ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.436

Effectiveness of Machine Learning tools for detecting phishing attacks: A systematic review

Campos Echavigurin, Felix Alejandro[1]; Morales Andía, Ciliani Iduví[2]; Ráez Martínez, Haymín Teresa[3]; Martínez Navarro, José Antonio[4]
Universidad Tecnológica del Perú, Perú, *u20204498@utp.edu.pe*, *u18215798@utp.edu.pe, c19240@utp.edu.pe, c18868@utp.edu.pe*

***Abstract*—In this new technological era, there are many threats through a simple internet search, and we are daily exposed to them. A lot of people don't know the risks and leads them to be a potential victim causing them serious consequences. The phishing is one of the most common threats that scams people all over the world. Due to that, this investigation wants to examinate the literatures existing about the based machine learning solutions for the phishing attacks. After the recompilation, that ended on 464 articles original from Scopus. But now the investigation has 30 open access articles that were carefully selected with the inclusion and exclusion rules to guarantee that these ones are closely related to the investigated subject. The results showed that the solutions have 90% or superior precision in most of the cases. With this information, it concluded that the machine learning techniques are very effective and a good choice to affront the problem. However, there are still some aspects that has to be considered before putting it on practice.**

***Keywords*— Machine learning, phishing, cybersecurity, cybersecurity threats, social engineering.**

Eficacia de herramientas con Machine Learning para la detección de ataques phishing: Una revisión sistemática

Campos Echavigurin, Felix Alejandro[1]; Morales Andía, Ciliani Iduví[2]; Ráez Martínez, Haymín Teresa[3]; Martínez Navarro, José Antonio[4]
Universidad Tecnológica del Perú, Perú, *u20204498@utp.edu.pe, u18215798@utp.edu.pe, c19240@utp.edu.pe, c18868@utp.edu.pe*

Resumen-En esta nueva era tecnológica, existen muchas amenazas a través de una simple búsqueda en Internet y diariamente estamos expuestos a ellas. Mucha gente desconoce los riesgos y los lleva a ser víctimas potenciales causándoles graves consecuencias. El phishing es una de las amenazas más comunes que estafa a personas de todo el mundo. Por eso, esta investigación pretende examinar la literatura existente sobre las soluciones basadas en machine learning para los ataques de phishing. Después de la recopilación, que terminó en 464 artículos originales de Scopus, pero ahora la investigación tiene 30 artículos de acceso abierto que fueron cuidadosamente seleccionados con las reglas de inclusión y exclusión para garantizar que estos están estrechamente relacionados con el tema investigado. Los resultados mostraron que las soluciones tienen una precisión del 90% o superior en la mayoría de los casos. Con estos datos, se concluyó que las técnicas de aprendizaje automático son muy eficaces y una buena opción para afrontar el problema. Sin embargo, todavía hay algunos aspectos que deben tenerse en cuenta antes de ponerlo en práctica.

Palabras clave- Machine learning, phishing, ciberseguridad, amenazas a la ciberseguridad, ingeniería social.

INTRODUCCIÓN

Hoy en día, en muchas partes del mundo se afronta un desafío crítico que ha traído consigo la modernización y rápida evolución de la tecnología: la ciberdelincuencia. Los criminales han conseguido explotar los novedosos avances desarrollando distintas metodologías de ataque como virus, ransomware, troyanos, ataques phishing, spyware, entre otros; que infectan dispositivos con el fin de dejarlos dañados, inoperables, secuestrados o vigilados.

Dentro de las más recurrentes se tiene al phishing que se considera como una de las amenazas de seguridad más peligrosas y severas dentro del campo de la ciberseguridad en los tiempos actuales [1]. El phishing es una práctica de la ingeniería social en la que el delincuente recopila información confidencial de la víctima como datos personales, detalles de sus tarjetas de crédito y demás, suplantando a una entidad legítima para engañar al usuario [2,3]. Este problema afecta sin discriminación tanto a personas naturales, como a compañías y/o organizaciones de manera global [4]. Un estudio en 2018 muestra que las organizaciones gubernamentales estadunidenses han sido uno de los blancos principales de los atacantes, incrementado su porcentaje de incidencia en un 40% en relación con años anteriores [3]. Si bien se considera un asunto relativamente reciente, su primer uso data del año 1996 y desde entonces se ha convertido en una de las formas de fraude más severas [5]. Esta cuestión ha cobrado aún más importancia en los últimos años con el inicio de la pandemia suscitada por el coronavirus (SARS-CoV-2). La coyuntura forzó a las entidades a digitalizar sus servicios implementando sitios web o aplicativos móviles en los cuales los usuarios pueden realizar sus operaciones; la modalidad de trabajo remoto agarró más popularidad y el uso del correo electrónico se volvió fundamental para la comunicación entre colaboradores [6]. Los delincuentes informáticos vieron en ello una enorme oportunidad y, a través del uso de URL maliciosas, robaron enormes cantidades de datos. Tan solo en julio de 2021 se registraron 260,642 casos de ataques phishing por correo electrónico [7].

A partir de ello, se contempló el uso de machine learning para la mitigación de casos de ataques de phishing en la red. El aprendizaje automático es un sector de la inteligencia artificial que consiste en el adiestramiento de las máquinas para la búsqueda de patrones en cantidades masivas de datos y la elaboración de predicciones más precisas. Estos modelos son excelentes para la predicción de casos de phishing [4], dado que los algoritmos de ML permiten un análisis con un alto nivel de detalle contra los links maliciosos y proporcionan una respuesta rápida, casi intuitiva ante ello [8]. Debido a lo anteriormente mencionado, se nota cómo estos ataques phishing son una gran amenaza para la seguridad porque mantener la confidencialidad e integridad de la privacidad de los usuarios es un gran desafío [9, 10]. Desafortunadamente, billones de cibernautas han sido diariamente expuestos a páginas web fraudulentas que solicitan información confidencial [9]. En la actualidad, esto ha ido en aumento, provocando grandes pérdidas de dinero, privación de documentos o la usurpación de identidad entre las más resaltantes. Para nuestra fortuna, existe la ciberseguridad. Esta rama asegura que toda nuestra información esté protegida de forma inmediata contra cualquier amenaza en la red que pueda comprometer estos datos [8]. En suma, contamos con las tecnologías avanzadas las cuales facilitan las tareas del día a día. Una de ellas es el ML, el cual a través del tiempo ha demostrado resultados fascinantes para las organizaciones y

personas [8]. Las técnicas de ML son requeridas para combatir este tipo de ataques [11]. Estos sistemas basados en aprendizaje automático proporcionan un enfoque de seguridad eficaz para los desafíos de seguridad cibernética, ayudando a investigar y detectar amenazas potenciales. En la actualidad, se cuenta con diseños o modelos que muestran hasta un 90% de eficacia en la detección de links maliciosos [12]. Por tanto, Podemos Confiar en que estas tecnologías proporcionan un alto nivel de protección.

Por ende, este estudio examinará la literatura existente para informar a los lectores sobre estas estafas en línea, así como también revisará el estado actual de las técnicas de aprendizaje automático, particularmente aquellas enfocadas en la detección de phishing. El objetivo es analizar el machine learning para la detección de ataques phishing.

Ante lo expuesto, queda evidenciado el importante rol que cumple el machine learning en la defensa contra los piratas informáticos [5] pues ha demostrado resultados prometedores de la aplicación de sus métodos en el campo de la ciberseguridad [14]. Ahondar en las bases de esta disciplina abrirá grandes posibilidades de solución frente otras situaciones críticas de la seguridad de la información. La creación de herramientas basadas en este enfoque ayudará a reforzar la protección de información sensible y prevenir a los usuarios de ser víctimas potenciales de robo. Asimismo, se cuenta con un gran repositorio de revisiones previas en idioma inglés que estudian los temas mencionados, pero el contenido para la comunidad hispanohablante se encuentra muy limitado. Por ello, el desarrollo de esta RSL contribuiría significativamente en expandir la información para revisiones futuras

El trabajo es relevante y aborda un tema actual con un enfoque tecnológico prometedor, pues aborda un problema crítico dentro del ámbito de la ciberseguridad: el phishing y su impacto global. Se presenta una perspectiva bien fundamentada sobre cómo el aprendizaje automático (ML) puede ser una solución eficaz para detectar y mitigar estos ataques.

I. METODOLOGÍA

La revisión sistemática realizada para este RSL se basó en la búsqueda exhaustiva de fuentes en Scopus, utilizando la estrategia PICO y variantes para recopilar estudios pertinentes a las preguntas de investigación y soluciones propuestas. La estrategia PRISMA se empleó para seleccionar los trabajos incluidos en la revisión.

A. Formulario de PICO

La revisión sistemática realizada para este RSL se basó en la búsqueda exhaustiva de fuentes en Scopus, utilizando la estrategia PICO y variantes para recopilar estudios pertinentes a las preguntas de investigación y soluciones propuestas. La estrategia PRISMA se empleó para seleccionar los trabajos incluidos en la revisión.

1) Identificación de componentes PICO

A partir de esto se especificaron los componentes del PICO, es decir: población, intervención, comparación y resultado.

Luego de la identificación de cada ítem de PICO, se propuso la pregunta de indagación en el formato planteado.

¿Qué soluciones con ML existen en la actualidad que presenten una alta eficacia en la detección de ataques phishing comparados con aquellas que no usan ML a nivel mundial durante el año 2024?

Para examinar estos artículos, las preguntas generales formuladas se desglosaron en preguntas específicas de los componentes de PICO. Por otro lado, se continuó con la búsqueda de palabras clave o palabras principales que correspondan a la pregunta formulada y contengan conexión o sincronicidad con el tema de investigación, posteriormente se incorporaron conexiones OR y el uso de comillas ("") para términos compuestos.

Esta información se muestra en la Tabla I.

TABLA I
COMPONENTES PICO PARA LA INVESTIGACIÓN

P	***Problema/ población***	***Phishing***	***¿Como se detectan los ataques de phishing?***	***Phishing, cyberattack, email, fraud, virus, malignant program***	***Phishing, cyberattack, email, fraud, virus, malignant program***
I	Intrevención	Soluciones con machines learnimg para la detección de ataques phishing	¿Qué soluciones con ML para la detección de phishing existen en la actualidad?	ML, machine learning, cyber security, security network, IA artificial	ML or "machine learning" or "cyber security" "security network" or "IA artificial"
C	Comparación	Métodos, soluciones o propuestas de ML para la detección del phishing	¿Qué tan eficaz han resultado estas soluciones en otras que no usan ML?	Solucions, machine, learning, ML methods	Solucions or "machine, learning" or ML or methods
O	Resultados	Efectividad de las soluciones con ML para la detección del phishing	¿Qué conclusiones se han obtenido de estas soluciones y cuál es su potencial?	ML, machine learning efectivity	ML or "machine learning" or efectivity
C	Contexto	A nivel mundial	¿Quiénes han sido afectado por este tipo de ataque?	Global, world, around the world	Global or world or "around the world"

T	Tiempo	2024	¿En qué año está situado esta problemática?	Solutions, machine learning, ML, methods, cyber security	Solutions or "machine learning" or ML or methods or "cyber security"

2) Sintaxis de la fórmula PICO

Finalmente, las palabras clave de cada pregunta se relacionan mediante el operador AND. Se utilizan OR y AND para refinar los resultados en una ecuación de búsqueda que muestra artículos relacionados con la investigación.

TABLA II
ECUACIÓN DE BÚSQUEDA DE PICO CON LAS PALABRAS CLAVES

Ecuación de búsqueda sin criterios de inclusión y exclusión
(TITLE-ABS-KEY (phishing OR "Cyber security threats" OR cyberattack OR "email scam" OR "mail-related fraud" OR virus OR "malignant program") AND TITLE-ABS-KEY (ml OR "machine learning" OR "cyber security" OR security OR "network security" OR ia OR "artificial intelligence" OR "anti-phishing") AND TITLE-ABS-KEY (solutions OR "machine learning" OR ml OR methods) AND TITLE-ABS-KEY (ml OR "machine learning" OR efectivity) AND TITLE-ABS-KEY (global OR world OR "around the world"))

Luego de formular la ecuación, se implementaron filtros específicos con el fin de reducir la cantidad de artículos y obtener documentos relacionados con la investigación. Existen los siguientes filtros:

- Para el idioma de los artículos: solo se aceptó aquellos que estén en español e inglés.
- El intervalo de tiempo debe estar entre 2019 y 2024
- Para los tipos de publicación: solo se usaron aquellas de tipo article, conference paper y review.
- Todos los artículos usados deben ser de libre acceso.

Una vez establecida la ecuación, se la introdujo en "Scopus" para visualizar los artículos encontrados de la búsqueda realizada. El primer resultado arrojo un total de 3912 artículos, esta cantidad es demasiado alta por lo cual, se aplicaron los primeros filtros. Después del primer filtrado nos quedaron 464 artículos. Por último, se aplicaron los criterios y esta búsqueda dio como resultado 30 artículos.

3) Criterios de búsqueda de empleados

Criterios de inclusión

- I1: Los estudios deben analizar, describir y aplicar métodos tecnológicos relacionados con machine learning y phishing.
- I2: Los estudios deben ser desarrollados en cualquier entorno.
- I3: Los estudios incluidos deben reportar propuestas, estadísticas y métodos para la detección de ataques phishing con machine learning.
- I4: Los estudios incluidos deben usar tecnologías modernas o que aun sigan vigentes.

Criterios de exclusión

- **E1:** Publicaciones en idiomas que no sean español e inglés.
- **E2:** Artículos publicados fuera del rango de años entre 2019 – 2024.
- **E3:** Estudios desarrollados que no presenten mejoras o información estadísticas sobre la machine learning dentro del entorno de la ciber seguridad y phishing.
- **E4:** Documentos de extensión menor a 5 páginas.
- **E5:** Documentos que hablen de métodos y propuestas que no utilicen ML

B. Prisma

En la selección de los artículos de la RSL, se utilizó la metodología PRISMA para seleccionar los artículos de la RSL, dividida en dos secciones: identificación de estudios y creación de un esquema matricial para recolectar y administrar datos de investigaciones.

1) Proceso de reconocimiento

La selección de artículos se realizó en cinco etapas: determinación de trabajos, criterios de duplicación, elegibilidad, elección y sesgo. Se utilizó netamente la web de Scopus por su respaldo y prestigio, que alberga una amplia variedad de materias científicas y garantiza la confiabilidad de los artículos.

2) Eliminar artículos duplicados

Durante esta fase de eliminación, no se encontraron artículos duplicados y todos se recuperaron del banco de datos de Scopus.

3) Proceso de elegibilidad

Este proceso se ha dividido en dos puntos: Para el primer punto, se excluyeron los artículos que no obedecían a la opción de acceso abierto y, para el siguiente punto, los artículos que no tenían un enfoque central relacionado con tema de machine learning para la detección de ataques phishing.

4) Proceso de selección

En la etapa final, se optaron por aquellos artículos que no seguían los criterios establecidos, limitando las opciones restantes a artículos de 5 páginas o más de extensión.

Finalmente, quedan 30 artículos que pasan a formar parte de la RSL.

PRISMA ha sido útil en esta investigación de seguridad informática, inteligencia artificial, desarrollo de software, ya que permitió sintetizar grandes volúmenes de información de manera estructurada y basada en evidencia.

Representa visualmente el proceso de selección de estudios, mostrando el número de registros identificados, excluidos y finalmente incluidos en la revisión.

En el siguiente gráfico (Fig. 1) se muestra el diagrama PRISMA.

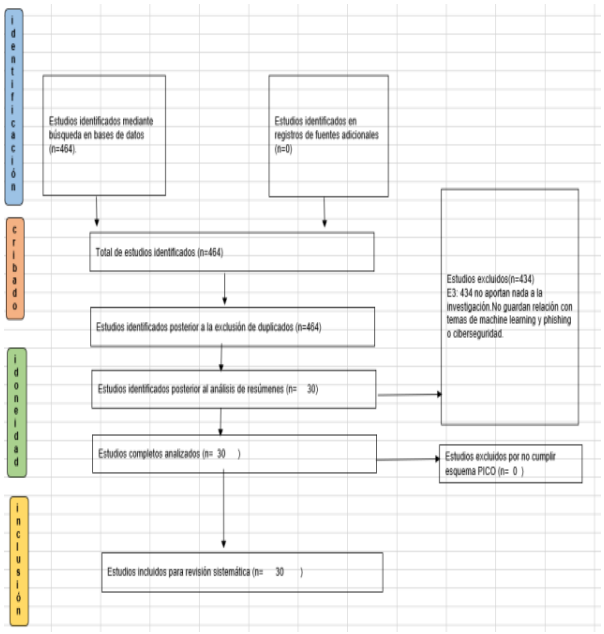

Fig. 1 Diagrama de PRSIMA

Este diagrama muestra el número de estudios identificados, evaluados y excluidos, así como las razones de exclusión durante cada fase de la revisión sistemática.

II. Resultados

El siguiente informe resume los datos extraídos de los distintos artículos analizados en las revisiones sistemáticas. La información está organizada en un formato tabular que proporciona información detallada sobre los resultados de la encuesta. El objetivo de este apartado es explicar las soluciones con machine learning que existen en la actualidad y en especial de aquellas que son especializadas en los campos de la detección de ataques phishing.

A. Ítems para tabla de extracción de datos

1) RQ1: ¿Cómo se detectan los ataques de phishing?

Este estudio se investigó las múltiples opciones con las que se cuenta actualmente para la detección de ataques phishing a través de varios métodos que implementan ML. Toda la información es respaldada gracias a la literatura revisada previamente que hablan de métodos convencionales, hasta otros los cuales son una evolución de estos y realizan un trabajo de mayor complejidad. Estos sistemas se basan en funciones o algoritmos capaces de analizar los datos y etiquetarlos como maliciosos o legítimos. Sin embargo, para comprender este campo de la tecnología se ha iniciado una investigación sobre el concepto de ML para la detección de ataques phishing. Por lo tanto, esta sección extrae información acerca de este tema dentro de los 30 artículos analizados para brindar una mejor comprensión del área en cuestión. A su vez, esta información ayuda a comprender los conceptos o ideas contenidas en el artículo relacionados con esta investigación. La explicación anterior se detalla a continuación:

CATEGORÍAS: Métodos de detección usados en los estudios analizados.

Estas investigaciones organizan las opciones para comprender los enfoques detallados relacionados con la detección de ataques phishing con ML que los autores desean expresar en sus artículos. La Figura 2 muestra los métodos más y menos utilizadas por los autores para este tema de investigación. La información recopilada se dividió en 2 definiciones tipos, resultando en los siguientes porcentajes: Por un lado, el 86,67 % de los autores hacen referencia a una "Técnica o método" [13], [14], [15], [17], [18], [19], [20], [3], [22], [23], [24], [25], [26], [28], [29], [31], [8], [32], [33], [34], [35], [36], [37], [38], [39], [11], por otro lado, el 13,33% hace más alusión a un "Aplicativo" [16], [21], [27], [30].

A continuación, los gráficos de la revisión.

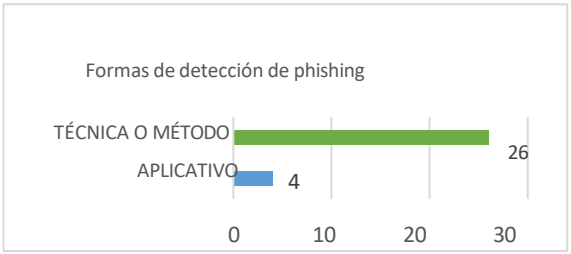

Fig. 2 Categorías de la RQ1

Esta gráfica muestra las tendencias y las formas de detección de phishing que se encontraron durante la RS.

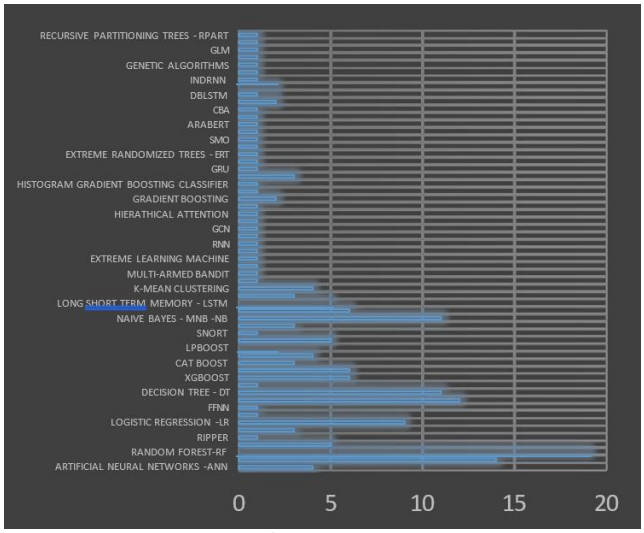

Fig. 2.1 Técnicas de machine learning
Este gráfica muestra las técnicas de machine learning y la frecuencia de uso de por parte de los autores en sus respectivas investigaciones.

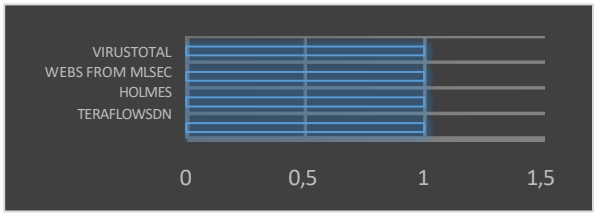

Fig. 2.1 Aplicaciones que usan machine learning
Este gráfica muestra cuales han sido las aplicaciones que usaron machine learning para su lógica encontradas durante la RS.

2) RQ2: ¿Qué soluciones con ML para la detección de phishing existen en la actualidad?

Esta sección proporciona información sobre los tipos de solución para la detección de ataques phishing con ML. De las 30 RSL, se recogieron los datos correspondientes para el tema. Estos brindan una idea acerca de cuáles serían algunas de los remedios para estos ataques informáticos. El siguiente artículo proporciona información como se muestra en la Figura 3. Del análisis se desprenden 3 categorías: En primer lugar, el 43,33% son "Análisis" [13], [15], [19], [21], [24], [30], [31], [8], [32], [37], [38], [39], [11], del tema y cómo podemos evitarlo. En segundo lugar, el otro 43,33 % referencian un "Modelo" [14], [17], [20], [3], [22], [23], [25], [26], [28], [33], [34], [35], el cual se podría implementar para hacer el trabajo de detección. Por último, el 13,33 % restante es "Modelo/Análisis" [16], [18], [29], [36], porque una mezcla de ambas categorías mencionadas anteriormente, ya que realiza un análisis del tema y propone un modelo basado en la investigación realizada

CATEGORÍAS: Soluciones referenciadas por los autores En la Figura 3, se muestran a mayores detalles las soluciones.

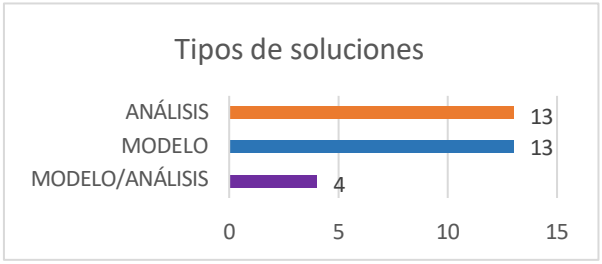

Fig. 3 Categorías de las RQ2
La gráfica muestra cuales son los tipos de soluciones más populares para el uso de machine learning en la detección de phishing.

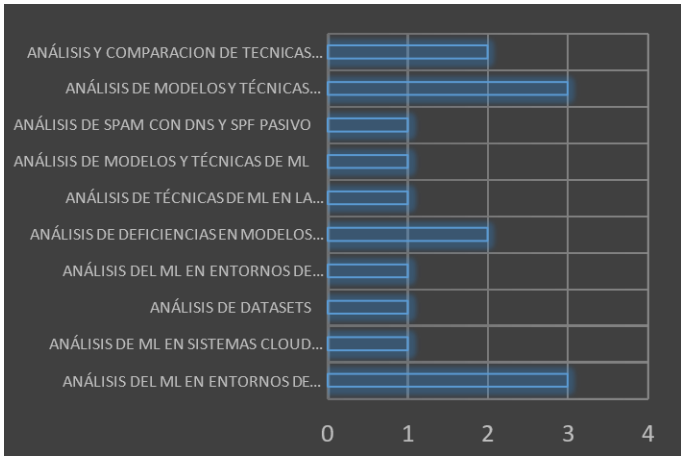

Fig. 3.1 Métodos usados en la detección de phishing
Esta gráfica nos dice cuáles son aquellos métodos/técnicas usados para la integración de un modelo de machine learning en un aplicativo.

3) RQ3: ¿Qué tan eficaz han resultado estas soluciones en comparación a otras que no usan ML?

Para esta pregunta, los datos obtenidos por cada autor fueron clasificados según el porcentaje de eficacia y estandarizados en rangos a los cuales se le asignaron etiquetas para una mejor comprensión y visualización. Por lo tanto, esta información se resume en 4 términos con la finalidad de vislumbrar cuan buenos son los resultados en comparación con otros. Después de identificar los elementos necesarios, la información se organizó en un gráfico de barras tal y como se muestra en la Figura 4. El 20 % de las RSL, "No presenta" [13], [15], [18], [21], [8], [38], un porcentaje de eficacia especifico, pero aportan conocimiento en general que ha sido beneficioso para el campo de la detección de ataques phishing con ML. El 6,67 % son aquellas que presentan una "Eficacia estándar" [3], [27], que se encuentra entre el 70% y 79% de exactitud y son similares a las que no usan tecnologías con IA. El otro 20 % corresponde a las soluciones con "Eficacia buena" [26], [30], [32], [34], [36], [11], las cuales

presentan una precisión promedio del 80% al 89%. Por último, el 53,33% presentan una "Eficacia excelente" [14], [16], [17], [19], [20], [22], [23], [24], [25], [28], [29], [31], [33], [35], [37], [39], estas tienen el más alto porcentaje de exactitud al momento de realizar un análisis en búsqueda de un posible ataque phishing, ya que se muestra un rango de entre 90% y 100%, lo que asegura una mayor protección contra estos problemas.

CATEGORÍAS: Datos recopilados de la eficacia de estas soluciones.

En la Figura 4, se visualizan los porcentajes de eficacia de estas soluciones con ML

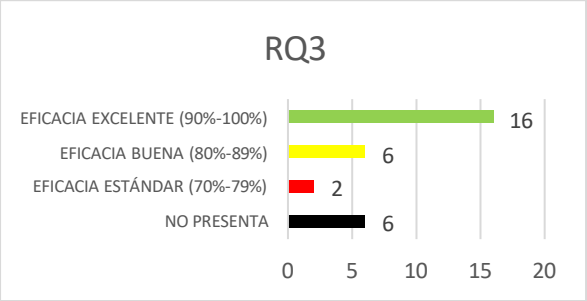

Fig. 4 Categorías de la RQ3
Gráfica de rangos de eficacia del uso de machine learning para la detección de phishing.

4) *RQ4: ¿Qué conclusiones se han obtenido de estas soluciones y cuál es su potencial?*

En este apartado se proporciona información acerca de las conclusiones obtenidas por los autores en base a las alternativas propuestas para la detección de amenazas phishing. Para ello se evaluó minuciosamente la efectividad y acierto de las herramientas implementadas en diferentes casos. A partir de este análisis, se categorizó el potencial que muestran estas soluciones para su aplicación en situaciones futuras. De un total de 30 artículos revisados, el 43.33% de ellos muestra un "Potencial alto" [14], [16], [17], [19], [20], [22], [23], [24], [25], [28], [29], [31], [33], [35], [37], [11]. Con un empate, un 20% de los artículos evidencian un "Potencial medio" [26], [30], [32], [34], [36], [39], y el otro 20% de ellos "No indica" claramente el potencial de su solución [13], [15], [18], [21], [8], [38]. Por último, solo el 6.67% de las fuentes muestra un "Potencial bajo" [3], [27].

CATEGORÍAS: Potencial de las soluciones propuestas.

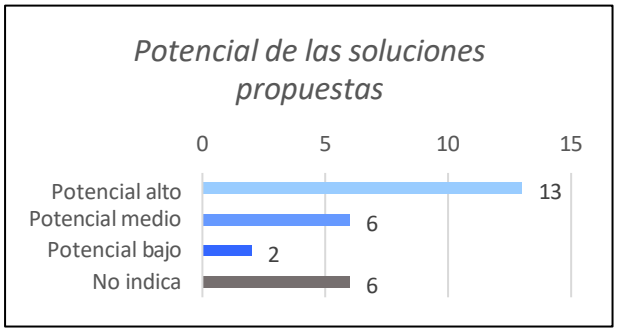

Fig. 5 Categorías de la RQ4
Esta gráfica mide el potencial de las soluciones y las clasifica en cuatro niveles.

5) *RQ5: ¿Quiénes han sido los afectados por este tipo de ataques?*

En esta sección se proporciona información acerca de las víctimas de los ciberdelincuentes y su práctica más utilizada: el phishing. Para ello, se identificó en un total de 30 revisiones previas quiénes fueron los principales afectados por estos actos criminales en la red, obteniendo así la siguiente categorización: En primer lugar, un 40% de las fuentes recopiladas mencionan que las "Personas naturales" [14], [17], [22], [25], [28], [31], [32], [33], [34], [35], [36], [11] son un blanco fácil para los atacantes. En segundo lugar, un 23.33% que indican que las "Organizaciones" son su objetivo principal [13], [18], [20], [24], [26], [27], [38]. Un 20% de los estudios revisados expresan que "Ambos", tanto organizaciones como personas naturales, son blanco de este tipo de ataques [15], [3], [23], [29], [8], [39]. Asimismo, un 13.33% de los estudios "No precisa" de manera clara quiénes son las principales víctimas [16], [21], [30], [37]. Por último, solo el 3.33% de estos menciona que los objetivos de los piratas de la red son "Otros" [19].

CATEGORÍAS: Afectados por ataques phishing.

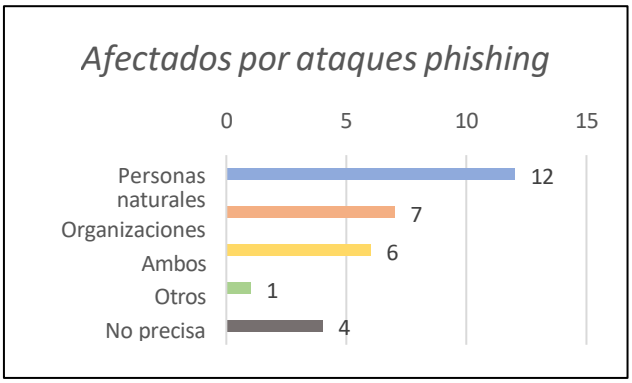

Fig. 6 Categorías de la RQ6
Esta grafica muestran la cantidad y cuales son los afectados de este tipo de ataques.

6) RQ6: ¿En qué año está situada la problemática planteada?

Para este estudio, se analizó un total de 30 artículos que guarden estrecha relación al tema de machine learning o afines y abarquen la problemática planteada. Para ello, fue requisito principal que las revisiones previas se hayan realizado a partir del año 2019 en adelante para garantizar que las soluciones propuestas sean las más actuales. Para el siguiente análisis, se clasificaron los estudios de acuerdo con sus años de publicación, obteniendo lo que se muestra en el gráfico a continuación

CATEGORÍAS: Años de los artículos utilizados en la RSL.

Fig. 7 Categorías de la RQ6
La gráfica muestra en que periodo de años están las investigaciones usadas para la RS y cuantas se usaron de cada año.

III. Discusión

Los resultados arrojan que un 86,67% de los estudios, se centran en las técnicas y métodos las cuales son usadas para usadas para entrenar a los modelos y que realicen la tarea de detección de ataques de phishing. El otro 13,33% nos habla de aplicativos que se encuentran y los analizan con la finalidad de revisar su efectividad y como es que funcionan. Es decir, que en la actualidad existen varias formas y maneras para la detección de ataques phishing e incluso aplicativos a los cuales ya se tiene acceso. Esto nos da a entender que las técnicas de ML para este campo están siendo aceptadas e irán ganando cada vez más terreno, logrando que su uso sea esencial y generalizado en al ámbito de la ciberseguridad. Dentro de las soluciones, los resultados muestran que un 43,33% enfocan su estudio en realizar un análisis detallado. Por un lado, algunos muestran cómo se debe abordar ciertos temas con la intención de lograr un mejor modelo que detecte ataques phishing con mayor efectividad y precisión. Por otro lado, los estudios toman un rumbo informativo para los usuarios y difunden información detallada del funcionamiento de los ataques, la gestión de datos o comparaciones de técnicas. El otro 43,33% muestra un modelo detallando todas sus características y su rendimiento al detectar ataques phishing. Por último, el 13.33% analiza las tecnologías actuales y en base a ello propone un modelo al cual se le hacen pruebas y se muestran los resultados para su rendimiento, precisión y eficacia detectando ataques phishing. Esto revela que a pesar de que en la actualidad podemos encontrar varias soluciones que se apoyan en el uso de ML, estas tecnologías aun no alcanzan todo su potencial y hay varias que solo quedan como modelos propuestos por el autor. Cabe destacar, que este es un tema amplio y que este campo está en cambio continuo, por lo tanto, siempre se debe estar a la vanguardia y conocer que soluciones existen. Gracias al análisis de las literaturas, se pudo concluir que el 91,67% de las tecnologías que hacen uso de ML muestran una mejora significativa en todos los aspectos en comparación con aquellas que no hacen uso de esta tecnología. Logrando eficacias de hasta un 99% dentro de las pruebas realizadas por los autores. Se deduce, que estas herramientas son muy útiles y que el ML junto con sus técnicas son esenciales y clave para realizar una mejor detección de ataques phishing, ofreciendo un gran aporte al campo de la ciberseguridad, mejorando las defensas contra amenazas y luchando contra los ataques cibernéticos. Se examinaron las 30 revisiones previas considerados para esta RSL con la finalidad de identificar a las principales víctimas de estas prácticas delincuenciales en la red. Se encuentra que, en la mayoría de los trabajos representando a un considerable 40%, se reconoce a las personas naturales como los más afectados por estos actos en comparación a otras revisiones [13], [18], [20], [24], [26], [27], [38] donde expresan que son las organizaciones el blanco favorito de los ciberdelincuentes. Este notable contraste en los números se atribuye a una mayor concientización que brindan las empresas a sus colaboradores respecto a los peligros en la red y sus consecuencias; además de llevar un mayor refuerzo en su seguridad. Caso contrario en la mayoría de los individuos que desconocen parcial o completamente acerca de este tipo de ataques. Entre las dificultades, por un lado, se denota la baja calidad de datasets [18], esto se refiere a que la mayoría de estos conjuntos de datos no cuentan con una estructura previamente establecida lo cual produce dificultades y retrasos durante adiestramiento de la máquina ya que estos trabajan con datos desorganizados y desestructurados, perdiendo oportunidades de mejora en la detección de phishing. Por otro lado, el alcance del entorno es poco realista, es decir que gran parte de estos modelos

propuestos por los autores realizan sus pruebas dentro de un área o espacio delimitado más no abarcan un entorno completo o una situación real, como la implementación del modelo en una empresa funcional, el uso del software libre para el público o la creación y difusión de un aplicativo.

A pesar de las significativas contribuciones del uso de machine learning en la detección de phishing, todavía existen áreas de mejora. Por ejemplo, la dependencia de datos de calidad, los conjuntos de datos desorganizados o incompletos pueden afectar el desempeño del modelo. Es crucial que se sigan desarrollando y mejorando los conjuntos de datos utilizados para entrenar estos modelos, a fin de mejorar su precisión y robustez. Además, de los desafíos de integración en entornos reales. Como se mencionó, muchos modelos de machine learning se prueban en entornos controlados, lo que limita su aplicabilidad en situaciones del mundo real. Las futuras investigaciones deben enfocarse en la integración de estos modelos en entornos dinámicos y con grandes volúmenes de datos. En suma, la implementación de estos modelos en entornos reales enfrenta desafíos significativos. Entre los principales obstáculos se incluyen:

1) *La interpretabilidad del modelo*: Muchos modelos de machine learning, como los basados en XGBoost, son considerados "cajas negras", lo que dificulta su interpretación y la comprensión de las decisiones tomadas por el sistema. En un entorno empresarial o gubernamental, la falta de transparencia puede generar desconfianza en los usuarios y en los responsables de la toma de decisiones.
2) *Sesgo de los datos*: La calidad de los datos es un factor crítico para el rendimiento del modelo. Los conjuntos de datos de entrenamiento pueden estar sesgados, lo que puede resultar en un modelo que no detecte correctamente todos los tipos de phishing. Este problema se agrava cuando los atacantes modifican sus técnicas para eludir las soluciones basadas en machine learning.
3) *Escalabilidad y adaptación al entorno real*: Muchos de los estudios revisados se realizaron en entornos controlados, sin considerar las variaciones y complejidades que se presentan en situaciones del mundo real. Esto puede afectar la efectividad de los modelos cuando se implementan en sistemas en vivo, como en el caso de empresas que procesan grandes volúmenes de datos. La escalabilidad es, por lo tanto, una barrera importante para la adopción generalizada de estos modelos.

A pesar de ello, las soluciones basadas en machine learning para la detección de phishing pueden jugar un papel fundamental en la mejora de la seguridad digital en diversas áreas:

1) *Políticas públicas*: Los gobiernos pueden adoptar tecnologías basadas en machine learning como parte de sus estrategias de ciberseguridad para proteger tanto a los ciudadanos como a las infraestructuras críticas. El desarrollo de sistemas de alerta temprana que utilicen machine learning puede prevenir que los usuarios accedan a sitios web maliciosos, especialmente si se combinan con campañas de concientización pública sobre los riesgos de phishing.
2) *Empresas*: Las organizaciones pueden integrar modelos de machine learning en sus sistemas de seguridad cibernética para proteger a sus empleados y clientes de ataques de phishing. Esta integración permitiría detectar y bloquear sitios maliciosos en tiempo real, reduciendo el riesgo de fraude y protegiendo la reputación corporativa. Además, las empresas podrían invertir en el entrenamiento de sus empleados para fortalecer la defensa interna contra este tipo de ataques.
3) *Sistemas educativos*: Los sistemas educativos también pueden beneficiarse de estas soluciones. La educación en ciberseguridad debe ser un componente clave en los programas académicos, y las tecnologías de detección basadas en machine learning pueden ser utilizadas para enseñar a estudiantes y personal sobre cómo identificar y prevenir ataques de phishing. Además, las universidades y colegios podrían colaborar con gobiernos y empresas para promover el uso de herramientas avanzadas de ciberseguridad en su comunidad.

IV. Conclusiones

La revisión sistemática de la literatura demuestra que, a pesar de las limitantes se manifiestan a partir de los datos no organizados y los entornos restrictivos, los modelos con machine learning logran sobreponerse ante estas adversidades y demuestran una alta precisión y eficacia en la detección de phishing. Sin embargo, para ponerlo en práctica, es necesario abordar y resolver una serie de problemas, como la interpretación del modelo, la capacidad de actualización, la integración con otros sistemas de seguridad u otros más amplios y la gestión eficiente de materiales actuales y futuros. Las investigaciones futuras deben centrarse en desarrollar modelos más eficientes que sean fáciles de entender y usar, así como en desarrollar técnicas para una mejor integración y adaptación continua en entornos de ciberseguridad que estén cambiando constantemente. Será esencial que los investigadores, profesionales del tema, las industria y los líderes políticos trabajen juntos para mejorar la seguridad en línea y luchar contra el phishing.

Asimismo, es importante reflexionar sobre cómo las soluciones basadas en técnicas de machine learning pueden ser integradas de manera efectiva en políticas públicas, entornos empresariales y sistemas educativos. Desde la perspectiva gubernamental, estas tecnologías pueden incorporarse en estrategias nacionales de ciberseguridad, promoviendo el desarrollo de herramientas automatizadas para proteger servicios digitales y ciudadanos ante amenazas de phishing. En el ámbito empresarial, la integración de modelos de

detección puede fortalecer las infraestructuras de seguridad informática, especialmente en sectores críticos como banca, salud y telecomunicaciones. Finalmente, en el sistema educativo, se pueden desarrollar programas de concientización y capacitación apoyados en estas tecnologías para fomentar una cultura digital segura desde edades tempranas. La combinación de innovación tecnológica y compromiso institucional resulta clave para reducir el impacto real de los ataques de phishing en la sociedad.

REFERENCIAS

[1] Sumitra Das Guptta, Khandaker Tayef Shahriar, H. Alqahtani, D. Alsalman, and I. H. Sarker, "Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques," Annals of data science, vol. 11, no. 1, pp. 217–242, Mar. 2022, doi: https://doi.org/10.1007/s40745-022-00379-8.

[2] R. Jayaraj, A. Pushpalatha, K. Sangeetha, T. Kamaleshwar, S. Udhaya Shree, and Deepa Damodaran, "Intrusion detection based on phishing detection with machine learning," Measurement. Sensors, vol. 31, pp. 101003–101003, Feb. 2024, doi: https://doi.org/10.1016/j.measen.2023.101003.

[3] Z. Fan, W. Li, Kathryn Blackmond Laskey, and K.-C. Chang, "Investigation of Phishing Susceptibility with Explainable Artificial Intelligence," Future internet, vol. 16, no. 1, pp. 31–31, Jan. 2024, doi: https://doi.org/10.3390/fi16010031.

[4] Maria Carla Calzarossa, P. Giudici, and Rasha Zieni, "Explainable machine learning for phishing feature detection," Quality and reliability engineering international, vol. 40, no. 1, pp. 362–373, Jul. 2023, doi: https://doi.org/10.1002/qre.3411.

[5] A. Karim, Mobeen Shahroz, Khabib Mustofa, Samir BrahimBelhaouari, and R. Kumar, "Phishing Detection System Through Hybrid Machine Learning Based on URL," IEEE access, vol. 11, pp. 36805–36822, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3252366.

[6] M. Dewis and T. Viana, "Phish Responder: A Hybrid Machine Learning Approach to Detect Phishing and Spam Emails," Applied system innovation, vol. 5, no. 4, pp. 73–73, Jul. 2022, doi: https://doi.org/10.3390/asi5040073.

[7] Bander Nasser Almousa and Diaa Mohammed Uliyan, "Anti-Spoofing in Medical Employee's Email using Machine Learning Uclassify Algorithm," International journal of advanced computer science and applications/International journal of advanced computer science & applications, vol. 14, no. 7, Jan. 2023, doi: https://doi.org/10.14569/ijacsa.2023.0140727.

[8] A. Alharbi et al., "Analyzing the Impact of Cyber Security Related Attributes for Intrusion Detection Systems," Sustainability, vol. 13, no. 22, pp. 12337–12337, Nov. 2021, doi: https://doi.org/10.3390/su132212337.

[9] A. B. Altamimi et al., "PhishCatcher: Client-Side Defense Against Web Spoofing Attacks Using Machine Learning", IEEE Access, p. 1, 2023. Accedido el 14 de abril de 2024. [En línea]. Disponible: https://doi.org/10.1109/access.2023.3287226.

[10] Baadel y J. Lu, "Data Analytics: Intelligent Anti-Phishing Techniques Based on Machine Learning", J. Inf. & Knowl. Manage., vol. 18, n. º 01, p. 1950005, marzo de 2019. Accedido el 14 de abril de 2024. [En línea]. Disponible: https://doi.org/10.1142/s0219649219500059.

[11] A. Suryan, C. Kumar, M. Mehta, R. Juneja y A. Sinha, "Learning Model for Phishing Website Detection", ICST Trans. Scalable Inf. Syst., p. 163804, julio de 2018. Accedido el 14 de abril de 2024. [En línea]. Disponible: https://doi.org/10.4108/eai.13-7-2018.163804.

[12] K. Omari, "Comparative Study of Machine Learning Algorithms for Phishing Website Detection," International journal of advanced computer science and applications/International journal of advanced computer science & applications, vol. 14, no. 9, Jan. 2023, doi: https://doi.org/10.14569/ijacsa.2023.0140945.

[13] M. Sulaiman, M. Waseem, A. N. Ali, G. Laouini y F. S. Alshammari, "Defense strategies for epidemic cyber security threats: modeling and analysis by using a machine learning approach", *IEEE Access*, p. 1, 2024. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1109/access.2024.3349660.

[14] M. Ghosh, D. Ghosh, R. Halder y J. Chandra, "Investigating the impact of structural and temporal behaviors in ethereum phishing users detection", *Blockchain: Res. Appl.*, p. 100153, julio de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1016/j.bcra.2023.100153.

[15] Y. Shang, "Detection and prevention of cyber defense attacks using machine learning algorithms", *Scalable Comput.: Pract. Experience*, vol. 25, n.º 2, pp. 760–769, febrero de 2024. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.12694/scpe.v25i2.2627.

[16] A. Mozo, A. Karamchandani, L. de la Cal, S. Gómez-Canaval, A. Pastor y L. Gifre, "A machine-learning-based cyberattack detector for a cloud-based SDN controller", *Appl. Sci.*, vol. 13, n.º 8, p. 4914, abril de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.3390/app13084914

[17] N. Nagy *et al.*, "Phishing urls detection using sequential and parallel ML techniques: Comparative analysis", *Sensors*, vol. 23, n. º 7, p. 3467, marzo de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.3390/s23073467

[18] I. Skula y M. Kvet, "A framework for preparing a balanced and comprehensive phishing dataset", *IEEE Access*, p. 1, 2024. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1109/access.2024.3387437

[19] A. S. M. Al-Ruwili y A. M. Mostafa, "Analysis of ransomware impact on android systems using machine learning techniques", *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, n.º 11, 2023. Accedido el 20 de junio de2024. [En línea]. Disponible: https://doi.org/10.14569/ijacsa.2023.0141178

[20] C. Thapa *et al.*, "Evaluation of federated learning in phishing email detection", *Sensors*, vol. 23, n. º 9, p. 4346, abril de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.3390/s23094346

[21] Y. Gao, B. M. Ampel y S. Samtani, "Evading anti-phishing models: A field note documenting an experience in the machine learning security evasion competition 2022", *Digit. Threats: Res. Pract.*, junio de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1145/3603507

[22] R. J. van Geest, G. Cascavilla, J. Hulstijn y N. Zannone, "The applicability of a hybrid framework for automated phishing detection", *Comput. & Secur.*, vol. 139, p. 103736, abril de 2024. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1016/j.cose.2024.103736

[23] L. R. Kalabarige, R. S. Rao, A. Abraham y L. A. Gabralla, "MLSELM:Multi-layer stacked ensemble learning model to detect phishing websites", *IEEE Access*, p. 1, 2022. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1109/access.2022.3194672

[24] P. J. Uppalapati, B. K. Gontla, P. Gundu, S. M. Hussain y K. Narasimharo, "A machine learning approach to identifying phishing websites: A comparative study of classification models and ensemble learning techniques", *ICST Trans. Scalable Inf. Syst.*, junio de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.4108/eetsis.vi.3300

[25] K. Ahammad y M. S. Shaiham, "An approach to detect phishing websites with features selection method and ensemble learning", *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, n.º 8, 2022. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.14569/ijacsa.2022.0130888

[26] A. D. Kulkarni, "Convolution neural networks for phishing detection", *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, n.º 4, 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.14569/ijacsa.2023.0140403

[27] D. Preuveneers, E. Lavens, y W. Joosen, "Applying Machine Learning to Use Security Oracles: A Case study in Virus and Malware Detection", 2022 IEEE European Symposium On Security And Privacy Workshops (EuroS&PW), jun. 2022, Disponible: https://doi.org/10.1109/EuroSPW55150.2022.00030

[28] S. S. Roy, A. I. Awad, L. A. Amare, M. T. Erkihun y M. Anas, "Multimodel phishing URL detection using LSTM, bidirectional LSTM, and GRU models", *Future Internet*, vol. 14, n. º 11, p. 340, noviembre de 2022. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.3390/fi14110340

[29] "Machine learning algorithms for phishing email detection", *J. Logistics, Inform. Service Sci.*, junio de 2023. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.33168/jliss.2023.0217

[30] P. Wu y H. Guo, "Holmes: An efficient and lightweight semantic based anomalous email detector", en *2022 IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Wuhan, China, 9–11 de diciembre de 2022. IEEE, 2022. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1109/trustcom56396.2022.00192

[31] M. Aljabri *et al*., "Detecting malicious urls using machine learning techniques: Review and research directions", *IEEE Access*, p. 1, 2022. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1109/access.2022.3222307

[32] S. Fernandez, M. Korczyński y A. Duda, "Early detection of spam domains with passive DNS and SPF", en *Passive and Active Measurement*. Cham: Springer Int. Publishing, 2022, pp. 30–49. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1007/978-3-030-98785-5_2

[33] Z. B. Siddique, M. A. Khan, I. U. Din, A. Almogren, I. Mohiuddin y S. Nazir, "Machine learning-based detection of spam emails", *Scientific Program.*, vol. 2021, pp. 1–11, diciembre de 2021. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1155/2021/6508784

[34] Y. B. Abushark *et al.*, "Cyber security analysis and evaluation for intrusion detection systems", *Comput., Mater. & Continua*, vol. 72, n.º 1, pp. 1765–1783, 2022. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.32604/cmc.2022.025604

[35] D. Mehanović y J. Kevrić, "Phishing website detection using machine learning classifiers optimized by feature selection", *Trait. Signal*, vol. 37, n.º 4, pp. 563–569, octubre de 2020. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.18280/ts.370403

[36] "Deep learning based-phishing attack detection", *Int. J. Recent Technol. Eng.*, vol. 8, n. º 3, pp. 8428–8432, septiembre de 2019. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.35940/ijrte.c6527.098319

[37] N. Syuhada Selamat y F. H. Mohd Ali, "Comparison of malware detection techniques using machine learning algorithm", *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 16, n.º 1, p. 435, octubre de 2019. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.11591/ijeecs.v16.i1.pp435-440

[38] P. Dixit, R. Kohli, A. Acevedo-Duque, R. R. Gonzalez-Diaz y R. H. Jhaveri, "Comparing and analyzing applications of intelligent techniques in cyberattack detection", *Secur. Communication Netw.*, vol. 2021, pp. 1–23, junio de 2021. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.1155/2021/5561816

[39] A. Kulkarni y L. L., "Phishing websites detection using machine learning", *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, n.º 7, 2019. Accedido el 20 de junio de 2024. [En línea]. Disponible: https://doi.org/10.14569/ijacsa.2019.0100702