UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR

POSGRADOS

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

OPCIÓN DE TITULACIÓN:

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

TEMA:

ESTUDIO DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO UTILIZADAS EN CIBERSEGURIDAD

AUTOR:

WILLIAM ANDRES VILLAVICENCIO BENALCAZAR

DIRECTOR:

RODOLFO XAVIER BOJORQUE CHASI

CUENCA – ECUADOR
2023

Autor:

William Andres Villavicencio Benalcazar
Ingeniero en Sistemas mención Telemática.
Candidato a Magíster en Seguridad de la Información por la Universidad Politécnica Salesiana – Sede Cuenca.
wvillavicenciob@gmail.com

Dirigido por:

Rodolfo Xavier Bojorque Chasi
Ingeniero de Sistemas.
Máster Universitario en Seguridad de las Tecnologías de la Información y Comunicación.
rbojorque@ups.edu.ec

DEDICATORIA

Quiero agradecer como siempre lo he hecho a mi familia, primero a mis padres William X. Villavicencio C. y Ruth K. Benalcazar M., por ser mi mayor inspiración y ejemplo por seguir, por siempre apoyarme en cada paso de mi vida y por jamás dejarme rendir, aunque muchas veces quise.

A mis hermanas Carolina y Ruth que son mis mejores amigas y han estado ayudándome y animándome cada día de mi vida.

A mis abuelos, especialmente, Papa Polo y Mama Lolita, quienes han estado conmigo apoyándome desde que empecé mi vida estudiantil en la universidad y posgrado.

A mis tíos y mi familia, especialmente a mi tía Fanny que siempre me apoyo y estuvo pendiente de mi en todos mis procesos como estudiante.

A la universidad por todos los conocimientos, experiencias y apoyo durante todo ese tiempo.

A mis mejores amigos, especialmente lo que la universidad me puso en el camino, sin ellos no hubiera podido porque no solo fueron una compañía sino un pilar fundamental en mi vida profesional.

Gracias y muchas gracias a todos por ser parte de mi vida.

AGRADECIMIENTO

Quiero agradecer a todas las personas que han hecho posible este logro. A mis padres, hermanas y abuelos por su amor incondicional y su apoyo constante. A mis amigos por su compañía y motivación. Y a todas aquellas personas que han contribuido en el desarrollo de este artículo, incluyendo colegas, mentores y editores. Gracias a todos por su invaluable contribución en mi camino hacia el éxito. ¡Esto es solo el comienzo!

Tabla de Contenido

Contenido

Estudio de técnicas de aprendizaje automático utilizadas en ciberseguridad

Autor(es):

WILLIAM ANDRES VILLAVICENCIO BENALCAZAR

1. Resumen

Con el crecimiento de las redes informáticas, el aumento de los servicios que brindan y la necesidad de mantener la confiabilidad, integridad y disponibilidad de la información transmitida, la seguridad de los sistemas informáticos se ha vuelto cada vez más importante. Por otro lado, los ataques a los sistemas van en aumento, convirtiéndose en un grave problema. Esta afirmación puede ser verificada por el Informe Anual de Seguridad 2014 de Cisco [1] que destaca el aumento de vulnerabilidades, el mayor desde 2000, aprovechando nuevos frentes ofensivos y técnicas innovadoras. El informe también destaca la disminución de la capacidad de las organizaciones para monitorear y proteger sus redes. Además, el 100% de las 30 redes corporativas más grandes del mundo dirigen el tráfico a sitios web que alojan malware, y el 96 % de las redes analizadas dirigen el tráfico a servidores "públicos" "hackeados" [1] , mientras que el 92 % envía tráfico a sitios sin contenido. Los ataques de denegación de servicio distribuido (DDoS), que afectan el tráfico dirigido o generado por sitios web pirateados, pueden paralizar a los proveedores de servicios de Internet, han ido en aumento a lo largo de estos años y no hay evidencia de que esto se vaya a detener, incluso cada día aumenta más su volumen. Los ataques simples que causan un daño manejable han dado paso a actividades de ciberdelincuencia más sofisticadas, patrocinadas y organizadas capaces de causar un daño económico y reputacional significativo para organizaciones públicas y privadas, atacando así la seguridad nacional de cualquier país.

Por otro lado, hay una mayor complejidad de amenazas y soluciones debido al crecimiento exponencial de los dispositivos móviles y los entornos de nube. Las nuevas clases de dispositivos inteligentes y la nueva infraestructura han ampliado el alcance de los atacantes que pueden aprovechar vulnerabilidades imprevistas y defensas inadecuadas. Los ciberdelincuentes han aprendido que aprovechar el

[1] Hackeados: proviene del verbo “hackeo”, el cual consiste en la aplicación de tecnología o conocimientos técnicos para superar alguna clase de problema u obstáculo, teniendo claro esto, hackeado, sería el resultado o la acción que se logra luego del hackeo. [29]

poder de la infraestructura de Internet les brinda más ventajas que simplemente acceder a computadoras o dispositivos individuales. Estos ataques a nivel de infraestructura buscan obtener acceso a los servidores principales que alojan sitios web, servidores de nombres y centros de datos, con el objetivo final de propagar amenazas a una multitud de activos individuales que dependen de estos recursos. Al atacar la infraestructura de Internet, los ciberdelincuentes socavan la confianza en cualquier cosa que dependa de esa infraestructura [1].

Las técnicas de Machine Learning (ML) se utilizan para neutralizar y tratar de contener estas amenazas en la seguridad informática como detección de malware, ingeniería social, seguridad en redes sociales (cyberbullying, incitación al odio o violencia), pruebas de penetración, asegurar y atacar data, detección de intrusos en redes informáticas y actualmente en la era del Big Data se cuenta con acceso a grandes cantidades de información que ha permitido la evolución y sofisticación de muchas técnicas que comprometen la seguridad como leaks de información, fake news, etc. [2] Datos para demostrar que estos problemas están en constante incremento lo podemos encontrar solamente con buscar en Google, esto nos permitirá tener una buena cantidad de referencias.

Según los antecedentes expuestos es de vital importancia ir a la par con los ciberdelicuentes, por lo que, bajo nuestra proyección, la manera más eficaz y concreta de seguir este ritmo, es bajo el manejo de técnicas de aprendizaje automatizado o Machine Learning, en donde las “máquinas” podrían realizar un trabajo de detección y prevención más eficiente que esperar que un humano lo haga.

Palabras clave:

Cyberseguridad, Machine Learning, Deep Learning, seguridad informática.

2. Abstract

With the growth of computer networks, the increase in services they provide, and the need to maintain the reliability, integrity, and availability of transmitted information, the security of computer systems has become increasingly important. On the other hand, attacks on systems are on the rise, becoming a serious problem. This statement can be verified by Cisco's 2014 Annual Security Report [1], which highlights the increase in vulnerabilities, the largest since 2000, taking advantage of new offensive fronts and innovative techniques. The report also highlights the decreasing ability of organizations to monitor and protect their networks. In addition, 100% of the world's 30 largest corporate networks direct traffic to websites hosting malware, and 96% of analyzed networks direct traffic to "hacked" "public" servers, while 92% direct traffic to content-less sites. Distributed denial of service (DDoS) attacks, which affect traffic directed or generated by hacked websites and can paralyze internet service providers, have been increasing over the years and there is no evidence that this will stop, even their volume is increasing every day. Simple attacks that cause manageable damage have given way to more sophisticated, sponsored, and organized cybercrime activities capable of causing significant economic and reputational damage to public and private organizations, thereby attacking the national security of any country.

On the other hand, there is greater complexity of threats and solutions due to the exponential growth of mobile devices and cloud environments. New classes of smart devices and new infrastructure have expanded the scope of attackers who can exploit unforeseen vulnerabilities and inadequate defenses. Cybercriminals have learned that harnessing the power of the Internet infrastructure gives them more advantages than simply accessing individual computers or devices. These infrastructure-level attacks seek to gain access to main servers hosting websites,

name servers, and data centers, with the goal of spreading threats to a multitude of individual assets that depend on these resources. By attacking the Internet infrastructure, cybercriminals undermine trust in anything that depends on that infrastructure [1].

Machine Learning (ML) techniques are used to neutralize and try to contain these threats in computer security, such as malware detection, social engineering, social network security (cyberbullying, incitement to hatred or violence), penetration testing, securing and attacking data, intrusion detection in computer networks, and currently in the era of Big Data, we have access to large amounts of information that has allowed the evolution and sophistication of many techniques that compromise security, such as information leaks, fake news, etc. [2] Data to demonstrate that these problems are constantly increasing can be found just by Googling, which will allow us to have a good number of references.

Based on the background, it is of vital importance to keep pace with cybercriminals, which is why, under our projection, the most effective and concrete way to keep up with this pace is through the use of machine learning techniques, where "machines" could perform more efficient detection and prevention work than waiting for a human to do it.

Keywords:
IT Security, IT, Machine Learning, Deep Learning, Cyber Security

3. Introducción

La llegada de Internet y su masificación han sido uno de los mayores avances tecnológicos de la era moderna. En un período de 21 años, se registró una expansión significativa en el número de usuarios, pasando de 14 millones en 1993 a cerca de 2.900 millones en julio de 2014 [3].

La incorporación de tecnologías por la sociedad ha generado desafíos para el sector público y privado, que han tenido que adaptarse a los cambios para cumplir con las necesidades y demandas del mercado. La rapidez con la que las nuevas tecnologías están evolucionando ha obligado a las organizaciones a estar al día con los avances. La transformación digital de la sociedad, organizaciones públicas y empresas ha traído consigo nuevos riesgos y desafíos, y requiere de un enfoque constante para mantenerse actualizado y protegido ante amenazas.[4].

Es evidente que la transformación digital trae cambios en todo el ámbito social y empresarial, por lo cual debemos estar preparados y que no nos tome por sorpresa que, así como debemos protegernos físicamente de ataques, también debemos hacerlo en el ámbito virtual.

Tal como nos dice Jeimy J. Cano en [5]: En el mundo actual, el aumento en la fuga de información, los ataques informáticos que han comprometido la seguridad, y las vulnerabilidades en tanto en el sector público como privado, muestran un panorama de amenazas y riesgos en el que la información se ha convertido en una herramienta estratégica y táctica que pone en duda la gobernabilidad de organizaciones y naciones.

El autor destaca la importancia de la ciberseguridad y cómo no es solo una responsabilidad de las personas encargadas de Tecnologías de la Información, sino que es una tarea que involucra a todos en la organización. Los ejercicios de identificación de riesgos y controles en las empresas para evaluar los activos de información críticos han evolucionado de ser una tarea exclusiva de seguridad a ser

una disciplina fundamental en la gestión de la información, lo que a su vez se convierte en una ventaja clave en el entorno de negocios. [5].

La ciberseguridad, según Agnese Carlini, es un tema relevante en la sociedad occidental, ya que muchos procesos industriales, bancarios, instalaciones de energía, entre otros, dependen en gran medida de los sistemas informáticos que están interconectados. Con el surgimiento del ciberespacio, se han planteado nuevos desafíos en cuanto a la evolución de medidas de seguridad y la regulación legal para hacer frente a las amenazas en este ámbito. [3].

4. Determinación del Problema

Objetivo General:

- Levantar un estado del arte de las técnicas de Machine Learning para Ciberseguridad en la actualidad.

Objetivos específicos:

- Revisar el estado del arte de las técnicas de machine learning para ciberseguridad.
- Establecer las principales técnicas de deep learning utilizadas para ciberseguridad en la era del Big Data.
- Implementar una técnica de ML para Ciberseguridad en un repositorio público para su difusión.

5. Marco Teórico

5.1. Aprendizaje Automático

El aprendizaje automático es un campo de la inteligencia artificial que posibilita a los computadores aprender a realizar una tarea sin tener que haber sido programados para ejecutarla[6].

5.1.1. Tipos de aprendizaje:

5.1.1.1. Según su salida

El aprendizaje automático se divide en supervisado y no supervisado.

- Supervisado: Estos son conocidos por saber los valores de la variable que genera la respuesta a los datos que se usan en el entrenamiento. Este algoritmo realiza una función que determina una relación entre las entradas y salidas del sistema. Poseen un conocimiento previo establecido en el conjunto de datos el cual está relacionado por sus características y su resultado, sobre los cuales se realizan predicciones, las misma que ayudan a tomar decisiones.

Por ejemplo, se utiliza una base de datos de correos electrónicos, todo ellos en la categoría de SPAM o NO SPAM, la entrada del sistema sería el texto del correo y la salida sería la determinación de si el texto es SPAM.

- No supervisado: Este aprendizaje en el cual los resultados se ajustan a las observaciones. Al contrario del supervisado, no tiene un conocimiento previo. Este algoritmo explora de forma autónoma las características de los datos de entrada, tratándolos como un conjunto de variables aleatorias, formando así una base para el conjunto de datos.

Por ejemplo, se tiene una base de datos de la navegación web de todos los usuarios, no existe una salida como tal, pero el sistema puede clasificar en base a los gustos y frecuencia de navegación diferentes tipos de grupos de usuarios.

- Semisupervisado: Se combinan los dos tipos anteriores.
- Por refuerzo: Esta basado en el método ensayo y error, las respuestas que obtiene de sus actividades son sus datos de entrada.

5.1.1.2. Según el tipo de tarea

Los tipos de aprendizaje automático se asocian en:

Tipo predictivo:

- Clasificación: Su objetivo es asignar a los registros no clasificados los valores de clase partiendo de ejemplos de entrenamiento. La variable que utiliza como respuesta se representa en categorías, es decir, se asigna un conjunto determinado de datos.
- Regresión: Su objetivo es definir al atributo de un registro un valor real, partiendo de valores de los demás atributos del registro. La variable de su resultado se representa cuantitativamente y tomas infinitos valores.

Tipo descriptivo:

- Clustering: Divide los registros en clusters, de esta manera los registros de cada cluster se identifican por su gran semejanza, lo cual hace que haya una mayor distancia entre grupos.
- Asociación: Identifica las relaciones no explicitas de registros no clasificados. Intenta identificar las relaciones que se dan de acuerdo con la ocurrencia mutua de varios atributos.

5.2. Algoritmos de Aprendizaje Automático

5.2.1. Regresión logística

Es una técnica en la cual se realiza la clasificación en binario, el valor binario obtenido esta entre 0 y 1. Es una red neuronal que posee de manera exacta una neurona, la cual realiza una función logística. Este algoritmo calcula la relación que hay entre la variable dependiente e independiente. Para esto utiliza una función logística, por ejemplo, la función Sigmoide, con la cual se determina la probabilidad de la variable dependiente. Normalmente se utiliza un margen que determina que si está por encima de 0.5 la sentencia es cierta y por debajo es falsa, la cual se puede gestionar para reducir la cantidad de falsos negativos o falsos positivos.

Esta técnica es sencilla y sus resultados son fáciles de interpretar, pero su desventaja es que el problema a resolver debe ser separable linealmente de otra forma no es eficiente[7].

5.2.2. Naïve Bayes (NB)

Esta técnica utiliza modelos probabilísticos para el desarrollo del aprendizaje automático. Está basada en el teorema de Bayes, el cual se utiliza para determinar la probabilidad de que ocurra un evento A, dado que un evento B ha ocurrido (probabilidad condicional). Utiliza la ecuación 1:

Ecuación 1. Teorema de Bayes.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

En donde A y B son eventos, P la probabilidad y P(B) es diferente de 0.

5.2.3. Árbol de decisión

Los modelos predictivos son una de las formas de emplear un árbol de decisión, relacionando observaciones de un artículo con conclusiones acerca de su valor objetivo.

Los árboles de clasificación son modelos de árbol que se utilizan cuando la variable de destino posee un grupo limitado de valores posibles, y en los que las hojas hacen referencia a etiquetas de clase mientras que las ramas reflejan la combinación de características que llevan a esas etiquetas de clase. Por su parte, los árboles de regresión son árboles de decisión en donde la variable de destino toma valores continuos.

En el análisis de decisiones, una forma de emplear un árbol de decisión es como una herramienta visual y explícita para representar la toma de decisiones.

En la minería de datos, un árbol de decisión describe los datos y no las decisiones, aunque el árbol de clasificación resultante se puede utilizar para tomar decisiones[7].

5.2.4. Random forest

El Bosque Aleatorio, es una combinación de árboles predictores en lo que cada árbol depende de los valores de un vector aleatorio ensayado independientemente y con la misma distribución para cada uno de estos. Es una reforma sustancial del bagging de Breiman que construyó: "una selección aleatoria de árboles de decisión con una variación controlada y luego los promedia."

Una cantidad considerable de problemas de rendimiento que tiene el algoritmo de bosque aleatorio se asemejan a la de los algoritmos de boosting, y es más simple de entrenar y ajustar. Por estos motivos es un algoritmo popular y ampliamente utilizado [7].

5.2.5. Support Vector Machine

La Máquina de Vectores de Soporte es un algoritmo que se utiliza en problemas de clasificación y regresión debido a su bajo costo computacional y alta precisión. Su objetivo principal es encontrar un hiperplano en un espacio N-dimensional (donde

N es el número de características) mediane el cual se clasifican los puntos de datos. Aunque existen muchos hiperplanos posibles para separar los datos en dos clases, SVM intenta encontrar el plano que tenga el margen máximo entre las dos clases, es decir, la distancia máxima que puede llegar a tener entre los puntos de datos de cada clase. Esto permite que los datos se clasifiquen con mayor precisión en una de las dos clases que SVM ha determinado. La figura 1 muestra cómo se separan los datos a clasificar, con una línea continua que representa el límite entre ambas clases [7].

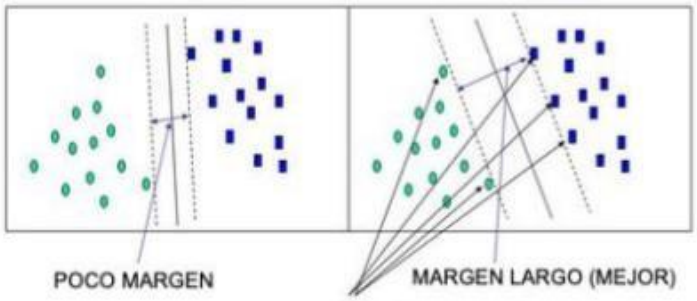

Figura 1 Vectores de soporte en hiperplanos

En ocasiones, SVM puede tener dificultades para lograr una separación adecuada en el plano, lo que se llama comúnmente como sobreajuste (overfitting). Para evitar esto, SVM utiliza un parámetro llamado C para poder controlar la compensación entre los márgenes rígidos y los errores de entrenamiento, y así crear un margen suave (soft margin) que permita algunos errores de clasificación, pero los penalice.

La ilustración anterior que se mostró es un caso ideal que se puede representar en dos dimensiones del plano, pero no siempre se puede aplicar el algoritmo SVM de esa manera, debido a esto se introduce la representación mediante funciones kernel. Estas funciones son las encargadas de proyectar la información en un espacio de características de mayor dimensión, lo que puede incrementar la capacidad computacional de las máquinas de aprendizaje lineal.

5.2.6. K vecinos más cercanos (K-Nearest Neighbors)

En el proyecto en cuestión se ha evaluado también el algoritmo K vecinos más cercanos, que al igual que los algoritmos anteriores es un algoritmo supervisado de aprendizaje automático. Su funcionamiento está basado en clasificar los valores utilizando los puntos de datos más cercanos en términos de distancia, los cuales son aprendidos durante el entrenamiento del modelo. La K en su nombre hace referencia al total de puntos vecinos que se tienen en cuenta para clasificar. A diferencia de otros algoritmos, KNN no aprende de manera explícita un modelo, sino que lo realiza memorizando las instancias en la fase de entrenamiento para su posterior predicción.

KNN calcula primero la distancia que existe entre el ítem a clasificar y el resto de los elementos del conjunto de datos que se utilizan en el proceso de entrenamiento. Luego, selecciona los "K" elementos que se encuentren más cercanos, utilizando una función de distancia determinada. Finalmente, para poder elegir se realiza un proceso de votación de mayoría entre los "K" puntos y la etiqueta que prevalezca se convertirá en la opción final. Para determinar la cercanía entre puntos, se suelen emplear medidas de similaridad, entre las más utilizadas tenemos la distancia Euclidiana o la Similitud Coseno, la cual se encarga de medir el ángulo entre los vectores.

Entre las ventajas de KNN se encuentran su rapidez de cálculo, facilidad de uso y la no necesidad de hacer presunciones sobre el conjunto de datos. Sin embargo, una desventaja es que la precisión de la clasificación depende de la calidad de los datos en el entrenamiento, la elección de un valor óptimo de "K" y la posibilidad de realizar una clasificación deficiente de los puntos de datos en un límite donde no se pueden clasificar de manera efectiva [7].

5.3. Aplicaciones del Machine Learning a la ciberseguridad

La aplicación de las técnicas de aprendizaje automático al campo de la ciberseguridad es más extensa con el paso del tiempo, en el cual su principal cometido es la detección de varios tipos de amenazas.

Las principales aplicaciones de machine learning en la seguridad de los sistemas de información son[8]:

5.3.1. Detección y prevención de intrusos.

Esta entre los principales sistemas informáticos en el cual su finalidad es detectar accesos de usuarios ajenos a la red, servidores o algún sistema en general es el IDS (Sistema de Detección de Intrusos). Al ser un sistema pasivo, reúne muchos datos mediante sensores virtuales, en su mayoría sniffers.

Otro sistema informático que protege contra ataques o intrusiones a aplicaciones web o servidores son los WAF (Web Application Firewall).

Ambos sistemas para poder realizar un análisis y poder diferencias entre las conexiones que son legítimas y las que no, se pueden lograr realizando diferentes aproximaciones usando machine learning, siendo las más utilizadas la basada en anomalías y la basada en firmas [8].

5.3.2. Aproximación basada en firmas.

Los IDS que se basan en firmas reúnen los paquetes que viajan por la red o a un servidor, los cuales son comparados con una base de datos de firmas, que son datos cargados previamente de patrones de ataques. Si alguno de estos paquetes coincide con los que se encuentran en la base, serán identificado como intento de intrusión.

Estos sistemas clasifican un paquete como malicioso o no malicioso. De acuerdo con el tipo o tipos de algoritmos de clasificación que se usan para esta técnica se pueden destacar:

- Clasificador único. Este tipo de técnica utiliza un solo algoritmo de machine learning para detectar intrusiones. Los más utilizados son: algoritmos de árboles de decisión, redes neuronales artificiales y máquinas de vectores de soporte.
- Clasificador hibrido. Este tipo de técnica utiliza una combina algoritmos diferentes para abarcar todas las fases del aprendizaje automático, desde normalizar los datos, hasta la clasificación. Un ejemplo es utilizar un algoritmo de clasificación como Naïve Bayes que marca con una bandera los paquetes sospechosos en la red y los envía como entrada a un algoritmo que utiliza el modelo oculto de Markov el cual se encarga de adicionar las direcciones IP de origen de ese paquete en una black list.
- Clasificador combinado. El objetivo es desarrollar un clasificador sólido a través de la combinación de múltiples clasificadores débiles. Los clasificadores débiles se caracterizan por tener una precisión mediocre, pero aun así son más precisos que los métodos de clasificación por especulación. Para lograr este objetivo, se emplea principalmente la técnica de boosting, que mezcla los resultados de varios clasificadores débiles para obtener un clasificador más robusto. Un ejemplo de esta técnica sería la combinación de algoritmos basados en redes neuronales profundas (DNN) y agrupamiento espectral [8]

5.3.3. Aproximación basada en anomalías

La estrategia consiste en vigilar un sistema e identificar cualquier desviación que se produzca con relación a su comportamiento normal. A continuación, se presenta un resumen del análisis de las técnicas de detección de intrusiones basadas en anomalías.

- Clasificador único. Algunos investigadores utilizan algoritmos y técnicas como máquinas de vectores de soporte, Naïve Bayes, árboles de decisión y depp learning para la detección de intrusiones. El deep learning tiene la ventaja de lograr una alta precisión en la detección, pero el proceso de aprendizaje requiere de un tiempo significativo. En cuanto a los árboles de decisión y Naïve Bayes, aunque tienen un mayor porcentaje de falsos positivos que el deep learning, su precisión mejora significativamente cuando se entrenan con casos reales de intrusiones ejecutadas en otros sistemas. El uso de máquinas de vectores de soporte puede ser muy efectivo, llegando a alcanzar un 100% de precisión cuando se utilizan datos no homogéneos en el conjunto de entrenamiento.
- Clasificador hibrido: Se han llevado a cabo investigaciones sobre IDS que utilizan el algoritmo C4.5 y k-means modificado como clasificadores. Estos algoritmos permiten procesar grandes cantidades de tráfico de red en tiempo real de manera eficiente, aunque la precisión obtenida no suele ser muy elevada [8].

5.3.4. Detección de Phising

De acuerdo con la compañía española Panda Security [9], especializada en soluciones de seguridad informática y recientemente adquirida por WatchGuard, el phishing se refiere a la práctica de enviar correos electrónicos que aparentan provenir de fuentes confiables (como bancos o compañías de energía), pero que en realidad buscan engañar al destinatario para obtener información confidencial. En la detección de estos de ataques, se pueden emplear técnicas de machine learning para identificar correos electrónicos y cuentas de redes sociales fraudulentas en comparación con las legítimas.

En lo que respecta a la detección de correos electrónicos fraudulentos, Hamid et al. [10] sugieren el uso de técnicas y algoritmos como AdaBoost y SMO, pero obtienen una tasa de error del 18%. Otros estudios, como el de Basnet et al. [11], proponen el uso de Naïve Bayes y logran reducir la tasa de error al 1.6%.

En cuanto a la detección de sitios web fraudulentos, Li et al.[12] plantean una estrategia fundamentada en el uso de un subtipo de máquinas de vectores de soporte conocidas como TSVM (transductive support vector machine).

5.3.5. Preservación de la privacidad

Para que se cumplan las normas como el GDPR2 de la Unión Europea y poder respetar la privacidad de los usuarios, el principal objetivo al utilizar técnicas de machine learning en datos de usuario es conseguir la mayor cantidad de información útil para el análisis sin comprometer la privacidad. Para lograr este equilibrio, se requiere una regla de compromiso adecuada. Se utilizan técnicas como máquinas de vectores de soporte [13] y agrupamiento de k-medias [14] en este ámbito.

5.3.6. Detección de spam

Se aplica el aprendizaje automático en la detección de mensajes no solicitados en diferentes canales como correo electrónico, SMS, redes sociales o comentarios en blogs. Sutta et al. [15] llevó a cabo un estudio comparando el rendimiento de varios sistemas clasificadores de correo electrónico, incluyendo diferentes tipos de máquinas de vectores de soporte, redes neuronales artificiales, k vecinos más cercanos, Naïve Bayes, regresión logística, árboles de decisión y Random Forest. Por otro lado, Chen et al. [16] estudió el empleo de Random Forest, C4.5, Naïve Bayes, KNN Bayes Network y máquinas de vectores de soporte para la detección en tiempo real de mensajes de spam en Twitter.

5.3.7. Análisis de riesgo

Se puede llevar a cabo un procedimiento denominado análisis de riesgos, según la norma ISO 9001:2015, el cual permite identificar los elementos o factores que tengan un alto grado de generar riesgos u oportunidades que afecten negativa o

positivamente a la organización. En este proceso, se pueden aplicar técnicas de machine learning, como el cálculo del nivel de riesgo (crítico, alto, medio o bajo) en el análisis cualitativo de los mismos. Eminagaoglu et al.[17] proponen una técnica en la cual los riesgos son clasificados en dos categorías: riesgo y sin riesgo. Se compararon un total de 68 algoritmos de clasificación, y se encontró que REPtree, un tipo de árbol de decisión arrojó los mejores resultados.

5.3.8. Detección de programa maligno

Kaspersky [18] define el malware como un software malicioso diseñado para infectar un ordenador y causar daños de diversas formas. El malware puede tomar muchas formas, como virus, gusanos, troyanos, spyware, entre otros, y puede infectar los dispositivos de los usuarios de diversas maneras. Por lo tanto, es importante que los usuarios sepan cómo detectar y protegerse contra el malware.

Para detectar el malware, se pueden aplicar diversas técnicas de aprendizaje automático, como la la detección basada en anomalías, detección basada en firmas y la detección basada en heurísticas. La mayoría de estas aplicaciones son software, pero en [19] presentan una solución basada en hardware y software que detecta el malware en tiempo real. Utilizando técnicas y algoritmos como J48, máquinas de vectores de soporte, Naïve Bayes, optimización mínima secuencial (SMO), regresión lineal y perceptrón multicapa, esta solución puede localizar hasta el 46% de los malware durante el primer 30% de la ejecución de la muestra observada y hasta el 97% al ejecutarla de manera completa, todo con una tasa de falsos positivos del 3%.

5.3.9. Testing de propiedades de seguridad

Esta aplicación se refiere a las pruebas que se emplean para verificar que se haya realizado de manera correcta la implementación de un protocolo criptográfico en un sistema distribuido, como puede ser utilizado en sistemas críticos, infraestructuras militares o comunicaciones móviles ad-hoc. Para ello se emplean diferentes métodos como pruebas de caja negra o pruebas en tiempo real. Además,

en este ámbito también se puede aplicar el machine learning. Un ejemplo de ello es el trabajo de Shu et al. [20], que propone el uso de algoritmos de aprendizaje supervisado junto con pruebas de caja negra para evaluar automáticamente y de manera sistemática la implementación de un protocolo de mensajería en un sistema distribuido, garantizando la confidencialidad de las comunicaciones.

5.4. Deep Learning y redes neuronales

Su fortaleza es que aprende a medida que se va entrenando en tiempo real y permite el desarrollo de nuevos criterios de clasificación sin necesidad de intervención humana. Por ejemplo, se está aplicando para combatir el software malicioso y el fraude en línea. ¿Por qué? Porque los ciberdelincuentes evolucionan rápidamente, creando amenazas que pueden adecuarse a la seguridad de los sistemas. Por lo tanto, el aprendizaje profundo es capaz de localizar y clasificar estas amenazas y resolverlas de manera eficiente y rápida[21], [22] Sin embargo, sus aplicaciones son interminables, por ejemplo, en el caso de FeedZai se utiliza como método de identificación, lo que les permite reconocer si el usuario es un ser humano o un robot, si un ciberdelincuente intenta imitar una identidad de usuario, o si un ciberdelincuente está interactuando con la cuenta de un usuario desde cualquier parte del mundo.

5.4.1. Perceptron multicapa (MLP Multilayer Perceptron)

El Perceptrón Multicapa (MPL) como muestra la figura 2, es la base de la arquitectura de las redes neuronales y aprendizaje profundo. Es una red de neuronas interconectada completamente entre ella que incluye una capa de entrada para recibir información, una capa de salida para tomar decisiones o hacer predicciones basadas en la señal de entrada, y una o más capas ocultas, que son el núcleo computacional de la red.[23]

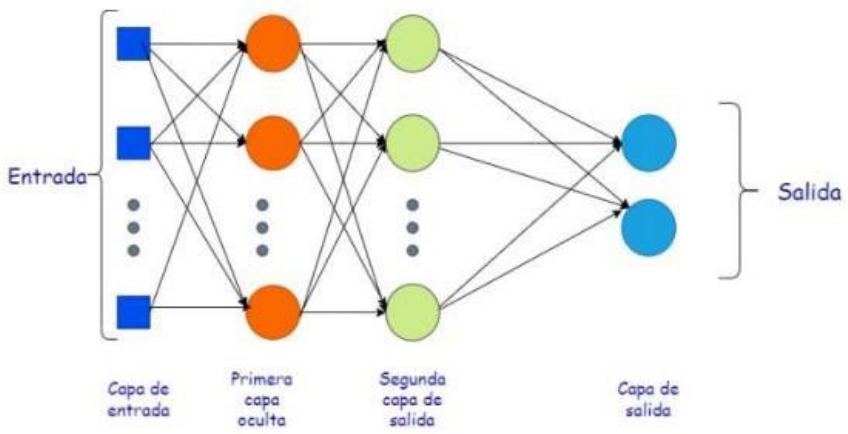

Figura 2 Red MPL. Fuente: Quirumbay Yagual, D., Castillo Yagual, C., & Coronel Suárez, I. (2022).

MPL está completamente conectado, las neuronas de una capa se conectan con un peso establecido a las neuronas de la siguiente capa. Para su activación se utilizan funciones como Rectified Linear Unit (Relu), que define la salida de una red, también se las conoce como funciones de transferencia, agregan a la red propiedades no lineales para adquirir mapas funcionales a partir del análisis de los datos. Además, se emplea una técnica de aprendizaje supervisado para realizar el entrenamiento, el cual tiene por nombre Backpropagation, el cual es elemental en una red neuronal; el cual es considerablemente utilizado las redes neuronales FeedForward. El principal propósito de este algoritmo es asociar con exactitud las entradas a las salidas a través de la optimización de los pesos de la red

En el proceso de entrenamiento se aplican varias técnicas para optimizar, como lo es el descenso gradiente estocástico. Estas redes neuronales se aplican en la elaboración de un modelo de detección de intrusos, sistemas de IoT fiables, analizar amenazas de seguridad.[24] MPL necesita varios hiper-parametros ya que es muy sensible al escalado de características, algunos de ellos son: número de capas que

se encuentran ocultas, las neuronas e interacciones que se deben ajustar, esto hace este modelo costoso computacionalmente en la resolución de problemas de seguridad complejos.

5.4.2. Red neuronal convolucional (CNN Convolutional neural network)

Es un modelo de red el cual aprende de los datos, sin tener que extraer de forma manual las características. Una CNN como se muestra en la figura 3, consta típicamente de una capa de entrada y una de salida en sus extremos, en el centro de estas consta de capas de agrupación, capas convolucionales y capas totalmente conectadas.

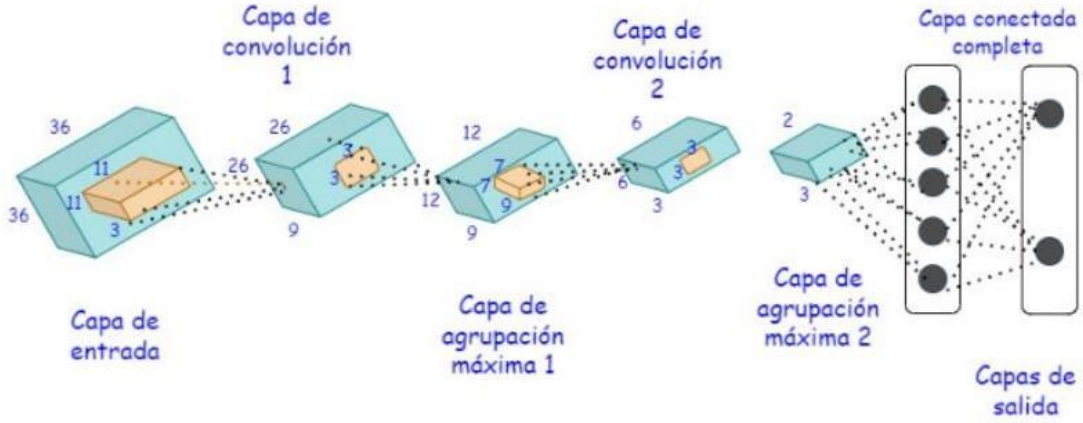

Figura 3 Red CNN. Fuente: Quirumbay Yagual, D., Castillo Yagual, C., & Coronel Suárez, I. (2022).

Cada una de estas capas maneja parámetros optimizados en la obtención de resultados relevantes, así como para disminuir la dificultad. Este tipo de redes neuronales están específicamente diseñadas para atender la gran variedad de imágenes en formato 2D. Además, se utilizan de manera amplia en el reconocimiento de video e imágenes, analizar imágenes médicas, sistemas de

recomendación y clasificación de imágenes, procesar el lenguaje natural, entre otros.

Estas redes pueden ser utilizadas en el entorno de la ciberseguridad, por ejemplo, un modelo de aprendizaje profundo basado en CNN el cual es utilizado para las detecciones de intrusos o en ataques de denegación de servicios, redes IoT, detecciones de malware, etc. Esta red neuronal artificial tiene una mayor carga computacional, pero su ventaja es que le permite detectar de manera automática las propiedades de mayor relevancia sin supervisión humana alguna, por lo que es considerada una gran opción en la generación de soluciones de seguridad informática aplicada.[25]

5.4.3. Red neuronal recurrente de memoria a corto plazo (LSTM Long Short Term Memory – RNN Recurrent Neural Networks)

Una red neuronal recurrente es una red artificial que permite procesar secuencias de entradas en el aprendizaje profundo y retener su estado al mismo tiempo que procesa la siguiente cadena de entradas. Las RNN poseen en su capa recurrente bucles de retroalimentación, lo que les da la capacidad de mantener la información en la memoria a lo largo del tiempo. Las redes de memoria a corto plazo (LSTM) son una variante de RNN que utilizan unidades especiales, además de las unidades estándar, para abordar el problema del desvanecimiento del gradiente. Las unidades LSTM puede almacenar datos por largos periodos de tiempo gracias a que tienen una celda de memoria y donde las "puertas de olvido", "puertas de entrada" y "puertas de salida" trabajan juntas para verificar el flujo de información en una unidad LSTM.

Las redes LSTM son muy convenientes para el aprendizaje y el estudio de datos secuenciales, como el procesamiento, clasificación y predicción basada en datos de series temporales, lo que las hace diferentes de otras redes convencionales. Son comúnmente aplicadas en áreas como detección de anomalías en series temporales, la predicción de series temporales, chatbots de respuesta a preguntas,

procesamiento de lenguaje natural, reconocimiento de voz, traducción automática, entre otros. Con el incremento de datos secuenciales de seguridad generados actualmente, como actividades maliciosas relacionadas con el tiempo, flujos de tráfico de red, entre otros., un modelo LSTM como muestra la figura 4, también puede ser relevante en el campo de la ciberseguridad, especialmente al estudiar soluciones de seguridad basadas en ella, como la detección de phishing,[26] y la detección y clasificación de aplicaciones maliciosas.[27]

Pese a que la principal ventaja de una red recurrente comparada con una red tradicional es la capacidad de modelar secuencias de datos, puede requerir una gran cantidad de recursos y tiempo para su entrenamiento. Por lo tanto, considerando esta ventaja, una red LSTM-RNN efectiva está en la capacidad de mejorar los modelos de seguridad para detectar amenazas, especialmente cuando los patrones de procedimiento de las amenazas manifiestan un comportamiento temporal dinámico.

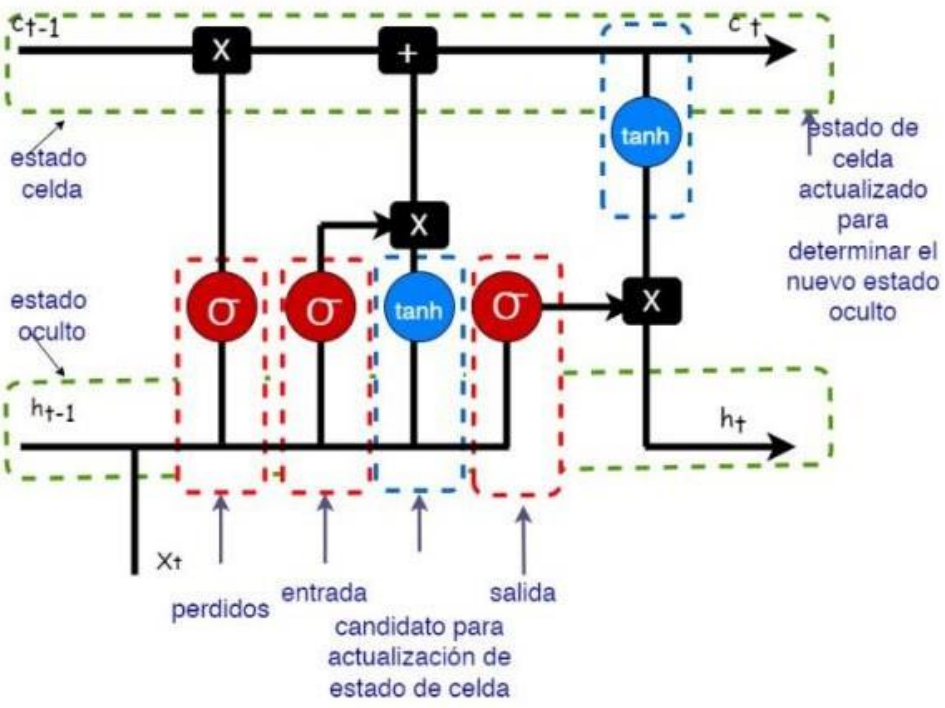

Figura 4 Red LSTM. Fuente: Quirumbay Yagual, D., Castillo Yagual, C., & Coronel Suárez, I. (2022).

5.4.4. Aprendizaje profundo por transferencia (DTL Deep Transfer Learning) o Deep TL

Este enfoque aborda el problema de la falta de datos de entrenamiento adecuados y permite entrenar modelos de Inteligencia Artificial (IA) con cantidades más pequeñas de datos como muestra la figura 5. Debido a la falta de datos etiquetados en la mayor parte de problemas actuales, el aprendizaje profundo por transferencia se ha vuelto muy habitual en el campo de la ciencia de datos. Hay tres categorías de aprendizaje por transferencia: aprendizaje por transferencia no supervisado, por transferencia transducido y por transferencia inductiva.

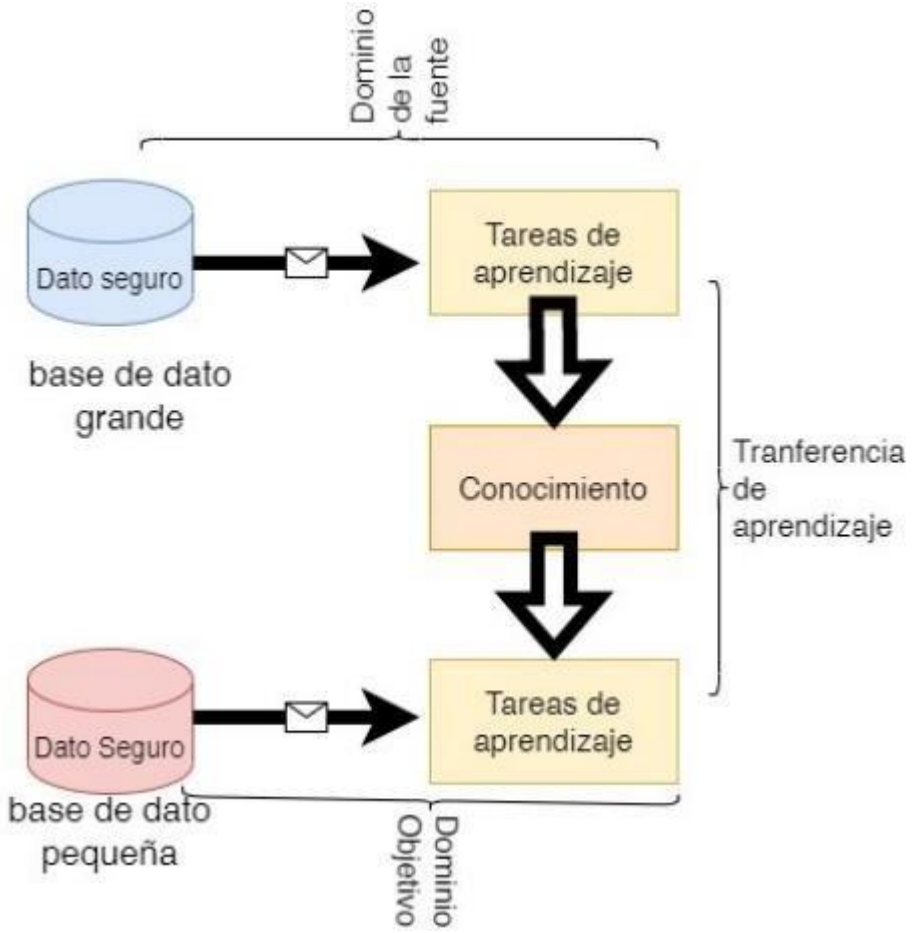

Figura 5 Red DTL. Fuente: Quirumbay Yagual, D., Castillo Yagual, C., & Coronel Suárez, I. (2022).

Además, el aprendizaje profundo por transferencia también es útil en la detección de fraudes financieros. Por ejemplo, utilizando el aprendizaje de transferencia, se pueden transferir los conocimientos adquiridos en la detección de fraudes en una

industria financiera a otras industrias similares. Esto permite mejorar la eficiencia en la detección de fraudes y reducir los tiempos de entrenamiento en comparación con los modelos tradicionales que requieren una gran suma de datos etiquetados.[28]

En el ámbito de la salud, el aprendizaje por transferencia se utiliza para el reconocimiento de enfermedades y la toma de decisiones clínicas. Por ejemplo, los modelos de aprendizaje de transferencia se han utilizado para la clasificación de imágenes médicas, como tomografías y radiografías, para ayudar en la detección de enfermedades.

En resumen, el aprendizaje profundo por transferencia es una técnica valiosa en el ámbito de la inteligencia artificial, y su uso en las distintas áreas han demostrado ser eficiente y efectiva en la solución de problemas complejos.

6. Materiales y Metodología

Para alcanzar cada objetivo específico, es necesario el uso de diferentes técnicas que nos ayuden a la obtención de un certero resultado a lo que nos hemos planteado.

- Revisar el estado del arte de las técnicas de machine learning para ciberseguridad.

  Este objetivo se lo realizara mediante una revisión sistemática de literatura que permita una explicación de las técnicas y métodos que se usan en la ciberseguridad, con el fin de mostrar y analizar las diferentes metodologías existentes para implementar dichas actividades en la vida cotidiana de las personas naturales y jurídicas. Se detallará las más recomendables técnicas y sus beneficios a corto y largo plazo en materia de protección de datos. La revisión sistemática de literatura se encuentra detallada en la sección 5.

- Establecer las principales técnicas de deep learning utilizadas para ciberseguridad en la era del Big Data.

  Se realizará una investigación documental en las principales fuentes de información referentes a ciberseguridad, que permitan detallar los trabajos relacionados entre ciberseguridad y Deep learning. El detalle de técnicas se ha desarrollado en la sección 5 del presente documento.

- Implementar una técnica de machine learning para ciberseguridad en un repositorio público para su difusión.

  Dicho objetivo se llevará a cabo mediante prácticas empíricas, que detallan de manera específica el alcance de estas técnicas en su ambiente de prueba. Con la ayuda de esta técnica, se espera como resultado los beneficios de las diferentes prácticas aplicables en esta rama, comparando sus resultados en cuanto a números, ahorro de dinero, nivel de seguridad, etc., y con esta información, designar bajo nuestro criterio, cuáles son las herramientas principales, y poder presentarlos como recurso para que cualquier estudiante o negocio en crecimiento puede acceder a estos y tener una guía del uso de estas herramientas en su entorno.

7. Resultados y Discusión

El principal resultado es la elaboración de un programa en Github que se encuentra disponible públicamente para su descarga y utilización, el cual se puede obtener del siguiente enlace https://github.com/wvillavicenciob/Isolation_Tree/tree/master . El mismo es un ejemplo de codificación sobre una técnica, la cual es definida como Isolation Tree, la misma que tiene sus siguientes características:

1. **Algoritmo Basado en Árboles**: El Isolation Forest es un algoritmo de detección de anomalías basado en árboles de decisión. Lo que lo diferencia de otros

métodos, es su enfoque en separar anomalías del conjunto en lugar de solo detectarlas.

2. **Uso de Subconjuntos Aleatorios**: Este algoritmo de detección crea arboles de decisión de manera aleatoria utilizando subconjuntos aleatorios de datos. Esto permite una detección más efectiva de anomalías, ya que las anomalías tienden a ser más susceptibles a la separación en comparación con las instancias normales.

3. **Métrica de Anomalía**: La puntuación de anomalía se calcula en función de la profundidad del árbol en el que se aísla una instancia. Las instancias que se aíslan en menos niveles son consideradas más anómalas.

Ventajas de Isolation Forest:

1. **Eficiencia**: Isolation Forest es especialmente eficiente para conjuntos de datos grandes debido a su capacidad para dividir el espacio de búsqueda rápidamente utilizando subconjuntos aleatorios.
2. **Manejo de Datos No Lineales:** A diferencia de algunos otros métodos de detección de anomalías, el Isolation Forest puede manejar datos no lineales de manera efectiva, lo que lo hace adecuado para una variedad de aplicaciones.
3. **Escalabilidad:** Es escalable y puede aplicarse a problemas en tiempo real.
4. **Parámetro Fácil de Ajustar**: El número de árboles y la profundidad máxima del árbol son dos parámetros principales que se pueden ajustar para controlar la sensibilidad del algoritmo a las anomalías.

Desventajas de Isolation Forest:

1. **Sensibilidad a Parámetros:** Si los parámetros no se ajustan adecuadamente, el rendimiento del Isolation Forest puede verse afectado. En particular, el número de árboles y la profundidad máxima deben ajustarse cuidadosamente.

2. **Puede Sobreestimar Anomalías**: En algunos casos, Isolation Forest puede sobreestimar la gravedad de una anomalía debido a la aleatoriedad en la creación de árboles.

Usos Ejemplificados de Isolation Forest:

1. **Detección de Fraude Financiero:** Isolation Forest se utiliza para identificar transacciones financieras fraudulentas, que son generalmente excepciones en un conjunto de datos.
2. **Detección de Intrusiones en Redes:** Puede utilizarse para detectar actividades inusuales en una red, lo que puede indicar intentos de intrusión.
3. **Control de Calidad de Manufactura:** Isolation Forest puede utilizarse para identificar productos defectuosos en una línea de producción basándose en mediciones anómalas.
4. **Detección de Enfermedades Raras:** En medicina, se puede aplicar para identificar enfermedades raras o condiciones médicas inusuales basadas en datos de pacientes.
5. **Detección de Comportamiento de Usuarios Anómalos:** En aplicaciones de seguridad cibernética y sistemas de recomendación, Isolation Forest puede detectar patrones de comportamiento de usuarios inusuales.

8. Conclusiones

La inteligencia artificial dentro del ámbito de la seguridad informática nos ha permitido reflexionar de que es algo de suma importancia y que cada día su evolución se acelera, sin embargo, también enfrenta un equilibrio constante entre los beneficios que brinda y las desventajas que pueden ser aprovechadas por los ciberdelincuentes para realizar ataques más complejos y fraudulentos.

Aunque los informes actuales sobre la inteligencia artificial en ciberseguridad son interesantes, su comprensión sigue siendo complicada para el público en general, ya que requieren cierto conocimiento previo.

Existen diversas herramientas de inteligencia artificial para la ciberseguridad, como la que se ha utilizado en este trabajo, la cual es fácil de usar y ha proporcionado resultados interesantes. Cada día se avanzan en herramientas con usos muy prometedores, pero también pueden ser utilizadas con malas intenciones, al fin y al cabo son los seres humanos quienes deciden utilizar las tecnologías para el bien o con otros fines.

Las técnicas de aprendizaje automático desempeñan un papel crucial en la ciberseguridad al proporcionar herramientas y enfoques efectivos para detectar, prevenir y responder a amenazas cibernéticas. Estas técnicas permiten una mejora significativa en la capacidad de las organizaciones para proteger sus sistemas y datos frente a ataques y vulnerabilidades. Algunos puntos clave a destacar gracias a la elaboración del presente trabajo son:

Detección más precisa, automatización de respuestas, adaptación continua, predicción de amenazas, mejora de la eficiencia. Estos aspectos en el ámbito de la seguridad de la información son aspectos clave en el diseño, desarrollo e implementación de soluciones, por tanto, el uso de estas técnicas es una condición indispensable en entornos de ciber seguridad.

Sin embargo, también es importante reconocer que el aprendizaje automático no es una solución única para todos los desafíos de ciberseguridad. Debe utilizarse en conjunto con otros enfoques de seguridad que deberían ser tomados en cuenta dentro de un análisis de riesgos, como la capacitación de empleados, la gestión de parches y la implementación de políticas de seguridad sólidas.

Además, la seguridad cibernética es un campo en constante evolución, y las amenazas se vuelven cada vez más sofisticadas. Por lo tanto, las organizaciones deben mantenerse actualizadas y estar dispuestas a ajustarse a medida que surgen nuevas amenazas y se desarrollan nuevas técnicas de aprendizaje automático para abordar estos desafíos. En resumen, el aprendizaje automático es una herramienta valiosa en la ciberseguridad, pero su éxito depende de una estrategia integral y una vigilancia constante.

9. Referencias

[1] "Cisco Annual Security Report Documents Unprecedented Growth of Advanced Attacks and Malicious Traffic". https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2014/m01/cisco-annual-security-report-documents-unprecedented-growth-of-advanced-attacks-and-malicious-traffic.html (consultado el 12 de septiembre de 2022).

[2] "Por qué el Machine Learning es un gran aliado para la ciberseguridad | WeLiveSecurity". https://www.welivesecurity.com/la-es/2021/12/10/por-que-machine-learning-aliado-para-ciberseguridad/ (consultado el 12 de septiembre de 2022).

[3] A. Carlini, "Ciberseguridad: un nuevo desafío para la comunidad internacional", jul. 2016.

[4] M. De y G. Avanzada, "MARCO DE REFERENCIA SOBRE LA CIBERSEGURIDAD EN ORGANIZACIONES Y EMPRESA".

[5] Jeimy J. Cano, "Ciberseguridad y ciberdefensa: dos tendencias emergentes en un contexto global", 2011. Consultado: el 12 de septiembre de 2022. [En línea]. Disponible en: https://acis.org.co/archivos/Revista/119/Editorial.pdf

[6] H. He, Y. Bai, E. A. Garcia, y S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning", *Proceedings of the International Joint Conference on Neural Networks*, pp. 1322–1328, 2008, doi: 10.1109/IJCNN.2008.4633969.

[7] I. H. Sarker, M. H. Furhad, y R. Nowrozy, "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions", *SN Comput Sci*, vol. 2, núm. 3, p. 173, 2021, doi: 10.1007/s42979-021-00557-0.

[8] J. M. Quesada Dueñas, "Aplicación de técnicas de machine learning a la ciberseguridad: Aprendizaje supervisado para la detección de amenazas web mediante clasificación basada en árboles de decisión", *Aplicación de técnicas de machine learning a la ciberseguridad*, p. 61, 2020, [En línea]. Disponible en: http://hdl.handle.net/10609/118166%0Ahttp://hdl.handle.net/10609/118166%0Ahttp://openaccess.uoc.edu/webapps/o2/bitstream/10609/118166/1/josedueñasTFM0620.pdf

[9] "Phishing", *Pandasecurity.com*.

[10] I. R. A. Hamid y J. H. Abawajy, "An approach for profiling phishing activities", *Comput Secur*, vol. 45, pp. 27–41, 2014, doi: https://doi.org/10.1016/j.cose.2014.04.002.

[11] A. H. and L. Q. Basnet Ram B. and Sung, "Feature Selection for Improved Phishing Detection", en *Advanced Research in Applied Artificial Intelligence*, W. and A. M. and W. X. Jiang He and Ding, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 252–261.

[12] Y. Li, R. Xiao, J. Feng, y L. Zhao, "A semi-supervised learning approach for detection of phishing webpages", *Optik (Stuttg)*, vol. 124, núm. 23, pp. 6027–6033, 2013, doi: https://doi.org/10.1016/j.ijleo.2013.04.078.

[13] Q. Jia, L. Guo, Z. Jin, y Y. Fang, “Preserving Model Privacy for Machine Learning in Distributed Systems”, *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, núm. 8, pp. 1808–1822, 2018, doi: 10.1109/TPDS.2018.2809624.

[14] G. Jagannathan y R. N. Wright, “Privacy-Preserving Distributed k-Means Clustering over Arbitrarily Partitioned Data”, en *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, en KDD ’05. New York, NY, USA: Association for Computing Machinery, 2005, pp. 593–599. doi: 10.1145/1081870.1081942.

[15] N. Sattu, “A study of machine learning algorithms on email spam classification”, Southeast Missouri State University, 2020.

[16] C. Chen *et al.*, “A Performance Evaluation of Machine Learning-Based Streaming Spam Tweets Detection”, *IEEE Trans Comput Soc Syst*, vol. 2, núm. 3, pp. 65–76, 2015, doi: 10.1109/TCSS.2016.2516039.

[17] M. Eminagaoglu, “A Qualitative Information Security Risk Assessment Model using Machine Learning Techniques”, en *Proceedings of the ICT2012 Second International Conference on Advances in Information Technologies and Communication, Amsterdam, The Netherlands*, 2018, pp. 27–29.

[18] “¿Qué es el malware y cómo puedes protegerte de él?” https://latam.kaspersky.com/resource-center/preemptive-safety/what-is-malware-and-how-to-protect-against-it (consultado el 22 de septiembre de 2023).

[19] S. Das, Y. Liu, W. Zhang, y M. Chandramohan, “Semantics-Based Online Malware Detection: Towards Efficient Real-Time Protection Against Malware”, *IEEE Transactions on Information Forensics and Security*, vol. 11, núm. 2, pp. 289–302, 2016, doi: 10.1109/TIFS.2015.2491300.

[20] G. Shu y D. Lee, “Testing Security Properties of Protocol Implementations - a Machine Learning Based Approach”, en *27th International Conference on Distributed Computing Systems (ICDCS ’07)*, 2007, p. 25. doi: 10.1109/ICDCS.2007.147.

[21] “Ciberseguridad Máster en Deep Learning - Universidad de Alcalá”. https://master-deeplearning.com/camino-deep-learning-ciberseguridad/ (consultado el 12 de septiembre de 2022).

[22] “Inteligencia Artificial: ventajas y riesgos de un mundo con máquinas ‘humanas’”. https://www.20minutos.es/noticia/2932536/0/inteligencia-artificial-ventajas-riesgos-mundo-maquinas-humanas/ (consultado el 12 de septiembre de 2022).

[23] I. H. Sarker, M. H. Furhad, y R. Nowrozy, “AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions”, *SN Comput Sci*, vol. 2, núm. 3, p. 173, 2021, doi: 10.1007/s42979-021-00557-0.

[24] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU)”, *CoRR*, vol. abs/1803.08375, 2018, [En línea]. Disponible en: http://arxiv.org/abs/1803.08375

[25] B. Susilo y R. F. Sari, “Intrusion Detection in IoT Networks Using Deep Learning Algorithm”, *Information*, vol. 11, núm. 5, 2020, doi: 10.3390/info11050279.

[26] J. Yan, Y. Qi, y Q. Rao, “Detecting Malware with an Ensemble Method Based on Deep Neural Network”, *Security and Communication Networks*, vol. 2018, p. 7247095, 2018, doi: 10.1155/2018/7247095.

[27] M. A. Adebowale, K. T. Lwin, y M. A. Hossain, “Intelligent phishing detection scheme using deep learning algorithms”, *Journal of Enterprise Information Management*, vol. ahead-of-print, núm. ahead-of-print, ene. 2020, doi: 10.1108/JEIM-01-2020-0036.

[28] P. Wu, H. Guo, y R. Buckland, “A Transfer Learning Approach for Network Intrusion Detection”, en *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, 2019, pp. 281–285. doi: 10.1109/ICBDA.2019.8713213.

[29] “¿Qué es el hackeo? | Definición de hackeo | Avast”. https://www.avast.com/es-es/c-hacker (consultado el 12 de septiembre de 2022).