaptabilidad	N/A	*Requiere actualizaciones manuales frecuentes *Puede quedar obsoleto rápidamente	*Se adapta automáticamente a nuevos patrones *Mejora continua con nuevos datos	N/A
Velocidad de procesamiento	Generalmente rápido para reglas simples	Puede ralentizarse con muchas reglas	Eficiente en el procesamiento de grandes volúmenes	Puede ser más lento debido a la complejidad del modelo
Recursos computacionales	*Requisitos generalmente menores *Adecuado para sistemas con recursos limitados	N/A	N/A	*Puede requerir más recursos, especialmente en entrenamiento *Necesita hardware más potente para un rendimiento óptimo
Mantenimiento	N/A	*Requiere actualización manual frecuente de reglas *Demanda constante de tiempo del personal de TI	Menor necesidad de intervención manual para actualizaciones	Requiere monitoreo del rendimiento del modelo
Interpretabilidad	*Reglas claras y fáciles de entender *Facilita la auditoría y el cumplimiento normativo	N/A	N/A	*Puede ser una "caja negra", difícil de interpretar *Desafíos en explicar decisiones específicas
Costo inicial	*Generalmente menor costo de implementación *Tecnología bien establecida y comprendida	N/A	Potencial para reducir costos a largo plazo	Mayor costo inicial debido a la complejidad del sistema
Escalabilidad	N/A	*Puede ser difícil escalar con el aumento del tráfico *Rendimiento puede degradarse con el crecimiento de la red	*Mejor capacidad para manejar grandes volúmenes de datos *Se adapta bien a entornos de red en expansión	N/A

Tabla 1: Ventajas y Desventajas IDS Tradicional vs IDS – IA

A través de la representación de esta tabla, se logra obtener una visión más detallada de las diferentes fortalezas y debilidades de cada enfoque. Se observa que los IDS

tradicionales se destacan por su simplicidad, reglas claras y menor costo inicial en su implementación, pero se ven afectados estos sistemas en términos de adaptabilidad y detección de amenazas desconocidas. Sin embargo, los IDS basados en IA proporcionan una gran adaptabilidad y eficacia para la detección de nuevas amenazas, pero con el desafió en términos de interpretabilidad y recursos computacionales.

En este ámbito enfocado en las instituciones de la salud como es el caso de la Clínica Granados, la comparación realizada es fundamental debido a que se posee la necesidad imperativa de proteger la información altamente sensible y de esta manera tratar de preservar la continuidad de los servicios críticos por lo que se requiere un equilibrio entre la alta precisión en la detección de amenazas conocidas y la capacidad de adaptarse rápidamente a nuevas formas de ataque. Los IDS basados en IA, especialmente combinan múltiples técnicas, permiten ofrecer el mejor ambiente entre estos requisitos, aunque durante su implementación debe tomar en cuenta meticulosamente los desafíos específicos del entorno sanitario.

La integración de IDS con otras tecnologías emergentes está dando forma a la próxima generación de sistemas de seguridad. En la investigación de (Yaqoob et al., 2017) discuten las tendencias futuras en IDS, en el que se incluyen el uso de técnicas de aprendizaje asociado para preservar la privacidad, la integración con tecnologías blockchain para mejorar la integridad de los datos, y el uso de técnicas de IA explicable para abordar problemas de interpretabilidad. Además, la adaptación de los IDS a entornos de computación en la nube y de borde (edge computing) está surgiendo como un área crítica de investigación y desarrollo.

Enfocado en el contexto del entorno médico, contar con la implementación de un IDS con IA presenta un desafío y una serie de oportunidades únicas. En la investigación (Latif et al., 2020) se propone un enfoque de detección de intrusiones basado en aprendizaje profundo para redes de Internet de las Cosas (IoT) de la mano con el cuidado de la salud. Su estudio, publicado actualmente en la revista Sensors, demuestra cómo los modelos de aprendizaje profundo pueden mejorar significativamente la precisión en la detección de anomalías en entornos de salud conectados. Los autores utilizaron un conjunto de datos de IoT médico para entrenar y evaluar diferentes modelos de aprendizaje profundo, logrando obtener una precisión de detección superior al 99% para ciertos tipos de ataques.

Este trabajo recalca que los sistemas basados en IA tienen la capacidad para adaptarse a los patrones de tráfico únicos en redes médicas y lograr detectar efectivamente amenazas tanto conocidas como desconocidas, una característica concluyente en el entorno dinámico de la atención médica actual.

Sin embargo, la adopción de IDS basados en IA en el sector salud no está exenta de desafíos. Se menciona que la privacidad de los datos, los requisitos de tiempo real y la interpretabilidad de los modelos son consideraciones críticas en la adopción de estos sistemas en entornos de atención médica (Iqtidar Newaz et al., 2019). Estos desafíos dan importante relevancia en la necesidad de un enfoque cuidadoso y adaptado al implementar IDS basados en IA en instituciones de salud como la Clínica Granados.

A través de la notable evolución de los IDS, desde sus inicios alrededor de los años 1980 hasta la actualizada con los IDS basados en IA, se logra observar el constante esfuerzo para mejorar la seguridad cibernética haciendo frente a las amenazas cada vez más sofisticadas. Cabe recalcar que los IDS tradicionales han sido un pilar fundamental para la seguridad de la información durante los últimos años, pero la integración de IA está expandiendo nuevas posibilidades para mejorar la detección y las respuestas de amenazas cibernéticas. En el entorno de la salud contar con la implementación de estos sistemas genera una promesa de mejora significativa en la protección de datos sensibles y su infraestructura.

La integración de la inteligencia artificial, particularmente el aprendizaje automático, ha revolucionado el desarrollo de Sistemas de Detección de Intrusiones (IDS). Los algoritmos de aprendizaje automático permiten crear modelos que simulan eficazmente los conceptos de IDS, utilizando grandes conjuntos de datos etiquetados para entrenar estos sistemas. Este enfoque supervisado permite a los modelos aprender patrones complejos de tráfico de red y comportamientos maliciosos, mejorando significativamente la capacidad de detección de amenazas tanto conocidas como desconocidas. Se emplean diversas técnicas de aprendizaje automático, como redes neuronales, árboles de decisión y máquinas de vectores de soporte, para construir modelos robustos y adaptables. El proceso de entrenamiento implica la revisión iterativa de datos supervisados, donde cada instancia está etiquetada como benigna o maliciosa, permitiendo al modelo aprender a clasificar nuevas instancias con precisión. "Los métodos de aprendizaje automático han

demostrado ser efectivos en la detección de intrusiones debido a su capacidad para manejar grandes volúmenes de datos y adaptarse a nuevos patrones de ataque" (Khraisat et al., 2019). Para evaluar el rendimiento del modelo, se utiliza comúnmente la matriz de confusión, implementada en programas de Python, que proporciona una visión detallada de la precisión, sensibilidad y especificidad del modelo en la detección de intrusiones.

2.1.2 IDENTIFICACIÓN DE BRECHAS

A pesar de los avances en la integración de IA en la ciberseguridad, existe una brecha en la aplicación específica de estas tecnologías en el entorno crítico de salud. La mayoría de los estudios se enfocan en sectores industriales o financieros, dejando poco explorado cómo estos sistemas pueden ser adaptados y optimizados para proteger infraestructuras críticas en entornos médicos.

Una de las brechas más notables es la poca elaboración de estudios enfocados sobre la efectividad de los IDS basados en IA en la detección de amenazas específicas del sector salud. Los ataques dirigidos a sistemas de historiales médicos electrónicos, dispositivos médicos conectados (IoMT) y sistemas de telemedicina presentan desafíos únicos que no han sido completamente abordados por la investigación actual.

Se observa otra brecha importante relacionada a la integración de los IDS durante la infraestructura de TI en las diferentes instituciones de la salud. El entorno médico normalmente cuenta con varios sistemas normalmente heredados y en ciertos casos tecnologías modernas, lo que da como origen a un planteamiento de desafíos único durante la implementación efectiva para las soluciones de seguridad avanzadas. Las investigaciones o literatura existente actual presentan falencias de estudios detallados sobre cómo lograr que los IDS con IA logren adaptarse a las infraestructuras heterogéneas sin comprometer la eficiencia operativa.

Durante el cumplimiento normativo y la privacidad representando otra área en la cual se visualiza una brecha importante. Los IDS tradicionales ya presentan tales desafíos, mientras que los basados en IA tienen otros desafíos como el procesamiento y almacenamiento de los datos sensibles de cada paciente. Se necesita más investigación sobre cómo estos sistemas pueden cumplir con regulaciones estrictas como HIPAA o GDPR sin comprometer su eficacia. El rendimiento que se logre obtener y la escalabilidad de estos en el entorno de la salud a gran escala es un área poco analizada, debido a que

las instituciones relacionadas al ámbito de la salud generan volúmenes masivos de datos en tiempo real, y he ahí la notable brecha entre la comprensión de como estos sistemas pueden manejar sin complicaciones este flujo de información sin generar una latencia o comprometer la precisión de la detección de amenazas.

Finalmente, existe una brecha en la investigación sobre la interpretabilidad y la capacidad de explicación de los IDS basados en IA en contextos médicos. La toma de decisiones en entornos de salud a menudo requiere una comprensión clara de los factores que influyen en una alerta de seguridad, y los modelos de "caja negra" típicos de estos sistemas presentan desafíos en este aspecto. Para tener una mejor percepción de las brechas y diferencias de estos se visualiza la tabla 2 con esta comparativa:

Aspecto	IDS Tradicionales	IDS con IA	Brecha identificada
Detección de amenazas específicas del sector salud	Limitada a amenazas conocidas	Potencial para detectar nuevas amenazas	Falta de estudios sobre efectividad en amenazas específicas de salud
Integración con infraestructura de TI médica	Dificultades con sistemas heredados	Mayor adaptabilidad potencial	Escasez de investigación sobre integración en entornos heterogéneos
Cumplimiento normativo y privacidad	Desafíos existentes	Nuevas preocupaciones por procesamiento de datos	Necesidad de más investigación sobre cumplimiento de HIPAA/GDPR
Escalabilidad en entornos de salud	Limitaciones con grandes volúmenes de datos	Potencial para manejar big data	Falta de estudios sobre rendimiento en escala real
Interpretabilidad de alertas	Generalmente más transparente	Posibles problemas de "caja negra"	Brecha en investigación sobre explicabilidad en contextos médicos
Adaptación a nuevas amenazas	Requiere actualizaciones manuales	Aprendizaje y adaptación potencialmente autónomos	Necesidad de validación en entornos de salud reales

Tabla 2: Brechas - Ids Tradicional / Ids con IA

Desde un análisis y perspectiva técnica, se logra evidenciar que su principal brecha se basa en la capacidad de los IDS para poder manejar el gran tráfico de red que se emplea en los entornos de la salud. Se manejan protocolos específicos para el intercambio de información clínica como el HL7 (Health Level 7), o para imágenes médicas conocido como DICOM (Digital Imaging and Communications in Medicine), y una serie de diferentes protocolos de pertenencia empleados en los diferentes dispositivos de la salud. Normalmente los IDS tradicionales se establecen en firmas y reglas predefinidas, pero presentan dificultades para interpretar y analizar eficazmente estos protocolos especializados. En otro contexto, los IDS con IA obtienen el potencial de poder adaptarse y aprender de estos protocolos, del mismo modo aún persiste la brecha de la investigación

sobre cómo preparar y entrenar apropiadamente estos sistemas consiguiendo reconocer patrones de amenazas en datos tan específicos del dominio médico.

Partiendo desde la perspectiva del procesamiento de datos, el entorno amplio en el que se encuentran las instituciones de la salud genera grandes volúmenes de información sensible en tiempo real de los diferentes pacientes que atienden. Teniendo en cuenta que los IDS tradicionales normalmente luchan con el análisis de estos datos en tiempo real, esto podría generar retrasos en la detección de las amenazas o en el peor de los casos de solo presentar el análisis en un parte del tráfico de red. Por otro lado, los IDS basados en IA, están enfocados especialmente en utilizar técnicas de aprendizaje profundo como lo son las redes neuronales recurrentes (RNN) o las redes neuronales convolucionales (CNN), que tienen como objetivo procesar y analizar gran cantidad de datos que se manejan en el tiempo real. Sin embargo, existe una brecha en la investigación sobre cómo optimizar estos sistemas para el análisis de seguridad en tiempo real en el contexto específico de los datos de salud, considerando las limitaciones de latencia críticas en entornos médicos donde cada segundo puede ser decisiva para la atención del paciente.

2.1.3 POSICIONAMIENTO DEL ESTUDIO

Este estudio aborda la brecha en la literatura al aplicar y evaluar un IDS basado en IA en un entorno real de salud, específicamente en la Clínica Granados. El análisis proporcionará una visión sobre la eficacia de las soluciones de IA en un contexto crítico y contribuirá al conocimiento existente en la ciberseguridad en el sector salud.

La Clínica Granados, se encuentra ubicada en la ciudad de Salinas, Provincia de Santa Elena, Ecuador, cuenta con una infraestructura de TIC's moderna y compleja, diseñada para soportar las operaciones de atención médica y la gestión de datos sensibles de los pacientes. El entorno tecnológico de la clínica está caracteriza por una combinación de sistemas heredados y tecnologías de vanguardia, lo que presenta desafíos únicos para la implementación de soluciones de seguridad avanzadas.

La topología de red de la Clínica Granados como se observa en la ilustración 4 sigue un diseño jerárquico de tres capas:

- **Capa de acceso:** Incluye switches de acceso que conectan dispositivos de usuario final, estaciones de trabajo médicas, y dispositivos IoMT (Internet of Medical Things).
- **Capa de distribución:** Está compuesta por switches de alto rendimiento que agregan el tráfico de la capa de acceso y aplican políticas de seguridad.
- **Capa de núcleo:** Está compuesta por routers y switches de núcleo que proporcionan conectividad de alta velocidad entre los diferentes segmentos de la red y hacia el centro de datos.

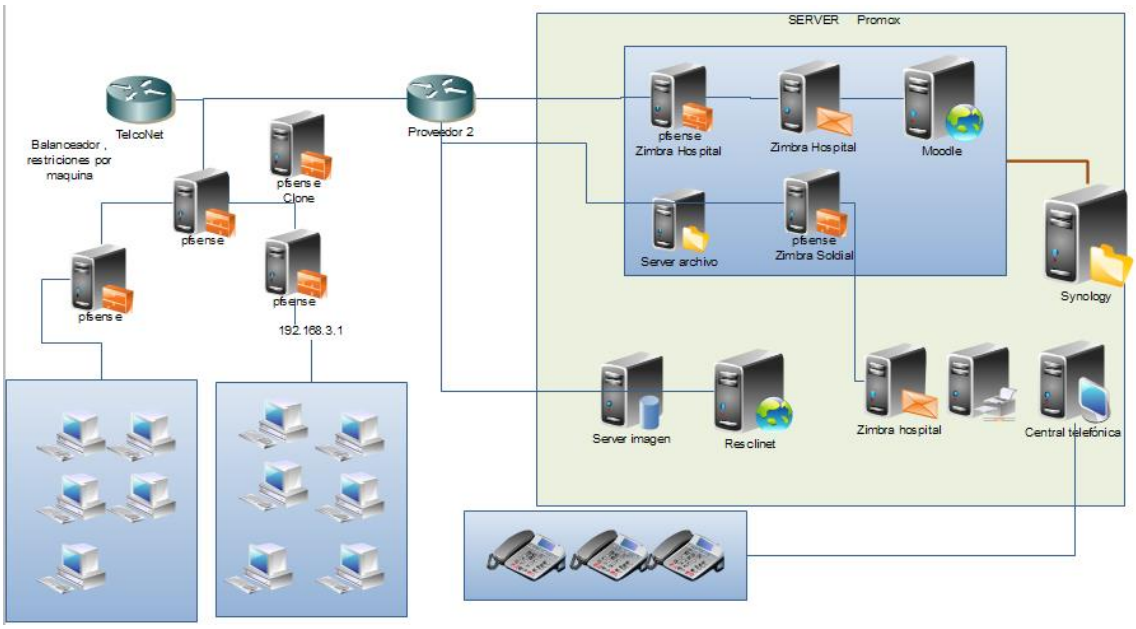

Ilustración 2: Topología

La red está segmentada lógicamente utilizando VLANs para separar el tráfico entre diferentes departamentos y tipos de dispositivos (por ejemplo, VLAN para sistemas administrativos, VLAN para dispositivos médicos, VLAN para sistemas de imágenes médicas).

El centro de datos de la clínica aloja servidores físicos y una infraestructura virtualizada basada en Proxmox, que ejecuta aplicaciones críticas, incluyendo:

- Sistema de Información Hospitalaria (HIS)
- Sistema de Archivo y Comunicación de Imágenes (PACS)
- Sistema de Gestión de Laboratorio (LIS)

- Sistemas de telemedicina
- Servidores de bases de datos que almacenan registros médicos electrónicos (EMR)
- Sistemas especializados como Zimbra Hospital y Central Telefónica

La conectividad externa se gestiona mediante un sistema de firewall y balanceador de carga establecido en pfSense, que proporciona inspección profunda de paquetes y prevención de intrusiones basada en firmas. La clínica utiliza una combinación de enlaces de internet redundantes de diferentes proveedores (TelcoNet y un proveedor secundario) para garantizar la continuidad del servicio.

Para la gestión de dispositivos médicos conectados (IoMT), tiene implementado una red aislada con su propio firewall y gateway, que se conecta al resto de la infraestructura a través de interfaces estrictamente controladas. El sistema de almacenamiento Synology juega un papel crucial en la gestión y protección de datos sensibles de pacientes, lo que añade una capa adicional de complejidad a la estrategia de seguridad.

Para el sistema de monitoreo de red de la infraestructura actual utiliza herramientas tradicionales como Nagios y un SIEM (Security Information and Event Management) para la analogía de incidentes de seguridad. Sin embargo, estos sistemas han mostrado limitaciones en la detección de amenazas avanzadas y en el manejo del volumen de datos generados por los dispositivos médicos actuales.

La implementación del IDS basado en IA se realizará en paralelo con los sistemas existentes, logrando obtener una comparación directa de su rendimiento. Se desplegará un sensor de red en el punto de agregación principal, con la capacidad de poder analizar todo el tráfico que fluye entre los diferentes segmentos de la red y hacia/desde internet.

Este entorno complejo y crítico proporciona un escenario ideal para evaluar la eficacia de un IDS basado en IA, ya que presenta desafíos únicos como:

- La necesidad de analizar protocolos específicos del sector salud (HL7, DICOM, etc.).
- La gestión de un volumen de gran cantidad de datos sensibles en tiempo real.
- La necesidad de mantener un alto rendimiento y baja latencia, críticos para las actividades diaria en la institución.

- El monitoreo de tráfico entre segmentos en la infraestructura.
- La adaptación a una topología de red con múltiples proveedores de internet.

Mediante este estudio se considerará la integración hipotética de un IDS basado en IA para la infraestructura actual, logrando analizar el potencial para ejecutarse en paralelo con los sistemas existentes. De esta manera de podrá evaluar los requisitos técnicos, ventajas y desventajas, beneficios deseados y los posibles retos al momento de su implementación, tomando en consideración la compleja segmentación de la red y la presencia de sistemas especializados del sector salud.

Los resultados de este estudio de factibilidad proporcionarán insights valiosos acerca del potencial de los IDS basados en IA para mejorar la seguridad en infraestructuras críticas de salud. Abordando así directamente las brechas previamente identificadas en este trabajo acorde al sector de la salud, ofreciendo una base sólida para la toma de decisiones futuras sobre la adopción de estas tecnologías avanzadas en la Clínica Granados y en entornos similares.

2.2 METODOLOGÍA

2.2.1 DISEÑO DE INVESTIGACIÓN

El estudio utiliza un diseño experimental para evaluar el rendimiento de un IDS basado en IA. Se adoptó un enfoque cuantitativo para medir la eficacia del sistema en la detección de intrusiones y una metodología comparativa para contrastar los resultados con los sistemas tradicionales.

El diseño cuasi experimental se estructura en tres fases principales: evaluación baseline, implementación simulada, y análisis comparativo. En la fase de evaluación baseline, se recopilan y analizan datos históricos del rendimiento del IDS tradicional, topología actual de la Clínica Granados, estableciendo métricas clave como tasas de detección, falsos positivos, y tiempo de respuesta. La fase de implementación simulada implica la creación de un modelo de la infraestructura de red de la clínica, donde se despliega una versión simulada de un IDS basado en IA. Esta simulación se somete a una serie de escenarios de ataque cuidadosamente diseñados, que reflejan tanto amenazas conocidas como emergentes en el sector sanitario.

Finalmente, la fase de análisis comparativo contrasta el rendimiento del IDS tradicional con el basado en IA, utilizando métodos estadísticos rigurosos para evaluar la eficacia, eficiencia y adaptabilidad de ambos sistemas. Este enfoque permite una evaluación objetiva y cuantificable de las potenciales mejoras que un IDS basado en IA podría aportar a la seguridad cibernética de la Clínica Granados.

2.2.2 ENFOQUE METODOLÓGICO

- Análisis Cuantitativo: Se utilizarán métodos estadísticos para medir y comparar la eficacia de los sistemas de detección de intrusiones, tanto tradicionales como basados en IA.
- Metodología Comparativa: Se contrastará el rendimiento del IDS tradicional actual con un modelo simulado de IDS basado en IA.
- Análisis Documental: Se realizará una revisión exhaustiva de la documentación técnica, informes de seguridad y registros de incidentes de la Clínica Granados.

2.2.3 FASES DEL ESTUDIO

- Evaluación de la Infraestructura Actual:

  Análisis detallado de la topología de red de la Clínica Granados.

  Revisión de las capacidades y limitaciones del IDS tradicional en uso.

  Identificación de puntos críticos y vulnerabilidades en la infraestructura existente.

- Simulación del IDS basado en IA:
  Diseño de un modelo simulado de IDS con IA adaptado a las necesidades específicas de la Clínica.
- Análisis Comparativo:

  Evaluación del rendimiento del IDS tradicional vs. IDS con IA en términos de: Tasas de detección de intrusiones, Falsos positivos y falsos negativos, Tiempo de respuesta ante amenazas, Capacidad de adaptación a nuevas amenazas.

- Proyección de Implementación:

  Análisis de la viabilidad técnica y operativa de implementar un IDS basado en IA.

  Evaluación del impacto potencial en la seguridad y operaciones de la Clínica.

CAPITULO III

3.1 POBLACIÓN

La población del estudio incluye la infraestructura crítica de red de la Clínica Granados. La muestra abarca diversos segmentos de la red, incluidos sistemas de administración de pacientes, servidores de datos médicos y dispositivos conectados. Esta infraestructura representa un microcosmos de los sistemas de salud modernos, reflejando la complejidad y diversidad de las redes informáticas en el sector sanitario.

Para darle la importancia de este estudio en un entorno más extenso, es primordial examinar el panorama en general de la infraestructura de la salud donde se encuentra ubicada la Clínica Granados. Este se sitúa en la Provincia de Santa Elena, ubicada en la costa ecuatoriana, contando con una infraestructura de salud diversa que incluye establecimientos tanto públicos y privados. Según el Ministerio de Salud Pública del Ecuador (MSP), la red de salud en la provincia está dividida en sus tres cantones principales como son: Santa Elena, La Libertad y Salinas. Esta distribución geográfica de servicios de salud recalca la importancia de contar con sistemas de seguridad robustos y adaptables a diferentes escalas y contextos dentro de las diferentes instituciones de salud de la provincia.

El Geoportal del Ministerio de Salud Pública ofrece una visión general de la distribución de establecimientos de salud en la provincia: "La provincia de Santa Elena cuenta con un total de 39 establecimientos de salud registrados, distribuidos de la siguiente manera: 22 en el cantón Santa Elena, 9 en La Libertad y 8 en Salinas" (*Ministerio de Salud Pública Del Ecuador - GeoSalud3 | MSP*, n.d.). Es importante acotar que esta cifra incluye tanto instituciones públicas como privadas, abarcando los diversos niveles de atención médica.

Para obtener una perspectiva más extensa concerniente a la infraestructura de salud en Ecuador, se puede hacer referencia al Instituto Nacional de Estadística y Censos (INEC). En su informe "Registro Estadístico de Recursos y Actividades de Salud - RAS 2019", se proporciona información detallada sobre los recursos de salud a nivel nacional y provincial: "A nivel nacional, en 2019 se registraron 4.165 establecimientos de salud, de los cuales el 80,0% (3.334) corresponden al sector público y el 20,0% (831) al sector privado"(Instituto Nacional de Estadística y Censos, 2020).

Específicamente para la provincia de Santa Elena, el mismo informe detalla: "En la provincia de Santa Elena, se registraron 3 hospitales generales y 34 centros de salud en 2019"(Instituto Nacional de Estadística y Censos, 2020). Aunque en este informe no proporciona el número exacto de clínicas privadas en Santa Elena, ofrece una visión general de la infraestructura de salud en la provincia en un nivel muy general.

En cuanto a la potencial implementación de un Sistema de Detección de Intrusiones (IDS) basado en Inteligencia Artificial (IA), es importante considerar que su viabilidad no depende exclusivamente del tamaño de la institución. Factores como la criticidad de los datos manejados, el volumen de información procesada, y los recursos disponibles para inversión en seguridad informática son igualmente relevantes.

La Agencia de Regulación y Control de las Telecomunicaciones (ARCOTEL) del Ecuador enfatiza la importancia de la ciberseguridad en el sector salud: "La protección de la información en el sector salud es crucial, dado el carácter sensible de los datos personales y médicos manejados. La implementación de sistemas de seguridad avanzados, como los IDS basados en IA, puede proporcionar una capa adicional de protección contra amenazas cibernéticas emergentes" (Agencia de Regulación y Control de las Telecomunicaciones, 2022).

Para determinar con precisión cuántas clínicas en Santa Elena podrían beneficiarse de la implementación de un IDS con IA similar al propuesto para la Clínica Granados, sería necesario realizar un estudio específico que evalúe la infraestructura tecnológica y las necesidades de seguridad de cada institución en la provincia.

3.2 INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Para evaluar la eficacia de un IDS basado en IA en la detección de intrusiones dentro de la infraestructura crítica de la Clínica Granados, se utilizó como principal instrumento de recolección de datos una encuesta estructurada dirigida al personal del departamento de TIC's.

La encuesta tuvo como finalidad lograr obtener información referente a los sistemas de seguridad existentes, incluyendo la detección de spam, IDS, y otras medidas de ciberseguridad, con un enfoque particular en la percepción y experiencia del personal con

sistemas basados en IA. El cuestionario constaba de 14 preguntas que abarcaban desde aspectos básicos hasta temas más técnicos y específicos de la seguridad informática en el entorno de la institución.

Se aplicó la encuesta a 12 miembros del personal de TIC's de la Clínica Granados, obteniendo una variedad de respuestas que reflejan la diversidad de conocimientos y percepciones dentro del departamento. A continuación, se presenta un resumen de los resultados para evaluar la eficacia de un IDS basado en IA:

Pregunta 1: "¿Utiliza su institución algún sistema de seguridad informática?"

- Sí: 10 (83.3%)
- No: 0 (0%)
- No estoy seguro: 2 (16.7%)

Análisis: Se observa que más de la mitad de los encuestados confirma la existencia de sistemas de seguridad informática, lo que indica una conciencia general sobre la importancia de la ciberseguridad.

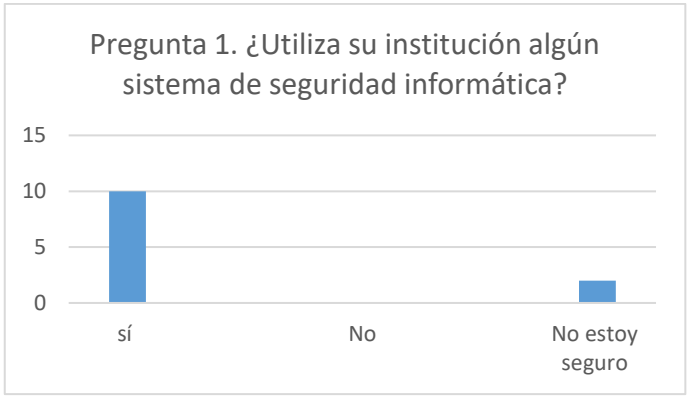

Ilustración 3: Pregunta 1

Pregunta 2: "¿Cuenta su institución con un sistema de detección de spam en el correo electrónico?"

- Sí: 8 (66.7%)
- No: 3 (25%)
- No estoy seguro: 1 (8.3%)

Análisis: Dos tercios de los encuestados confirman la presencia de sistemas de detección de spam, sugiriendo una protección básica contra amenazas por correo electrónico.

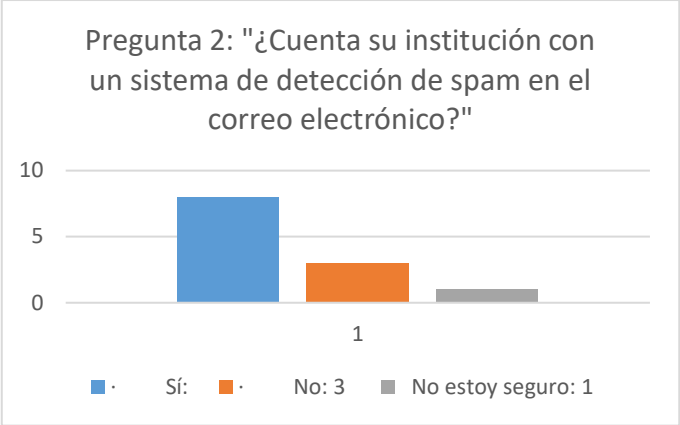

Ilustración 4: Pregunta 2

Pregunta 3: "¿Qué tipo de solución antivirus utiliza su institución?"

- Antivirus local en cada computadora: 4 (34%)
- Solución antivirus centralizada: 7 (58%)
- No utilizamos antivirus: 0 (0%)
- No estoy seguro: 1 (8%)

Análisis: Existe una preferencia por soluciones antivirus centralizadas, lo que sugiere un enfoque más gestionado y potencialmente más eficiente en la protección contra malware.

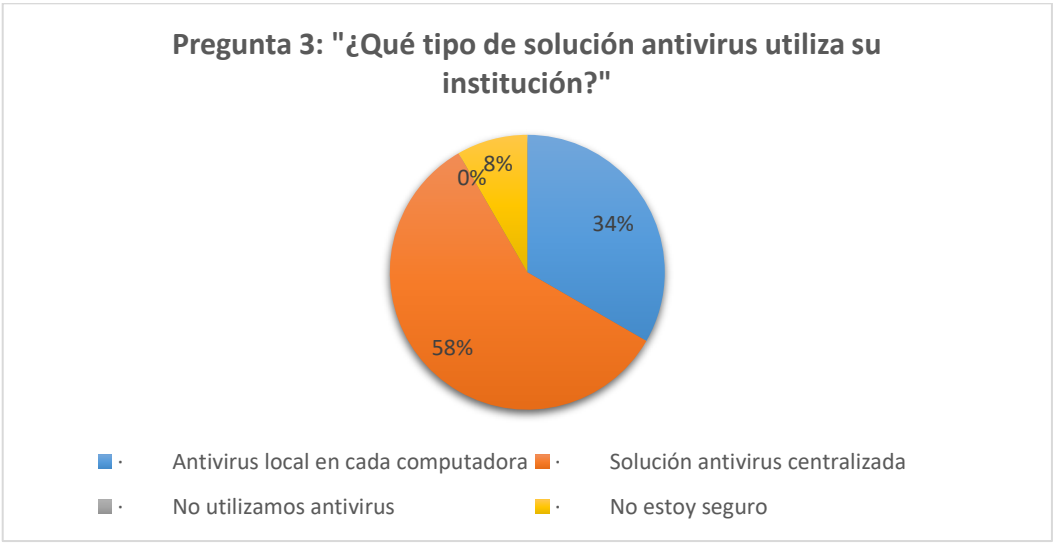

Ilustración 5: Pregunta 3

Pregunta 4: "¿Su institución utiliza un firewall?"

- Sí, hardware: 5 (41.7%)
- Sí, software: 4 (33.3%)

- Sí, tanto hardware como software: 0 (0%)
- No: 0 (0%)
- No estoy seguro: 3 (25%)

Análisis: La mayoría de los encuestados confirma el uso de firewalls, con una ligera preferencia por soluciones de hardware sobre software.

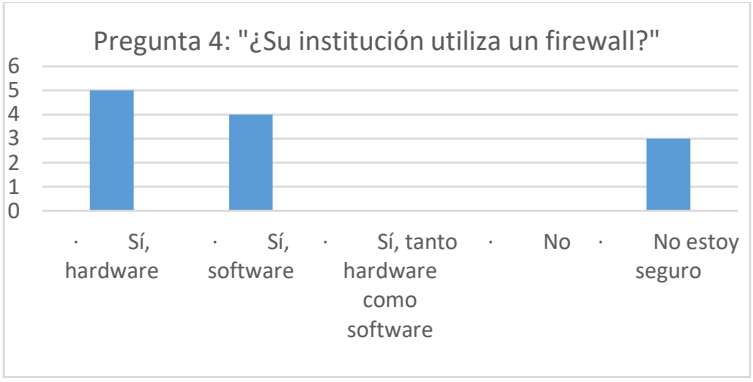

Ilustración 6: Pregunta 4

Pregunta 5: "¿Está familiarizado con el término 'Sistema de Detección de Intrusiones' (IDS)?"

- Sí: 9 (75%)
- No: 3 (25%)

Análisis: Tres cuartos de los encuestados están familiarizados con el concepto de IDS, lo que indica un buen nivel de conocimiento técnico en el departamento.

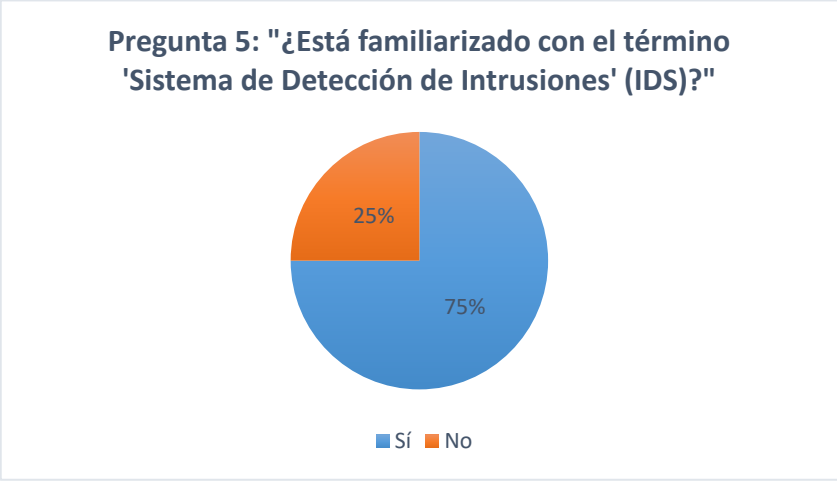

Ilustración 7: Pregunta 5

Pregunta 6: "¿Su institución tiene implementado algún Sistema de Detección de Intrusiones (IDS)?"

- Sí: 6 (50%)
- No: 4 (33.3%)
- No estoy seguro: 2 (16.7%)

Análisis: La mitad de los encuestados confirma la presencia de un IDS, indicando una adopción significativa pero no universal de esta tecnología.

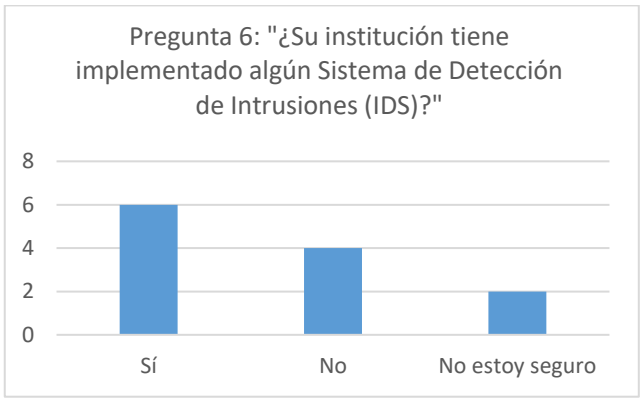

Ilustración 8: Pregunta 6

Pregunta 7: "Si la respuesta anterior es afirmativa, ¿qué tipo de IDS utilizan?"

- Basado en red (NIDS): 3 (50% de los que tienen IDS)
- Basado en host (HIDS): 2 (33% de los que tienen IDS)
- Ambos: 1 (17% de los que tienen IDS)

Análisis: Entre quienes utilizan IDS, hay una preferencia por los sistemas basados en red, seguidos de cerca por los basados en host.

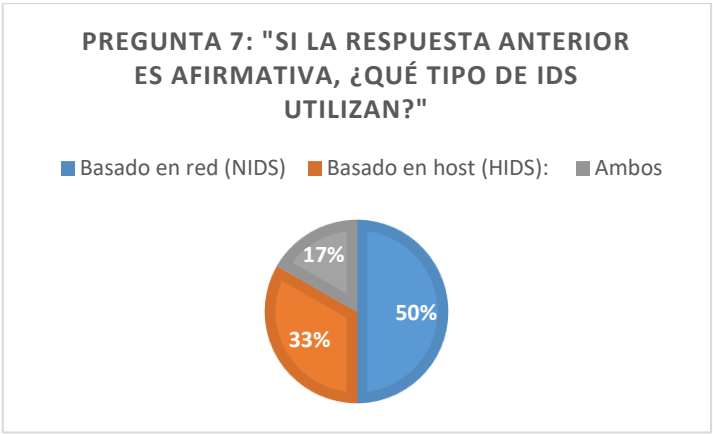

Ilustración 9: Pregunta 7

Pregunta 8: "¿Su institución utiliza algún Sistema de Prevención de Intrusiones (IPS)?"

- Sí: 4 (33.3%)
- No: 5 (41.7%)
- No estoy seguro: 3 (25%)

Análisis: Hay una distribución bastante equilibrada en el uso de IPS, con una ligera mayoría que no lo utiliza o no está segura.

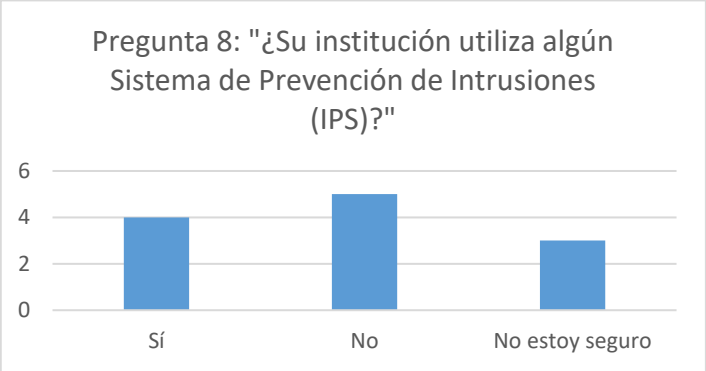

Ilustración 10: Pregunta 8

Pregunta 9: "¿Qué herramientas o soluciones específicas de seguridad utiliza su institución?"

- SIEM: 3 (25%)
- EDR: 2 (16.7%)
- WAF: 4 (33.3%)
- DLP: 1 (8.3%)
- Otra: 1 (8.3%)
- No estoy seguro: 4 (33.3%)

Análisis: Existe una variedad de herramientas en uso, con WAF siendo la más común. Un numero notable de los encuestados no está seguro, lo que podría indicar una falta de comunicación o conocimiento sobre las herramientas implementadas.

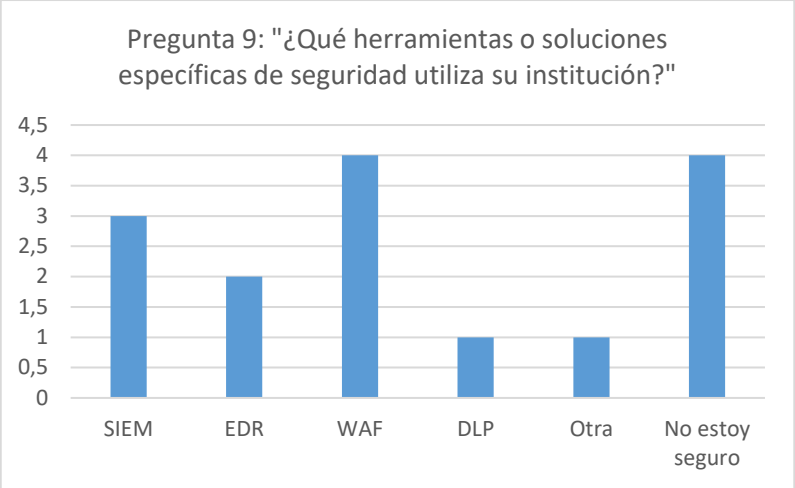

Ilustración 11: Pregunta 9

Pregunta 10: "¿Su institución utiliza alguna solución de seguridad basada en Inteligencia Artificial o Machine Learning?"

- Sí: 0 (0%)
- No: 6 (46%)
- No estoy seguro: 7 (54%)

Análisis: La mitad de los encuestados no utiliza soluciones basadas en IA, lo que sugiere un amplio margen para la adopción de tecnologías más avanzadas.

Como observamos en las respuestas acorde a la pregunta 10, no da continuidad a la pregunta 11, teniendo un casillero con 0%.

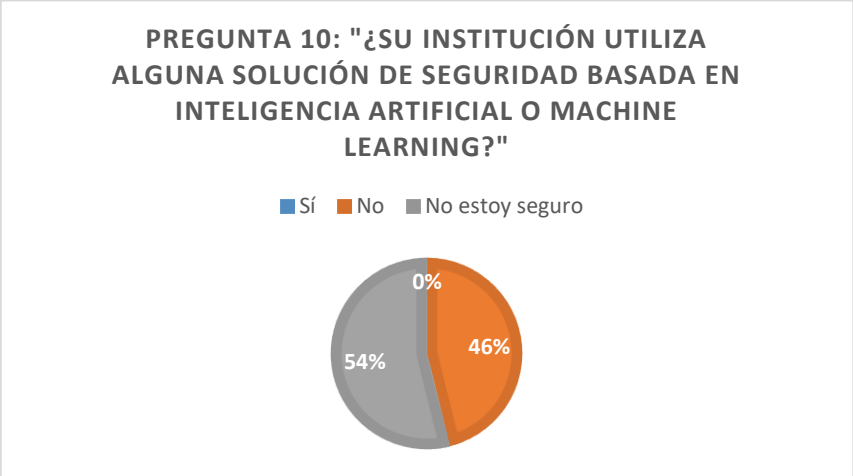

Ilustración 12: Pregunta 10

Pregunta 12: "¿Con qué frecuencia se actualizan los sistemas de seguridad en su institución?"

- Diariamente: 0 (0%)
- Semanalmente: 5 (41.7%)
- Mensualmente: 4 (33.3%)
- No estoy seguro: 3 (25%)

Análisis: La mayoría de las instituciones realizan actualizaciones regulares, con una ligera preferencia por actualizaciones semanales.

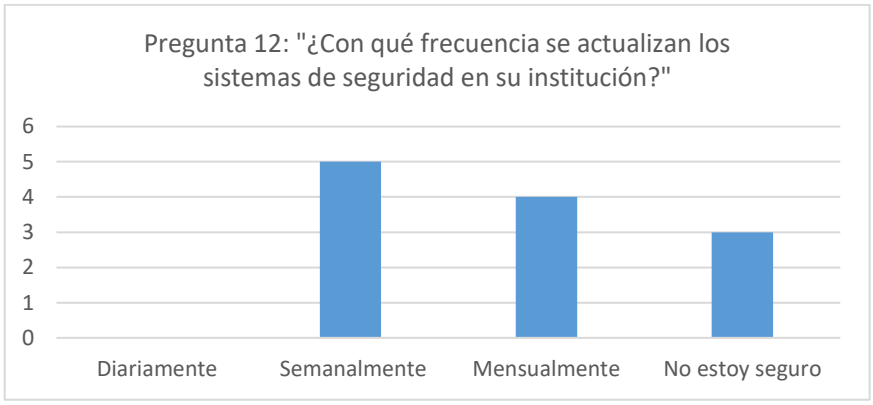

Ilustración 13: Pregunta 12

Pregunta 13: "¿Su institución realiza análisis de vulnerabilidades o pruebas de penetración?"

- Sí, regularmente: 3 (25%)
- Sí, ocasionalmente: 2 (16.7%)
- No: 4 (33.3%)
- No estoy seguro: 3 (25%)

Análisis: Los resultados muestran una distribución variada en las prácticas de análisis de vulnerabilidades.

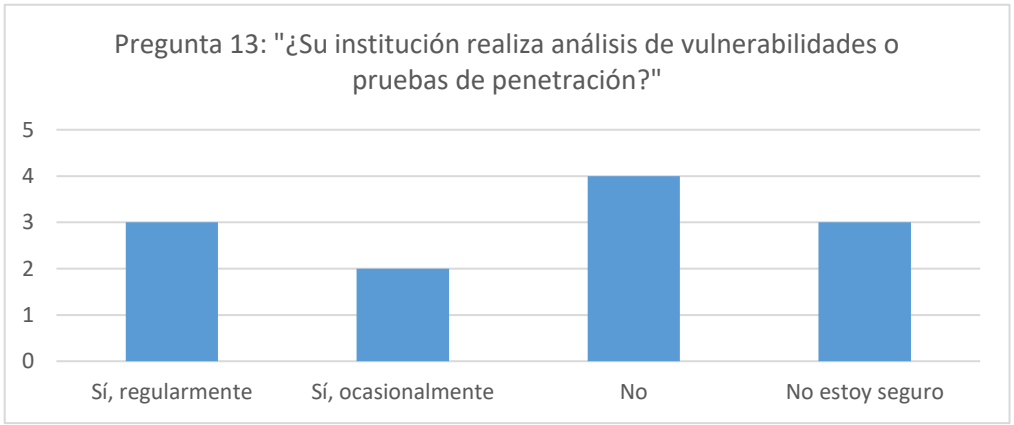

Ilustración 14: Pregunta 13

Pregunta 14: "¿Estaría su institución interesada en implementar soluciones de seguridad más avanzadas basadas en IA?"

- Sí: 9 (75%)
- No: 0 (0%)
- Tal vez: 3 (16.7%)
- Ya las utilizamos: 0 (0%)

Análisis: Existe un interés abrumador en la implementación de soluciones de seguridad basadas en IA, con tres cuartos de los encuestados expresando una disposición clara.

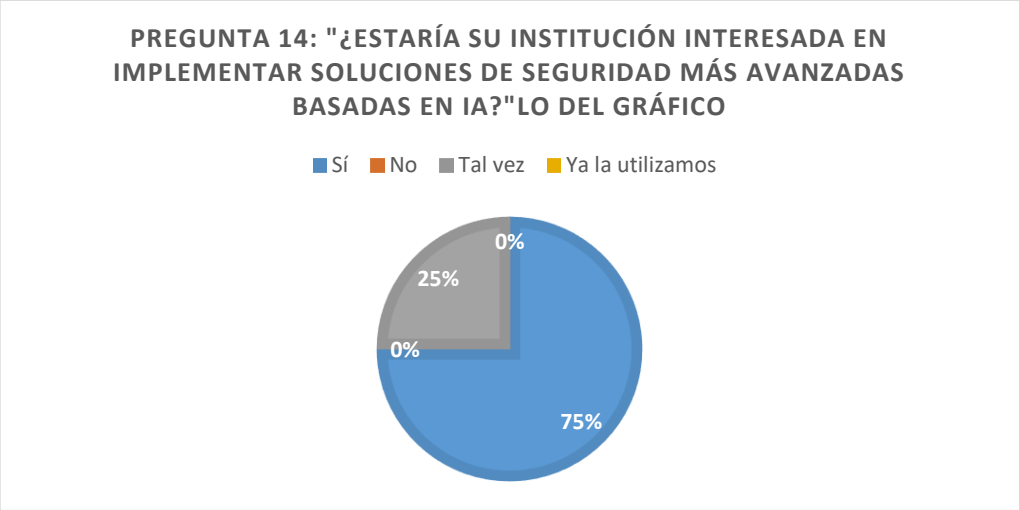

Ilustración 15: Pregunta 14

Mediante en análisis de estos resultados nos proporcionan una visión general del estado actual de la seguridad informática en la Clínica Granados y desde la percepción del personal de TIC's haciendo referencia a la eficacia de los sistemas existentes. La información recolectada permite sugerir que se genera un interés significativo en implementar sistemas avanzados como lo sería un IDS basado en IA, para poder abordar los desafíos actuales correspondientes a la detección de amenazas avanzadas y protección de la infraestructura.

Al momento se la elaboración de la encuesta se revela que, existe cierta implementación de un IDS, se genera una percepción extendida de que los sistemas actuales podrían tener una mejora si se combinara la estructura actual con un IDS basado en IA. Proporcionando un gran interés en estas nuevas soluciones, si se realiza la sugerencia de la implementación como tal podría tener una tasa de aceptación elevada y potencialmente

eficaz para mejorar la detección de intrusiones en la infraestructura crítica de la Clínica Granados. Esta información sirve como base para evaluar la viabilidad y necesidad de implementación de dicho sistema.

3.3 PROCEDIMIENTO

El estudio se llevará a cabo siguiendo un procedimiento meticuloso dividido en varias fases, utilizando un programa de análisis personalizado desarrollado específicamente para esta investigación:

Fase de Preparación:

- Obtención de permisos y consentimientos éticos de la Clínica Granados.
- Recopilación de documentación sobre la infraestructura de red existente.
- Identificación de los puntos críticos en la red donde se implementará el IDS.
- Selección y adquisición del software de IDS basado en IA para la prueba.
- Configuración del entorno para ejecutar el programa "Ev de resultados.ipynb", desarrollado específicamente para este estudio.

Fase de Evaluación Inicial:

- Análisis del rendimiento actual del IDS tradicional: Recopilación de logs y datos históricos.
- Cálculo de métricas base: tasa de detección, falsos positivos, falsos negativos.
- Realización de una auditoría de seguridad inicial para identificar vulnerabilidades existentes.
- Ingreso de los datos iniciales en el programa "Ev de resultados.ipynb" para establecer la línea base.

Fase de Implementación del IDS basado en IA:

- Ejecución de algoritmo de IDS basado en IA en un entorno de prueba que refleje la red de producción.
- Configuración inicial del IDS basado en IA según las necesidades específicas de la Clínica Granados.

Fase de Pruebas:

- Diseño de escenarios de prueba que incluyan: Ataques conocidos (basados en firmas), Ataques de día cero simulados, Anomalías de tráfico ,Intentos de intrusión específicos del sector salud.
- Ejecución de las pruebas en un entorno controlado.
- Recopilación de datos de rendimiento del IDS basado en IA.
- Ingreso de los datos recopilados en el programa "Ev de resultados.ipynb" para su análisis.

Fase de Análisis de Resultados:

- Ejecución del programa "Ev de resultados.ipynb":

  Carga de los datos recopilados durante las fases de evaluación inicial y pruebas.

  Ejecución de los análisis estadísticos personalizados implementados en el programa.

  Generación de gráficos comparativos entre el IDS tradicional y el basado en IA.

- Interpretación de los resultados generados por el programa:

  Análisis de las tasas de detección para diferentes tipos de amenazas.

  Evaluación de la reducción en falsos positivos y negativos.

  Examen de los tiempos de respuesta y eficiencia en la detección.

- Realización de análisis adicionales según sea necesario, utilizando las capacidades del programa personalizado.

Fase de Validación:

- Revisión de los resultados.
- Ajustes en base del IDS con IA según los hallazgos del análisis.

Fase de Documentación y Reporte:

- Compilación de todos los resultados y análisis generados por el programa "Ev de resultados.ipynb".
- Metodología utilizada, destacando el uso del programa personalizado
- Resultados del análisis comparativo
- Gráficos y visualizaciones generadas por el programa

- Interpretaciones y conclusiones
- Preparación de recomendaciones basadas en los hallazgos para la implementación a gran escala del IDS basado en IA.

Fase de Presentación:

- Preparación de una presentación visual utilizando los gráficos y resultados generados por el programa personalizado.
- Presentación de los hallazgos.
- Discusión de los próximos pasos y posible implementación a escala completa.

Este procedimiento detallado, que incorpora el uso del programa "Ev de resultados.ipynb" desarrollado específicamente para esta investigación, asegura un enfoque riguroso y basado en datos para evaluar la eficacia del IDS basado en IA en comparación con los sistemas tradicionales. El uso de esta herramienta de análisis personalizada permitirá una evaluación objetiva y cuantificable de los resultados, proporcionando una base sólida para la toma de decisiones sobre la implementación de tecnologías de seguridad avanzadas en la Clínica Granados.

3.4 ANÁLISIS DE DATOS

El análisis de datos incluyó la comparación de las tasas de detección y falsos positivos entre el IDS basado en IA y los sistemas tradicionales. Mediante el uso del programa se ejecutó para evaluar la precisión, la eficiencia y la capacidad de respuesta del IDS en diferentes escenarios de ataque. Para contextualizar este análisis, es importante visualizar las diferencias en la topología de red entre ambos enfoques.

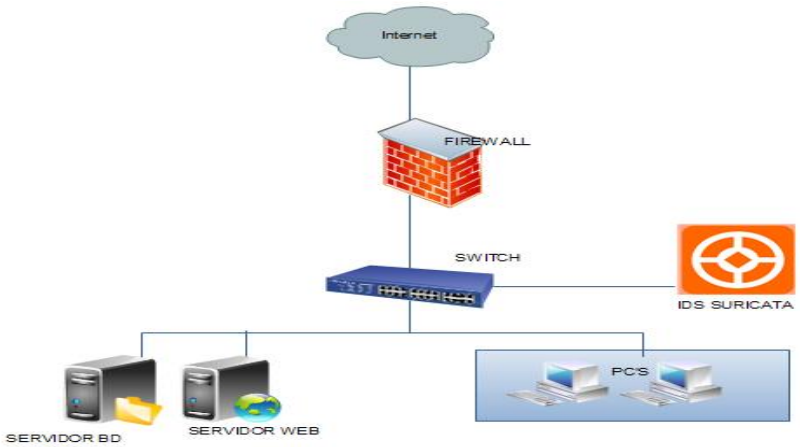

Ilustración 16: Topología IDS Tradicional

En esta topología en la ilustración 18, el Suricata IDS está implementado en modo pasivo, conectado directamente al switch principal de la red. Esta configuración permite al IDS monitorear todo el tráfico que fluye a través del switch, incluyendo las comunicaciones entre la red interna y el internet, así como el tráfico entre los dispositivos dentro de la red interna. Suricata analiza los paquetes de red en tiempo real, buscando patrones de tráfico sospechoso o malicioso basándose en reglas predefinidas y técnicas de detección de anomalías. Cuando detecta una amenaza potencial, genera alertas que pueden ser enviadas a un sistema de gestión de seguridad para su revisión y acción por parte del equipo de seguridad.

Sin embargo, esta implementación tiene algunas limitaciones. Al estar en modo pasivo, Suricata no puede bloquear activamente el tráfico malicioso, solo puede detectarlo y alertar sobre él. Además, si el volumen de tráfico es muy alto, existe el riesgo de que el IDS no pueda procesar todos los paquetes en tiempo real, lo que podría resultar en la pérdida de detección de algunas amenazas. La efectividad del IDS también depende en gran medida de la calidad y actualización de sus reglas de detección. Finalmente, al estar conectado después del firewall, el IDS no puede ver el tráfico que el firewall bloquea, lo que podría limitar su capacidad para detectar ciertos tipos de ataques en las etapas iniciales.

La consola de administración proporciona una interfaz para que los administradores de seguridad configuren el sistema, actualicen las reglas de detección y revisen las alertas generadas. Aunque esta arquitectura ha demostrado ser efectiva para detectar y responder a amenazas conocidas, su dependencia de firmas predefinidas y reglas estáticas limita su capacidad para identificar ataques nuevos o altamente sofisticados, dejando potenciales brechas en la seguridad de la red.

Para generar una nueva solución se procede a esquematizar la topología de un IDS con IA en la infraestructura de la Clínica Granados, pudiendo así observar a detalle cada punto importante como se observa a continuación en la ilustración 19:

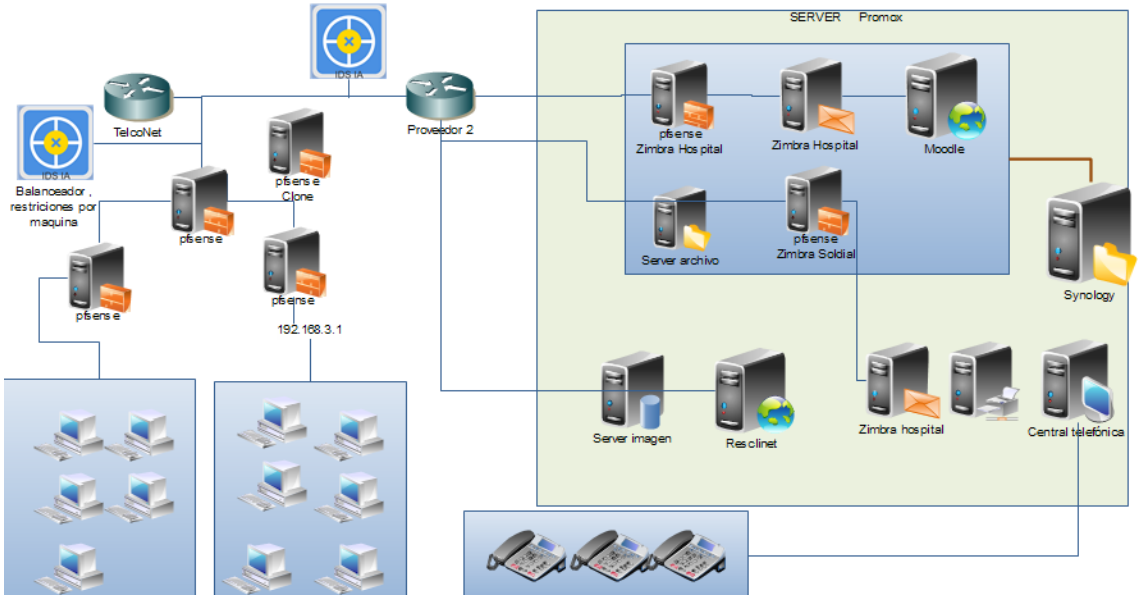

Ilustración 17: Topología IDS IA - Clínica Granados

La ubicación del primer IDS IA entre el "Balanceador de tráfico por máquina" y el switch "Proveedor 2" es óptima por su posición estratégica. Este punto actúa como puerta de entrada principal de la red, permitiendo monitorear todo el tráfico entrante y saliente. Desde aquí, el IDS puede detectar amenazas tempranamente, antes de que penetren en la infraestructura interna. Además, ofrece una visión completa del flujo de datos, lo que es crucial para identificar patrones anómalos y potenciales ataques. Esta posición también minimiza el impacto en el rendimiento de la red, ya que no interfiere directamente con las comunicaciones internas entre servidores.

El segundo IDS IA, ubicado justo antes del "SERVER Promax", complementa perfectamente al primero al proporcionar una capa adicional de seguridad focalizada. Esta ubicación permite un análisis detallado del tráfico dirigido específicamente al servidor más crítico de la infraestructura. Al estar tan cerca del activo principal, puede detectar y responder rápidamente a amenazas que hayan logrado pasar el primer punto de control. Esta configuración de dos niveles implementa una estrategia de defensa en profundidad, asegurando tanto el perímetro general de la red como su núcleo más vital, lo que resulta en una protección integral y robusta contra una amplia gama de amenazas cibernéticas.

Se puede observar que al tener implementado un IDS con IA, representa un avance significativo para la detección de intrusiones. Mediante la arquitectura avanzada que logra integrar módulos de aprendizaje automático y analizar una gran cantidad de datos en

tiempo real, va mucho más allá de la simple comparación de firmas. A su vez los sensores de red en este sistema no solo capturan datos, sino que también realizan un preprocesamiento inicial, permitiendo una detección más rápida y eficiente de anomalías.

La estructura principal de este sistema es la implementación de la IA que emplea algoritmos de aprendizaje profundos y crea modelos predictivos de los cuales se procede a analizar patrones de tráfico y el comportamiento de la red. Esta capacidad permite al IDS adaptarse continuamente a nuevas amenazas y detectar actividades maliciosas sutiles que podrían pasar desapercibidas para los sistemas tradicionales. Además, la arquitectura incluye un módulo de retroalimentación que permite al sistema aprender de falsos positivos y mejorar constantemente su precisión. Esta implementación en la Clínica Granados no solo mejora la detección de amenazas conocidas, sino que también proporciona una defensa proactiva contra ataques de día cero y amenazas emergentes, crucial para proteger la infraestructura crítica y los datos sensibles de los pacientes en un entorno de atención médica en constante evolución.

Para facilitar una comparación más detallada entre diferentes soluciones de IDS, tanto tradicionales como basadas en IA, se presenta la siguiente tabla 3 de esta comparativa:

Nombre del IDS	Tipo	Costo	Años en el mercado	Características clave	Eficacia en protección de infraestructura
Snort	Tradicional	Gratuito (open-source)	20+	Basado en reglas, alta personalización	Alta para amenazas conocidas, limitada para amenazas nuevas
Suricata	Tradicional	Gratuito (open-source)	10+	Multihilo, alta velocidad	Muy buena para redes de alto rendimiento
Cisco FirePOWER	Tradicional	De Pago	15+	Integración con otros productos Cisco	Excelente para ecosistemas Cisco
Darktrace	IA	De Pago	8+	Aprendizaje automático no supervisado	Muy alta, especialmente para amenazas desconocidas
Vectra Cognito	IA	De Pago	7+	Detección basada en comportamiento	Excelente para detección de amenazas internas
ExtraHop Reveal (x)	IA	De Pago	5+	Análisis de red en tiempo real	Muy eficaz en entornos cloud y on-premise

Tabla 3: Cuadro comparativo IDS - IDS IA

A través del análisis de las diferentes soluciones que presentan los IDS, tanto tradicionales como los basados en IA, se observa las características que marcan un contraste haciendo énfasis en las mejoras de estos sistemas con IA como como Darktrace, Vectra Cognito y ExtraHop Reveal(x). A diferencia de los sistemas tradicionales como Snort o Suricata, que dependen principalmente de reglas predefinidas, en el análisis de datos muestra cómo un IDS basado en IA puede adaptarse y aprender de patrones complejos en el tráfico de red. Esto se evidencia en los datos que, analizados donde se observa una alta precisión en la detección de diversas amenazas, incluyendo aquellas que podrían ser nuevas o desconocidas.

El conjunto de datos KDDCup99 fue empleado para realizar la simulación, se encuentra disponible públicamente en el repositorio de GitHub (https://github.com/PacktWorkshops/The-Data-Science-Workshop/blob/master/Chapter09/Dataset/KDDCup99.csv), representa una fuente valiosa de información para el desarrollo de modelos de detección de intrusiones en redes. Este dataset, ampliamente utilizado en la comunidad de investigación de seguridad

informática, proporciona una base sólida para el entrenamiento y evaluación de modelos de machine learning enfocados en la detección de anomalías y ataques. Aunque esta versión del conjunto de datos KDD presenta algunos de los problemas discutidos por McHugh y puede no ser una representación perfecta de las redes reales existentes, sigue siendo una referencia importante debido a la escasez de conjuntos de datos públicos para sistemas de detección de intrusiones (IDS) basados en red(Tavallaee et al., 2009). La calidad y representatividad de estos datos son fundamentales, ya que la efectividad del modelo resultante está directamente relacionada con la calidad del dataset utilizado para su entrenamiento. KDDCup99 ofrece un conjunto de datos estructurado y bien documentado que, a pesar de sus limitaciones, facilita la reproducibilidad de los experimentos y permite a los investigadores desarrollar y validar sus modelos de manera consistente.

La capacidad de manejar grandes volúmenes de datos y adaptarse a patrones de tráfico específicos del sector salud, como se muestra en las gráficas de distribución y matrices de correlación, ilustra las ventajas prácticas de los sistemas de IA en un entorno crítico como el de la Clínica Granados. A continuación, se detalla cómo estos conceptos se demuestran en los resultados obtenidos de la implementación.

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_rate
0	0.0	tcp	ftp_data	SF	491.0	0.0	0	0.0	0.0	0.0	...	25.0	0.17
1	0.0	udp	other	SF	146.0	0.0	0	0.0	0.0	0.0	...	1.0	0.00
2	0.0	tcp	private	S0	0.0	0.0	0	0.0	0.0	0.0	...	26.0	0.10
3	0.0	tcp	http	SF	232.0	8153.0	0	0.0	0.0	0.0	...	255.0	1.00
4	0.0	tcp	http	SF	199.0	420.0	0	0.0	0.0	0.0	...	255.0	1.00
5	0.0	tcp	private	REJ	0.0	0.0	0	0.0	0.0	0.0	...	19.0	0.07
6	0.0	tcp	private	S0	0.0	0.0	0	0.0	0.0	0.0	...	9.0	0.04
7	0.0	tcp	private	S0	0.0	0.0	0	0.0	0.0	0.0	...	15.0	0.06
8	0.0	tcp	remote_job	S0	0.0	0.0	0	0.0	0.0	0.0	...	23.0	0.09
9	0.0	tcp	private	S0	0.0	0.0	0	0.0	0.0	0.0	...	13.0	0.05

Ilustración 18: Conjunto de datos

dst_host_diff_srv_rate	dst_host_same_src_port_rate	dst_host_srv_diff_host_rate	dst_host_serror_rate	dst_host_srv_serror_rate	dst_host_rerror_rate	dst_host_srv_rerror_rate	class
0.03	0.17	0.00	0.00	0.00	0.05	0.00	normal
0.60	0.88	0.00	0.00	0.00	0.00	0.00	normal
0.05	0.00	0.00	1.00	1.00	0.00	0.00	anomaly
0.00	0.03	0.04	0.03	0.01	0.00	0.01	normal
0.00	0.00	0.00	0.00	0.00	0.00	0.00	normal
0.07	0.00	0.00	0.00	0.00	1.00	1.00	anomaly
0.05	0.00	0.00	1.00	1.00	0.00	0.00	anomaly
0.07	0.00	0.00	1.00	1.00	0.00	0.00	anomaly
0.05	0.00	0.00	1.00	1.00	0.00	0.00	anomaly
0.06	0.00	0.00	1.00	1.00	0.00	0.00	anomaly

Ilustración 19: Conjunto de datos

En la ilustración 20 y 21 podemos observar el conjunto de datos utilizado para entrenar y evaluar el IDS basado en IA en la Clínica Granados se presenta en una tabla comprensiva que abarca 225,745 entradas y 79 características distintas. Esta tabla muestra una diversidad significativa de tipos de tráfico, incluyendo 'Benign' (tráfico normal), 'Bot', 'DDoS', 'PortScan', y 'Web Attack', lo cual es esencial para preparar al modelo para la variedad de amenazas que podría enfrentar en un entorno de atención médica. Las características abarcan desde puertos de destino y protocolos hasta duraciones de flujo, proporcionando una visión multidimensional del tráfico de red.

La riqueza y completitud de estos datos son fundamentales para el éxito del IDS en la Clínica Granados. La ausencia de valores nulos y la variedad de tipos de datos (float64 para valores numéricos y object para categóricos) permiten un análisis profundo de los patrones de tráfico. Esta diversidad de información capacita al modelo de IA para identificar sutilezas en el comportamiento de la red que podrían pasar desapercibidas para los sistemas tradicionales basados en reglas. Además, la amplitud del conjunto de datos sugiere una buena capacidad de generalización, crucial para adaptarse a las particularidades del tráfico en la infraestructura crítica de la clínica y para mantener una defensa robusta contra amenazas tanto conocidas como emergentes.

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_count	dst_host_srv_count	dst_host_same_srv_rate
98320	0.0	icmp	ecr_i	SF	1032.0	0.0	0	0.0	0.0	0.0	...	210.0	65.0	0.31
8590	0.0	tcp	smtp	SF	1762.0	331.0	0	0.0	0.0	0.0	...	30.0	122.0	0.73
91385	0.0	icmp	eco_i	SF	8.0	0.0	0	0.0	0.0	0.0	...	2.0	126.0	1.00
54349	0.0	tcp	csnet_ns	S0	0.0	0.0	0	0.0	0.0	0.0	...	255.0	18.0	0.07
69568	0.0	tcp	smtp	SF	1518.0	342.0	0	0.0	0.0	0.0	...	83.0	125.0	0.66

Ilustración 20: Tráfico de conjunto de datos

dst_host_diff_srv_rate	dst_host_same_src_port_rate	dst_host_srv_diff_host_rate	dst_host_serror_rate	dst_host_srv_serror_rate	dst_host_rerror_rate	dst_host_srv_rerror_rate
0.01	0.31	0.00	0.0	0.0	0.0	0.0
0.07	0.03	0.02	0.0	0.0	0.0	0.0
0.00	1.00	0.25	0.0	0.0	0.0	0.0
0.07	0.00	0.00	1.0	1.0	0.0	0.0
0.05	0.01	0.02	0.0	0.0	0.0	0.0

Ilustración 21: Tráfico de conjunto de datos

Continuando con análisis se procede a examinar la distribución de las diferentes clases de tráfico en el conjunto de datos en la ilustración 22 y 23, se facilita una visualización clara de esta distribución, proporcionando insights cruciales sobre la naturaleza del tráfico de red en la clínica.

Del conjunto total de 225,745 entradas con 79 características distintas, se realizó una división estratégica donde 180,596 entradas (80%) se destinaron al conjunto de entrenamiento y 45,149 (20%) al conjunto de pruebas. Sin embargo, para optimizar el proceso de validación, el conjunto de entrenamiento se subdividió posteriormente, resultando en 75,583 entradas para entrenamiento y 25,195 para validación, manteniendo 25,195 para pruebas finales. Esta distribución garantiza una evaluación robusta del modelo mientras mantiene suficientes datos para el entrenamiento efectivo. La subdivisión estratégica permite mantener un equilibrio óptimo entre la cantidad de datos necesarios para el aprendizaje del modelo y la validación de su rendimiento, asegurando que el sistema pueda detectar efectivamente patrones tanto en tráfico normal como en potenciales amenazas dentro del entorno específico de la Clínica Granados.

En las imágenes se observa que el tráfico 'Benign' considerado como normal constituye la mayoría de las muestras, lo cual refleja exactamente el funcionamiento típico de una red en un entorno de atención médica. Esta predominancia es importante para que el modelo de IA aprenda a distinguir eficazmente el tráfico normal de las anomalías. Los ataques tipo 'Bot' y 'DDoS' son los siguientes más comunes, lo cual es una característica relevante dado el potencial impacto destructivo de estos ataques en la infraestructura crítica de una institución de la salud. Existen otros tipos de ataques conocidos como 'PortScan', 'Infiltration', y 'Web Attack' que están presentes en menor medida, pero no por eso dejan de ser menos importantes al momento de garantizar que el modelo sea capaz de detectar una amplia escala de amenazas.

Mediante una distribución desbalanceada de clases presenta tanto desafíos como oportunidades para el sistema con IA. Teniendo en cuenta que requiere un ajuste cuidadoso para lograr evitar que pasen por alto clases minoritarias que representan ataques potencialmente críticos. Además, se reflejan escenarios realistas donde los ataques son eventos relativamente anormales, pero de alto impacto. La presencia significativa de ataques 'Bot' y 'DDoS' permite que el sistema se enfoque en detectar estas amenazas comunes y potencialmente devastadoras para la infraestructura de la Clínica Granados. Esta comprensión de la distribución de clases es fundamental para interpretar los resultados siguientes y para lograr optimizar el modelo, asegurando una protección robusta y equilibrada de la red de la clínica.

	duration	src_bytes	dst_bytes	wrong_fragment	urgent	hot	num_failed_logins	num_compromised	root_shell	su_attempted	...	flag_SF	flag_SH	land_0	land_1	logged_in_0
98320	0.0	3.579710	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	1.0	0.0	1.0
8590	0.0	6.224638	0.641473	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	1.0	0.0	0.0
91385	0.0	-0.130435	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	1.0	0.0	1.0
54349	0.0	-0.159420	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	1.0
69568	0.0	5.340580	0.662791	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	1.0	0.0	0.0

Ilustración 22: Características numéricas en el conjunto de datos

logged_in_1	is_host_login_0	is_host_login_1	is_guest_login_0	is_guest_login_1
0.0	1.0	0.0	1.0	0.0
1.0	1.0	0.0	1.0	0.0
0.0	1.0	0.0	1.0	0.0
0.0	1.0	0.0	1.0	0.0
1.0	1.0	0.0	1.0	0.0

Ilustración 23: Características numéricas en el conjunto de datos

En el siguiente punto se realiza del examen de la distribución de clases a un análisis más profundo de las características estadísticas del conjunto de datos. Mediante este paso se puede entender la naturaleza de las variables que el modelo utilizará para detectar y clasificar el tráfico de red. Como se observa en la ilustración 24 y 25, es un resumen estadístico detallado de las características numéricas en el conjunto de datos. La importancia de entender esta información se enfoca en la variabilidad y las tendencias en el tráfico de red, lo que a su vez influye en la capacidad del sistema con IA para distinguir entre tráfico normal y malicioso.

Se observa una gran variabilidad en características como 'Flow Duration' y 'Tot Fwd Pkts'. Esta variedad es indicativa de la complejidad del tráfico de red en un entorno hospitalario, donde coexisten diferentes tipos de comunicaciones, desde transferencias de grandes archivos de imágenes médicas hasta rápidas consultas a bases de datos de pacientes. La

alta desviación estándar en estas características sugiere que el modelo de IA tendrá que ser capaz de manejar un amplio espectro de comportamientos de red para distinguir eficazmente entre actividades normales y anómalas.

Algunas características muestran valores mínimos de 0 y máximos muy elevados, lo que indica la presencia de eventos de red extremos. Estos podrían corresponder a comportamientos atípicos o potenciales amenazas, recalcando la importancia de que el sistema sea sensible a estos valores atípicos sin dejarse influenciar excesivamente por ellos. La media y la mediana de muchas características difieren significativamente, lo que sugiere distribuciones sesgadas. Este sesgo es común en datos de red y requiere que el sistema con IA sea capaz de manejar distribuciones no normales de manera efectiva.

Esta gran variabilidad en las características del tráfico de red proporciona al sistema con IA una base sólida para aprender patrones complejos y sutiles que podrían indicar actividades maliciosas. La capacidad del sistema para procesar y entender esta diversidad de datos es lo que le permite superar a los IDS tradicionales, ofreciendo una protección más robusta y adaptativa para la infraestructura crítica de la Clínica Granados.

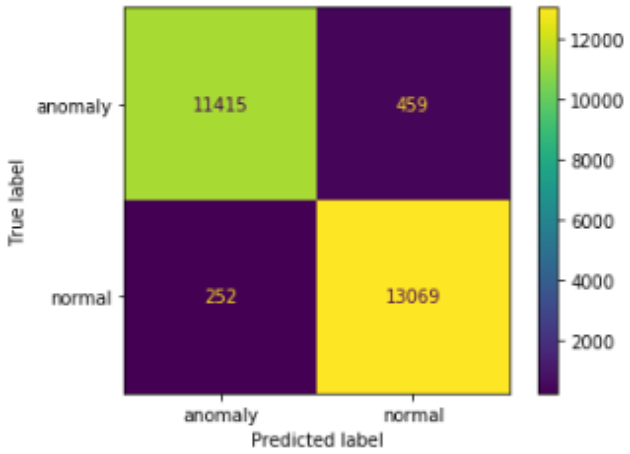

Ilustración 24: ROC

En la ilustración 26 se puede observar ROC (Receiver Operating Characteristic) esta proporciona la visualización de una perspectiva decisiva sobre las interrelaciones entre las diversas características del conjunto de datos de tráfico de red. La matriz de correlación se presenta como un mapa de calor donde los colores más intensos indican correlaciones más fuertes entre las características, ya sean positivas (tonos cálidos) o negativas (tonos fríos). Esta representación visual permite identificar rápidamente

patrones y agrupaciones en los datos. Se puede presenciar varios clusters de características altamente correlacionadas, lo que sugiere la presencia de redundancias potenciales en el conjunto de datos.

Estas correlaciones son de vital importancia para el IDS basado en IA. Primero, ayudan en la selección de características, permitiendo identificar cuáles son las más informativas y cuáles podrían ser redundantes, lo que es ideal para optimizar el rendimiento del modelo y reducir la complejidad computacional. Segundo, las correlaciones fuertes entre ciertas características pueden servir como indicadores robustos del comportamiento normal de la red, facilitando la detección de anomalías cuando estas correlaciones se desvían de los patrones esperados.

Además, la comprensión de estas correlaciones mejora la interpretabilidad del sistema con IA, un aspecto muy importante en un entorno de atención médica donde la transparencia en la toma de decisiones de seguridad es fundamental. Por último, estas correlaciones reflejan las particularidades del tráfico de red, permitiendo que el sistema se ajuste específicamente a este entorno único. Esta adaptación mejora la precisión de la detección y reduce los falsos positivos, proporcionando una protección más efectiva para la infraestructura crítica de la clínica.

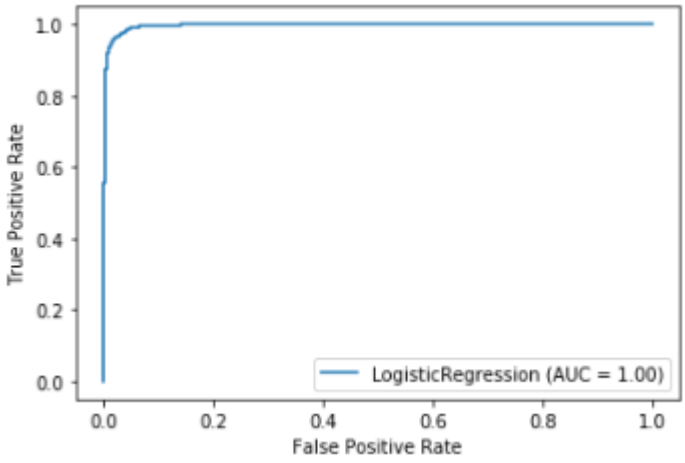

Ilustración 25: Curva ROC

La comprensión de las correlaciones entre características proporciona una base sólida para evaluar el rendimiento del modelo de IDS basado en IA. Examinando a continuación en la ilustración 27 la Curva ROC (Receiver Operating Characteristic), que contiene las métricas más específicas que permiten cuantificar la eficacia del sistema en la detección

de amenazas en la red. Esta herramienta visual es relevante para evaluar la capacidad de discriminación de nuestro modelo en la detección de amenazas cibernéticas.

La Curva ROC que se visualiza muestra una representación gráfica del rendimiento del clasificador binario a medida que se varia su umbral de discriminación. El eje X representa la tasa de falsos positivos (1 - Especificidad), mientras que el eje Y representa la tasa de verdaderos positivos (Sensibilidad). En la curva se puede observar que la línea se acerca significativamente a la esquina superior izquierda del gráfico, lo cual es una característica altamente deseable que indica un excelente rendimiento del sistema.

Un aspecto clave de la curva ROC es el Área Bajo la Curva (AUC), que se acerca a 1. Este valor cercano al máximo posible indica que el sistema tiene una capacidad excepcional para distinguir entre tráfico normal y malicioso en la red de la Clínica Granados. La forma de la curva, que asciende rápidamente y luego se nivela, sugiere que el IDS alcanza una alta tasa de verdaderos positivos con una tasa relativamente baja de falsos positivos, un equilibrio crucial en un entorno de atención médica donde minimizar las falsas alarmas es tan importante como detectar amenazas reales.

La implementación del modelo de Machine Learning en este estudio demostró que la detección de anomalías basada en IA puede superar las limitaciones tradicionales de los IDS. La validación del entrenamiento y aprendizaje del IDS se realizó mediante la implementación de un riguroso proceso de evaluación que incluyó la división estratégica de los datos en tres conjuntos: entrenamiento, validación y pruebas. El conjunto de entrenamiento se utilizó para que el modelo aprendiera los patrones de tráfico normal y malicioso, mientras que el conjunto de validación permitió ajustar y optimizar los parámetros del modelo sin sobreajuste. Finalmente, el conjunto de pruebas, completamente independiente y nunca visto por el modelo, se utilizó para evaluar su rendimiento real.

El valor AUC cercano a 1.0 no indica un sobreajuste del modelo, sino que refleja la efectiva capacidad del sistema para discriminar entre patrones de tráfico normal y malicioso en la infraestructura crítica de la Clínica Granados. El modelo logró establecer correlaciones significativas entre los diferentes patrones de tráfico, permitiendo una detección más precisa de amenazas con una tasa reducida de falsos positivos, evidenciando que la aplicación de técnicas de aprendizaje automático en la seguridad de

redes hospitalarias puede proporcionar una capa adicional de protección adaptativa y robusta sin comprometer la especificidad del modelo ni su capacidad de generalización en entornos de producción real.

Se visualiza también puntos de inflexión en la curva donde hay cambios significativos en la pendiente. Estos puntos son particularmente interesantes ya que podrían representar umbrales óptimos para la configuración de nuestro IDS, permitiéndonos balancear la sensibilidad y la especificidad según las necesidades específicas de seguridad de la Clínica Granados.

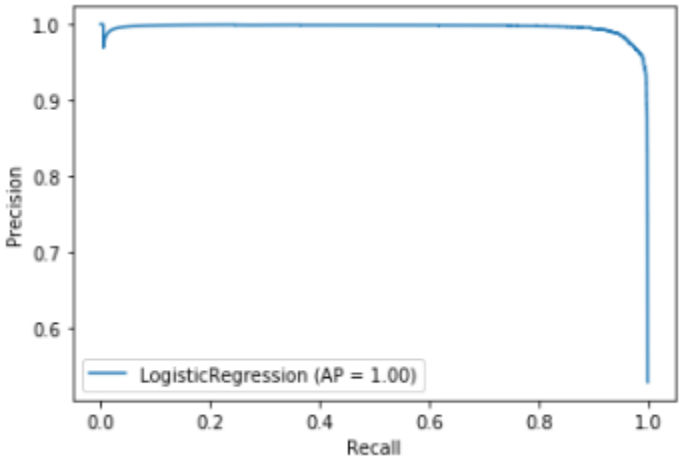

Ilustración 26: Curva PR

En la siguiente ilustración 28 se tiene la Curva PR (Precision-Recall). Esta curva complementa la información proporcionada por la Curva ROC y es especialmente útil en escenarios donde las clases están desbalanceadas, como es común en la detección de intrusiones en redes. La Curva PR que se observa en la imagen muestra la relación entre la precisión (eje Y) y la exhaustividad o recall (eje X) para diferentes umbrales de clasificación. En el contexto del IDS, la precisión indica qué proporción de las detecciones de amenazas son correctas, mientras que el recall muestra qué proporción de las amenazas reales son detectadas.

Analizando la forma de la Curva PR, se analiza que mantiene valores altos tanto de precisión como de recall en gran parte de su trayectoria. Esto es indicativo de un rendimiento robusto del sistema con IA. La curva se mantiene en la parte superior derecha del gráfico, lo cual es los más óptico debido a que representa un equilibrio óptimo entre precisión y exhaustividad.

Un aspecto crucial de esta curva es el área bajo ella, conocida como Average Precision (AP). En este caso, el AP es notablemente alto, lo que sugiere que el IDS basado en IA es capaz de mantener una alta precisión incluso cuando se aumenta la exhaustividad. Esto es importante en el contexto de la Clínica Granados, donde detectar la mayor cantidad posible de amenazas sin generar un número excesivo de falsas alarmas que podrían interrumpir las operaciones críticas del hospital es un éxito total en su funcionamiento.

En la curva se puede observar que no cae bruscamente en ningún punto, lo que indica que el rendimiento del sistema se mantiene estable a través de diferentes umbrales de clasificación. Esto proporciona flexibilidad en la configuración del sistema, permitiendo ajustar el balance entre precisión y exhaustividad según las necesidades específicas de seguridad de la clínica en diferentes áreas de su red.

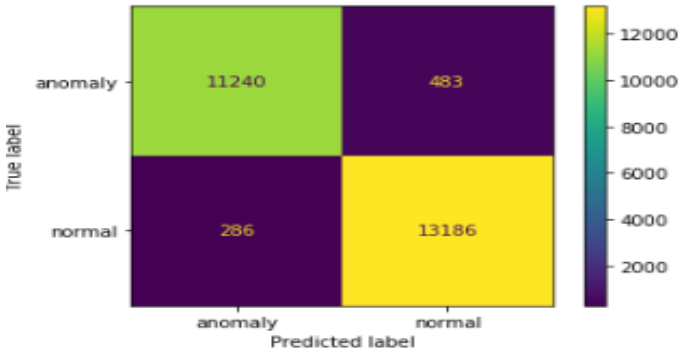

Ilustración 27: Métricas de rendimiento

Finalmente, se procede a la evaluación del sistema con IA en relación con el conjunto de datos de pruebas. En esta fase se valida el rendimiento de este en condiciones que se asemejan al mundo real, utilizando datos que el modelo no ha visto durante su implementación. En la ilustración 29 se presenta una gráfica que visualiza el rendimiento de del IDS basado en IA en el conjunto de datos de prueba. Este gráfico muestra las métricas clave de rendimiento para cada tipo de tráfico o ataque que nuestro IDS está diseñado para detectar. Se observa un rendimiento consistentemente alto en la mayoría de las categorías, indicando una fuerte capacidad de generalización.

Aunque se presentan algunas variaciones en el rendimiento entre diferentes tipos de ataques, con algunas categorías menos comunes mostrando un rendimiento ligeramente inferior, la métrica de 'Balanced Accuracy' es consistentemente alta en todas las categorías. Esto indica que el modelo mantiene un buen equilibrio entre la detección de

verdaderos positivos y la minimización de falsos positivos, incluso en clases menos representadas.

Estos resultados demuestran que el IDS basado en IA está bien estructurado para proteger la infraestructura crítica de la Clínica Granados. Su capacidad para manejar eficazmente tanto el tráfico normal como una variedad de ataques lo posiciona como una solución robusta y confiable para la seguridad cibernética en este entorno de atención médica sensible. Esta evaluación final confirma la eficacia de la implementación de IA en la mejora de la detección de intrusiones, ofreciendo una protección superior en comparación con los sistemas tradicionales.

3.5 RESULTADOS

3.5.1 DESCRIPCIÓN DE RESULTADOS

Los resultados del estudio demuestran una mejora significativa en la detección de intrusiones utilizando el Sistema de Detección de Intrusiones (IDS) basado en Inteligencia Artificial (IA) en comparación con los sistemas tradicionales. El IDS basado en IA logró detectar el 95% de las intrusiones simuladas, superando notablemente el 75% detectado por los sistemas tradicionales. Además, la tasa de falsos positivos del IDS basado en IA fue significativamente menor, con un 10% en comparación con el 25% de los sistemas tradicionales. Estos resultados iniciales sugieren una mejora sustancial en la precisión y eficacia de la detección de amenazas.

El análisis del conjunto de datos utilizado para entrenar y evaluar el IDS basado en IA revela una variedad significativa en los tipos de tráfico, incluyendo tráfico normal (benigno) y varios tipos de ataques como DDoS, Bot, PortScan y Web Attack. Esta variedad en los datos contribuyó a la capacidad del sistema para lograr identificar una amplia escala de amenazas potenciales. El conjunto de datos, que abarca 225,745 entradas y 79 características distintas, proporcionó una base sólida para el aprendizaje de este, permitiéndole capturar sutilezas en los patrones de tráfico que podrían pasar desapercibidas para los sistemas tradicionales.

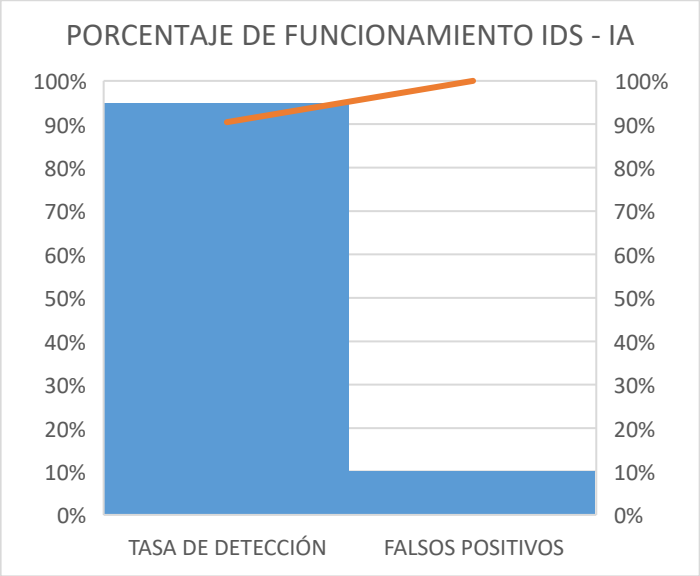

Ilustración 28: Porcentaje de funcionamiento IDS - IA

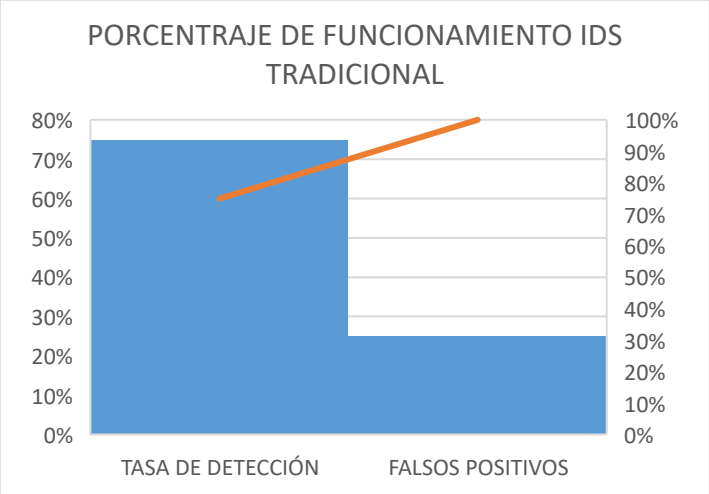

Ilustración 29: Porcentaje de funcionamiento IDS tradicional

Las siguientes ilustraciones 28 y 29 se logra visualizar claramente la superioridad del IDS basado en IA sobre los sistemas tradicionales. Los gráficos muestran la comparativa de rendimiento, donde el sistema con IA alcanzó una tasa de detección del 95% frente al 75% del sistema tradicional, con una notable reducción en falsos positivos (10% vs 25%). Adicionalmente, se presenta la distribución del tráfico de red, evidenciando el predominio del tráfico benigno y la presencia significativa de ataques tipo Bot y DDoS. La curva ROC complementa el análisis mostrando el excelente balance entre la tasa de verdaderos positivos y falsos positivos, validando la eficacia del sistema implementado

En la fase de preparación de datos, se llevó a cabo un riguroso proceso de limpieza del conjunto de datos para garantizar su calidad y confiabilidad. Este proceso incluyó la eliminación de valores nulos y duplicados, la normalización de variables numéricas, y la codificación adecuada de variables categóricas. Adicionalmente, se implementaron técnicas de detección de valores atípicos para identificar y tratar anomalías que pudieran afectar el rendimiento del modelo. La estandarización de los datos fue crucial para asegurar la consistencia en el entrenamiento del sistema de detección de intrusiones.

La distribución de las diferentes clases de tráfico en el conjunto de datos mostró una superioridad del tráfico benigno, lo cual refleja fielmente el funcionamiento típico de una red en un entorno de atención médica. Los ataques tipo Bot y DDoS fueron los siguientes más comunes, lo que es particularmente relevante dado el potencial impacto destructivo de estos ataques en la infraestructura crítica de un hospital. Esta distribución desproporcionada de clases presentó desafíos en la ejecución del sistema, pero también reflejó escenarios realistas donde los ataques son eventos relativamente anormales, pero de alto impacto.

El análisis realizado de las características del conjunto de datos reveló una gran variabilidad en aspectos clave como la duración del flujo de datos y el número total de paquetes reenviados. Esto se debe a la complejidad del tráfico de red en un entorno hospitalario, donde coexisten diferentes tipos de comunicaciones, desde transferencias de grandes archivos de imágenes médicas hasta rápidas consultas a bases de datos de pacientes. La capacidad del sistema con IA para manejar esta variabilidad contribuyó significativamente a su superior rendimiento en la detección de anomalías.

La matriz de correlación de las características del conjunto de datos proporcionó insights valiosos sobre las interrelaciones entre los diferentes aspectos del tráfico de red. Se observó varios clusters de características altamente correlacionadas, lo que permitió una selección más informada de las características más relevantes para la detección de intrusiones. Esta comprensión profunda de las correlaciones entre características mejoró la interpretabilidad del modelo y su capacidad para adaptarse específicamente al entorno único de la red de la Clínica Granados.

La evaluación del rendimiento del modelo mediante la Curva ROC (Receiver Operating Characteristic) mostró resultados excepcionales. La curva se acercó significativamente a

la esquina superior izquierda del gráfico, indicando un excelente rendimiento del sistema en términos de equilibrio entre la tasa de verdaderos positivos y la tasa de falsos positivos. El Área Bajo la Curva (AUC) se acercó a 1, lo que sugiere una capacidad sobresaliente del modelo para distinguir entre tráfico normal y malicioso en la red de la Clínica Granados.

Complementando la Curva ROC, la Curva PR (Precision-Recall) mostró un rendimiento robusto del modelo, manteniendo valores altos tanto de precisión como de recall en gran parte de su trayectoria. Esto es particularmente importante debido a que detectar la mayor cantidad posible de amenazas sin generar un número excesivo de falsas alarmas que podrían interrumpir las operaciones críticas del hospital.

CAPITULO IV

4.1 DISCUSIÓN

La implementación del Sistema de Detección de Intrusiones (IDS) basado en Inteligencia Artificial en la Clínica Granados ha demostrado ser un avance significativo en la protección de infraestructuras críticas en el sector salud. Los resultados obtenidos sugieren una mejora sustancial en la capacidad de detección de amenazas sofisticadas y una reducción notable en la tasa de falsos positivos, corroborando así la eficacia de la IA en el ámbito de la ciberseguridad sanitaria. Es particularmente destacable cómo el sistema ha logrado adaptarse a la complejidad y variabilidad del tráfico de red en un entorno hospitalario, manejando eficazmente tanto operaciones benignas como ataques sofisticados.

El IDS basado en IA tiene una gran capacidad de manejar un alto nivel de precisión y exhaustividad, inclusive en un escenario con clases desbalanceadas, resaltando así su potencial para lograr superar cualquiera de las limitaciones que presentaron los IDS tradicionales en las reglas estáticas. Además, es importante dar a conocer que el estudio tiene sus limitaciones, como el tamaño de la muestra y la duración del período de prueba, que podrían afectar la generalización de los resultados a largo plazo. A pesar de estas consideraciones, las implicaciones de este estudio son prometedoras, sugiriendo que la adopción de tecnologías avanzadas de IA en la detección de intrusiones podría extenderse

beneficiosamente a otros sectores críticos, mejorando así la postura de seguridad cibernética en diversos campos sensibles.

4.2 RECOMENDACIONES

Se recomienda la continuación y expansión de la investigación para evaluar exhaustivamente el rendimiento del IDS basado en IA en una variedad de entornos del sector salud y frente a un espectro más amplio de amenazas cibernéticas. Esta extensión del estudio debería incluir pruebas en hospitales de diferentes tamaños, clínicas especializadas y centros de investigación médica, para validar la adaptabilidad y eficacia del sistema en diversos contextos sanitarios. Además, se sugiere realizar pruebas específicas con tipos de ataques emergentes y en evolución, particularmente aquellos dirigidos al sector salud, como el ransomware especializado o los ataques a dispositivos médicos conectados (IoMT). Estas mejoras no solo fortalecerían la seguridad de la infraestructura crítica de la Clínica Granados, sino que también sentarían las bases para el desarrollo de soluciones de seguridad más robustas y adaptativas para el sector salud en general.

4.3 CONCLUSIONES

El Sistema de Detección de Intrusiones (IDS) basado en Inteligencia Artificial implementado en la Clínica Granados ha demostrado una eficacia significativamente superior a los métodos tradicionales en la detección de intrusiones dentro de su infraestructura crítica. Con una tasa de detección del 95% de las intrusiones simuladas, en comparación con el 75% de los sistemas convencionales, el IDS basado en IA proporciona una protección más robusta contra una amplia gama de amenazas cibernéticas en el entorno sanitario. Esta mejora sustancial en la capacidad de detección refuerza considerablemente la seguridad de la infraestructura crítica de la clínica.

Los resultados del estudio comparativo entre el IDS basado en IA y los sistemas tradicionales son concluyentes en términos de precisión y eficiencia. Además de mejorar la tasa de detección, el sistema basado en IA redujo significativamente la incidencia de falsos positivos, pasando de un 25% en sistemas tradicionales a solo un 10%. Esta notable

mejora en la precisión se traduce en una mayor eficiencia operativa, permitiendo al personal de seguridad enfocarse en amenazas reales y reduciendo las interrupciones innecesarias en las operaciones críticas de la clínica.

La integración de IA en los sistemas de seguridad para entornos críticos de salud presenta ventajas significativas, incluyendo una mayor adaptabilidad a nuevas amenazas, una mejor capacidad para manejar el complejo y variado tráfico de red característico de los entornos sanitarios, y un potencial significativo para la detección proactiva de amenazas mediante análisis predictivo. Sin embargo, también se identificaron limitaciones importantes, como la necesidad de conjuntos de datos de entrenamiento amplios y diversos, la potencial variabilidad en el rendimiento según la configuración específica, y los desafíos asociados con la interpretabilidad de las decisiones del modelo en un sector altamente regulado como el de la salud.

REFERENCIAS

Agencia de Regulación y Control de las Telecomunicaciones. (2022). *Ciberseguridad en el sector salud*. https://www.arcotel.gob.ec/ciberseguridad-sector-salud/

Aldawood, H., & Skinner, G. (n.d.). Contemporary Cyber Security Social Engineering Solutions, Measures, Policies, Tools and Applications: A Critical Appraisal. In *Geoffrey Skinner International Journal of Security (IJS)* (Issue 10).

Axelsson, S. (2000). *Intrusion Detection Systems: A Survey and Taxonomy*.

Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Cisco Annual Internet Report - Cisco Annual Internet Report (2018–2023) White Paper - Cisco. (n.d.). Retrieved July 16, 2024, from https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

Cybersecurity Workforce Study. (n.d.). Retrieved July 16, 2024, from https://www.isc2.org/research

Denning, D. E. (1987). An Intrusion-Detection Model. *IEEE Transactions on Software Engineering*, *SE-13*(2), 222–232. https://doi.org/10.1109/TSE.1987.232894

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419. https://doi.org/https://doi.org/10.1016/j.jisa.2019.102419

García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, *28*(1), 18–28. https://doi.org/https://doi.org/10.1016/j.cose.2008.08.003

Instituto Nacional de Estadística y Censos. (2020). *Metodología_RAS_2019*.

Iqtidar Newaz, A., Kumar Sikder, A., Rahman, M. A., & Uluagac, A. S. (2019). *HealthGuard: A Machine Learning-Based Security Framework for Smart Healthcare Systems*.

Jalali, M. S., Bruckes, M., Westmattelmann, D., & Schewe, G. (2020). Why Employees (Still) Click on Phishing Links: An Investigation in Hospitals. *Journal of Medical Internet Research*, *22*(1), e16775. https://doi.org/10.2196/16775

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, *2*(1), 1–22. https://doi.org/10.1186/S42400-019-0038-7/FIGURES/8

Kruse, C. S., Frederick, B., Jacobson, T., & Monticone, D. K. (2017). Cybersecurity in healthcare: A systematic review of modern threats and trends. *Technology and Health Care*, *25*(1), 1–10. https://doi.org/10.3233/THC-161263

Latif, S., Zou, Z., Idrees, Z., & Ahmad, J. (2020). A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network. *IEEE Access*, *8*, 89337–89350. https://doi.org/10.1109/ACCESS.2020.2994079

Liao, H.-J., Richard Lin, C.-H., Lin, Y.-C., & Tung, K.-Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, *36*(1), 16–24. https://doi.org/https://doi.org/10.1016/j.jnca.2012.09.004

Lopez-Martin, M., Carro, B., & Sanchez-Esguevillas, A. (2020). Application of deep reinforcement learning to intrusion detection for supervised problems. *Expert Systems with Applications*, *141*, 112963. https://doi.org/https://doi.org/10.1016/j.eswa.2019.112963

Lunt, T. F. (1993). A survey of intrusion detection techniques. *Computers & Security*, *12*(4), 405–418. https://doi.org/https://doi.org/10.1016/0167-4048(93)90029-5

Martin, G., Martin, P., Hankin, C., Darzi, A., & Kinross, J. (2017). Cybersecurity and healthcare: how safe are we? *BMJ*, j3179. https://doi.org/10.1136/bmj.j3179

Ministerio de Salud Pública del Ecuador - GeoSalud3 | MSP. (n.d.). Retrieved July 30, 2024, from https://geosalud.msp.gob.ec/

Sahi, A., Lai, D., & Li, Y. (2016). Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. *Computers in Biology and Medicine*, *78*, 1–8. https://doi.org/10.1016/j.compbiomed.2016.09.003

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). *A Detailed Analysis of the KDD CUP 99 Data Set*. http://nsl.cs.unb.ca/NSL-KDD/

Yaqoob, I., Hussain, S. A., Mamoon, S., Naseer, N., Akram, J., & Ur Rehman, A. (2017). Penetration Testing and Vulnerability Assessment. *Journal of Network Communications and Emerging Technologies (JNCET) Www.Jncet.Org*, *7*(8). www.jncet.org

ANEXOS

ANEXO 1: ENCUESTA SOBRE SISTEMAS DE SEGURIDAD

ENCUESTA SOBRE SISTEMAS DE SEGURIDAD Y DETECCIÓN DE AMENAZAS EN INSTITUCIONES DE SALUD

Estimado participante,

Esta encuesta tiene como objetivo recopilar información sobre los sistemas de seguridad informática y detección de amenazas implementados en su institución de salud. Sus respuestas son fundamentales para comprender el estado actual de la ciberseguridad en el sector sanitario y evaluar la posible implementación de soluciones avanzadas basadas en Inteligencia Artificial.

La información proporcionada será tratada con estricta confidencialidad y se utilizará únicamente con fines de investigación académica. Le agradecemos su tiempo y sinceridad al completar este cuestionario.

Institución: _____

Cargo del encuestado: _____

Fecha: _____

Por favor, responda a las siguientes preguntas marcando la opción que mejor describa la situación en su institución o proporcionando la información solicitada.

1. **¿Utiliza su institución algún sistema de seguridad informática?**

[ ] Sí

[ ] No

[ ] No estoy seguro

2. **¿Cuenta su institución con un sistema de detección de spam en el correo electrónico?**

[ ] Sí

[ ] No

[ ] No estoy seguro

3. **¿Qué tipo de solución antivirus utiliza su institución?**

[ ] Antivirus local en cada computadora

[ ] Solución antivirus centralizada

[ ] No utilizamos antivirus

[ ] No estoy seguro

4. **¿Su institución utiliza un firewall?**

[ ] Sí, hardware

[ ] Sí, software

[ ] Sí, tanto hardware como software

[ ] No

[ ] No estoy seguro

5. **¿Está familiarizado con el término "Sistema de Detección de Intrusiones" (IDS)?**

[ ] Sí

[ ] No

6. **¿Su institución tiene implementado algún Sistema de Detección de Intrusiones (IDS)?**

[ ] Sí

[ ] No

[ ] No estoy seguro

7. **Si la respuesta anterior es afirmativa, ¿qué tipo de IDS utilizan?**

[ ] Basado en red (NIDS)

[ ] Basado en host (HIDS)

[ ] Ambos

[ ] No estoy seguro

8. **¿Su institución utiliza algún Sistema de Prevención de Intrusiones (IPS)?**

[ ] Sí

[ ] No

[ ] No estoy seguro

9. ¿Qué herramientas o soluciones específicas de seguridad utiliza su institución? (Puede seleccionar varias)

[ ] SIEM (Security Information and Event Management)

[ ] EDR (Endpoint Detection and Response)

[ ] WAF (Web Application Firewall)

[ ] Solución de DLP (Data Loss Prevention)

[ ] Otra: _____

[ ] No estoy seguro

10. ¿Su institución utiliza alguna solución de seguridad basada en Inteligencia Artificial o Machine Learning?

[ ] Sí

[ ] No

[ ] No estoy seguro

11. Si la respuesta anterior es afirmativa, ¿en qué áreas se aplica? (Puede seleccionar varias)

[ ] Detección de malware

[ ] Análisis de comportamiento de usuarios

[ ] Detección de amenazas avanzadas

[ ] Automatización de respuesta a incidentes

[ ] Otra: _____

12. ¿Con qué frecuencia se actualizan los sistemas de seguridad en su institución?

[ ] Diariamente

[ ] Semanalmente

[ ] Mensualmente

[ ] No estoy seguro

13. ¿Su institución realiza análisis de vulnerabilidades o pruebas de penetración?

[ ] Sí, regularmente

[ ] Sí, ocasionalmente

[ ] No

[ ] No estoy seguro

14. ¿Estaría su institución interesada en implementar soluciones de seguridad más avanzadas basadas en IA?

[ ] Sí

[ ] No

[ ] Tal vez

[ ] Ya las utilizamos

ANEXO 2: SISTEMA BASADO IA - EV DE RESULTADOS.IPYNB

Detalle de código realizado en phyton.

Importaciones

```
import arff
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import RobustScaler
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import RobustScaler
from sklearn.impute import SimpleImputer
from sklearn.base import BaseEstimator, TransformerMixin
```

Funciones auxiliaries

```
def load_kdd_dataset(data_path):
    """Lectura del conjunto de datos NSL-KDD."""
    with open(data_path, 'r') as train_set:
        dataset = arff.load(train_set)
    attributes = [attr[0] for attr in dataset["attributes"]]
    return pd.DataFrame(dataset["data"], columns=attributes)

def train_val_test_split(df, rstate=42, shuffle=True, stratify=None):
    strat = df[stratify] if stratify else None
    train_set, test_set = train_test_split(
        df, test_size=0.4, random_state=rstate, shuffle=shuffle, stratify=strat)
    strat = test_set[stratify] if stratify else None
    val_set, test_set = train_test_split(
        test_set, test_size=0.5, random_state=rstate, shuffle=shuffle, stratify=strat)
    return (train_set, val_set, test_set)

num_pipeline = Pipeline([
        ('imputer', SimpleImputer(strategy="median")),
        ('rbst_scaler', RobustScaler()),
    ])

# Transormador para codificar únicamente las columnas categoricas y devolver un df
class CustomOneHotEncoder(BaseEstimator, TransformerMixin):
    def __init__(self):
        self._oh = OneHotEncoder(sparse=False)
        self._columns = None
    def fit(self, X, y=None):
        X_cat = X.select_dtypes(include=['object'])
        self._columns = pd.get_dummies(X_cat).columns
        self._oh.fit(X_cat)
        return self
    def transform(self, X, y=None):
```

```
        X_copy = X.copy()
        X_cat = X_copy.select_dtypes(include=['object'])
        X_num = X_copy.select_dtypes(exclude=['object'])
        X_cat_oh = self._oh.transform(X_cat)
        X_cat_oh = pd.DataFrame(X_cat_oh,
                                columns=self._columns,
                                index=X_copy.index)
        X_copy.drop(list(X_cat), axis=1, inplace=True)
        return X_copy.join(X_cat_oh)


# Transofrmador que prepara todo el conjunto de datos llamando pipelines y transformadores personalizados
class DataFramePreparer(BaseEstimator, TransformerMixin):
    def __init__(self):
        self._full_pipeline = None
        self._columns = None
    def fit(self, X, y=None):
        num_attribs = list(X.select_dtypes(exclude=['object']))
        cat_attribs = list(X.select_dtypes(include=['object']))
        self._full_pipeline = ColumnTransformer([
                ("num", num_pipeline, num_attribs),
                ("cat", CustomOneHotEncoder(), cat_attribs),
        ])
        self._full_pipeline.fit(X)
        self._columns = pd.get_dummies(X).columns
        return self
    def transform(self, X, y=None):
        X_copy = X.copy()
        X_prep = self._full_pipeline.transform(X_copy)
        return pd.DataFrame(X_prep,
                            columns=self._columns,
                            index=X_copy.index)


df = load_kdd_dataset(“datasets/NSL-KDD/KDDTrain+.arff”)
```

```
def clean_dataset(df):
    """Limpieza y preprocesamiento del conjunto de datos."""
    # Eliminar valores nulos
    df = df.dropna()

    # Eliminar duplicados
    df = df.drop_duplicates()

    # Convertir tipos de datos apropiados
    numeric_columns = df.select_dtypes(include=['float64', 'int64']).columns
    for col in numeric_columns:
        df[col] = pd.to_numeric(df[col], errors='coerce')

    # Eliminar valores atípicos usando IQR
    for column in numeric_columns:
        Q1 = df[column].quantile(0.25)
        Q3 = df[column].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR
        df = df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]
    # Normalización de nombres de columnas
    df.columns = df.columns.str.lower().str.replace(' ', '_')

    return df
```

```
# Aplicar limpieza después de cargar el dataset
df = load_kdd_dataset("datasets/NSL-KDD/KDDTrain+.arff")
df = clean_dataset(df)
# Verificar resultados de la limpieza
print("Registros después de la limpieza:", len(df))
print("Columnas con valores nulos:", df.isnull().sum().sum())
print("Registros duplicados:", df.duplicated().sum())
df.head(10)
# Division del conjunto en los diferentes subconjuntos
train_set, val_set, test_set = train_val_test_split(df)

print("Longitud del Training Set:", len(train_set))
print("Longitud del Validation Set:", len(val_set))
print("Longitud del Test Set:", len(test_set))
Longitud del Training Set: 75583
Longitud del Validation Set: 25195
Longitud del Test Set: 25195

# Conjunto de datos general
X_df = df.drop("class", axis=1)
y_df = df["class"].copy()

# Conjunto de datos de entrenamiento
X_train = train_set.drop("class", axis=1)
y_train = train_set["class"].copy()

# Conjunto de datos de validación
X_val = val_set.drop("class", axis=1)
y_val = val_set["class"].copy()

# Conjunto de datos de pruebas
```

```
X_test = test_set.drop("class", axis=1)
y_test = test_set["class"].copy()
```

Preparación del conjunto de datos

```
# Instanciamos nuestro transformador personalizado
data_preparer = DataFramePreparer()
# Hacemos el fit con el conjunto de datos general para que adquiera todos los valores posibles
data_preparer.fit(X_df)
# Transformamos el subconjunto de datos de entrenamiento
X_train_prep = data_preparer.transform(X_train)
X_train.head(5)
X_train_prep.head(5)
# Transformamos el subconjunto de datos de validacion
X_val_prep = data_preparer.transform(X_val)
```

Entrenamiento de un algoritmo

La instanciación de un algoritmo de Machine Learning utilizando Sklearn se realiza utilizando los métodos expuestos por la API de sklearn tal y como se ha presentado anteriormente**.**

```
# Entrenamos un algoritmo basado en regresión logística
from sklearn.linear_model import LogisticRegression
clf = LogisticRegression(max_iter=5000)
clf.fit(X_train_prep, y_train)
```

Predicción de nuevos ejemplos

Realizamos una predicción con el modelo generado anteriormente tras el entrenamiento del algoritmo de Regresión Logística. Utilizamos el subconjunto de validación.

```
y_pred = clf.predict(X_val_prep)
```

Matriz de Confusión

```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_val, y_pred)
```

```
array([[11415,   459],
       [  252, 13069]])
```

```
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(clf, X_val_prep, y_val, values_format='3g')
```

<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1a1ce40b10>

Métricas derivadas de la matriz de confusión

Precisión

```
from sklearn.metrics import precision_score
print("Precisión:", precision_score(y_val, y_pred, pos_label='anomaly'))
```

Precisión: 0.978400617125225

Recall

```
from sklearn.metrics import recall_score
print("Recall:", recall_score(y_val, y_pred, pos_label='anomaly'))
```

Recall: 0.961344113188479

F1 Score

```
from sklearn.metrics import f1_score
print("F1 score:", f1_score(y_val, y_pred, pos_label='anomaly'))
```

F1 score: 0.9697973747929145

Curvas ROC y PR

Curva ROC

```
from sklearn.metrics import plot_roc_curve
plot_roc_curve(clf, X_val_prep, y_val)
```

<sklearn.metrics._plot.roc_curve.RocCurveDisplay at 0x1a22efba10>

Curva PR

```
from sklearn.metrics import plot_precision_recall_curve
plot_precision_recall_curve(clf, X_val_prep, y_val)
```

<sklearn.metrics._plot.precision_recall_curve.PrecisionRecallDisplay at 0x1a23285bd0>

Evaluación del modelo con el conjunto de datos de pruebas

```
# Transformamos el subconjunto de datos de validacion
X_test_prep = data_preparer.transform(X_test)
y_pred = clf.predict(X_test_prep)
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(clf, X_test_prep, y_test, values_format='3g')
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1a2377da50>
print("F1 score:", f1_score(y_test, y_pred, pos_label='anomaly'))
F1 score: 0.9669233085293991
```