Tipo de artículo: Artículo original

Detección de anomalías en minería de datos: un enfoque de aprendizaje profundo

Anomaly detection in data mining: a deep learning approach

Adiane Cueto Portuondo [1], https://orcid.org/0009-0009-7669-8904
Mónica Delgado Hernández [2], https://orcid.org/0000-0002-2527-0037
Heidy Rodríguez Malvarez [3], https://orcid.org/0000-0002-5154-1339
Mónica Peña Casanova [4*], https://orcid.org/0000-0003-2500-4510

[1] Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. La Habana. Cuba. adianecp@estudiantes.uci.cu
[2] Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. La Habana. Cuba. monicadh@estudiantes.uci.cu
[3] Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. La Habana. Cuba. heidyrm@estudiantes.uci.cu
[4] Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. La Habana. Cuba. monica@uci.cu

*** Autor para correspondencia: monica@uci.cu**

Resumen

La detección de anomalías es crucial en el ámbito de la ciberseguridad, especialmente ante la creciente sofisticación de los ciberataques. Este artículo explora las técnicas basadas en aprendizaje profundo, que han ganado notoriedad por su capacidad para identificar patrones complejos en grandes volúmenes de datos y ofrecer soluciones eficaces en la detección de intrusiones, gestión de vulnerabilidades y seguridad de redes. Los resultados destacan la importancia de la detección de anomalías en la ciberseguridad y la efectividad de las técnicas basadas en aprendizaje profundo en diversas aplicaciones, como la detección de intrusiones, la gestión de vulnerabilidades y la seguridad en redes. Además, se presentan investigaciones relacionadas que abordan distintos aspectos de la detección de anomalías utilizando aprendizaje profundo. El trabajo identifica áreas clave de mejora y nuevas direcciones de investigación, como la interpretación de modelos de aprendizaje profundo, la integración de contexto y conocimiento del dominio, y la consideración de aspectos éticos y de privacidad en la detección de anomalías

Palabras clave: detección de anomalías; ataques cibernéticos; aprendizaje profundo; seguridad en redes

Abstract

Anomaly detection is crucial in the cybersecurity domain, especially in the face of the increasing sophistication of cyberattacks. This paper explores deep learning-based techniques, which have gained notoriety for their ability to identify complex patterns in large volumes of data and offer effective solutions in intrusion detection, vulnerability management, and network security. The results highlight the importance of anomaly detection in cybersecurity and the effectiveness of deep learning-based techniques in various applications, such as intrusion detection, vulnerability management, and network security. In addition, related research addressing different aspects of anomaly detection using deep learning is presented. The paper identifies key areas for improvement and new research directions, such as interpreting deep learning models, integrating context and domain knowledge, and considering ethical and privacy issues in anomaly detection.

Keywords: *anomaly detection; cyber attacks; deep learning; network security*

Recibido: 05/09/2024
Aceptado: 20/11/2024
En línea: 01/12/2024

Introducción

En la era digital actual, la detección de anomalías se ha consolidado como un componente esencial en diversos sectores, con un enfoque particular en el ámbito de la ciberseguridad. La proliferación de tecnologías interconectadas y la creciente dependencia de los sistemas digitales han sido acompañadas por un aumento alarmante de ataques cibernéticos que se vuelven cada vez más complejos y personalizados (Lan et al., 2018). Según un informe de Cybersecurity Ventures, se prevé que el costo global del delito cibernético alcance los 10.5 billones de dólares anuales para 2025, destacando la necesidad urgente de contar con técnicas robustas y eficientes que permitan identificar y mitigar estas amenazas en tiempo real.

En el mundo actual, se almacenan y transfieren enormes cantidades de datos de un lugar a otro. Cuando se transfieren o almacenan, los datos están expuestos a ataques. Aunque existen diversas técnicas o aplicaciones para proteger los datos, existen lagunas. Por ello, para analizar los datos y determinar diversos tipos de ataques, han surgido técnicas de minería de datos que los hacen menos vulnerables (Agrawal & Agrawal, 2015). La detección de anomalías utiliza estas técnicas de minería de datos para detectar el comportamiento sorprendente oculto en los datos, lo que aumenta las posibilidades de sufrir intrusiones o ataques. También se han realizado varios enfoques híbridos para detectar ataques conocidos y desconocidos con mayor precisión, tal es el caso del aprendizaje profundo.

La minería de datos y el aprendizaje profundo (*deep learning*) han emergido como enfoques vanguardistas en el campo de la detección de anomalías (Nguyen et al., 2019). Estos métodos, que forman parte de la inteligencia artificial, ofrecen la habilidad de procesar y analizar grandes volúmenes de datos, descubriendo patrones complejos que podrían indicar la ocurrencia de actividades sospechosas o maliciosas (Agrawal & Agrawal, 2015). A diferencia de los métodos tradicionales de detección, que a menudo dependen de reglas predefinidas y análisis manual, los algoritmos de aprendizaje profundo son capaces de aprender de los datos de manera autónoma, mejorando su precisión y adaptabilidad con el tiempo (Behrad & Abadeh, 2022).

Investigaciones recientes han demostrado que las técnicas de aprendizaje profundo, como las redes neuronales profundas (DNN) (Pang et al., 2021) y los autoencoders (Narayana et al., 2011), son particularmente efectivas en la identificación de anomalías en datos no estructurados y estructurados, lo que las convierte en herramientas invaluables para las organizaciones que buscan fortalecer su postura de seguridad (Wang et al., 2020). Como resultado, estas tecnologías no solo permiten una detección más rápida y precisa de incidentes de seguridad, sino que

también facilitan una respuesta proactiva frente a amenazas emergentes, subrayando así su relevancia en el panorama actual de la ciberseguridad (Hernández-Blanco et al., 2019).

Este artículo explora las técnicas basadas en aprendizaje profundo, que han ganado notoriedad por su capacidad para identificar patrones complejos en grandes volúmenes de datos y ofrecer soluciones eficaces en la detección de intrusiones, gestión de vulnerabilidades y seguridad de redes.

Materiales y métodos

Este estudio emplea una combinación de métodos empíricos y teóricos para explorar la efectividad de las técnicas de detección de anomalías en ciberseguridad, particularmente aquellas basadas en aprendizaje profundo. En primer lugar, los métodos empíricos están diseñados para recopilar datos cuantitativos y cualitativos que sustenten las afirmaciones sobre el rendimiento de estas técnicas. Se realizó un análisis exhaustivo de diversas bases de datos académicas y fuentes relevantes, utilizando criterios de inclusión específicos para garantizar la relevancia y calidad de los estudios revisados. Esta revisión permitió identificar aplicaciones concretas de técnicas de aprendizaje profundo en escenarios de ciberseguridad, así como su efectividad en la identificación y mitigación de ataques.

Las técnicas de minería de datos también se implementaron en este análisis empírico. Se exploraron algoritmos de aprendizaje supervisado y no supervisado para evaluar su capacidad de detectar anomalías en conjuntos de datos de tráfico de red. Se llevaron a cabo experimentos orientados a comparar el rendimiento de diferentes modelos de aprendizaje profundo, tales como redes neuronales convolucionales (CNN) y autoencoders. Estos experimentos no solo proporcionaron información sobre la precisión y la robustez de las técnicas, sino que también permitieron identificar patrones en los datos que podrían indicar actividades maliciosas.

Los métodos teóricos utilizados en el presente estudio se basan en una revisión crítica de la literatura existente sobre detección de anomalías y aprendizaje profundo. Se abordaron conceptos fundamentales, como la definición de anomalías en el contexto de la ciberseguridad y la evolución de los modelos de aprendizaje automático a lo largo del tiempo. Esta revisión no solo situó el estudio en un contexto académico más amplio, sino que también facilitó la identificación de lagunas en la investigación actual y áreas potenciales para futuras exploraciones.

Resultados y discusión

La detección de anomalías es una técnica de Minería de Datos con un amplio espectro de aplicaciones enfocadas en la seguridad social, como, por ejemplo: el análisis de redes informáticas y sociales, análisis de transacciones bancarias, y análisis de datos sensoriales, entre otros (Gadal & Mokhtar, 2017). Esta técnica permite el reconocimiento de patrones que no se comportan de la manera esperada en los datos. En una red informática, patrones de comportamiento inusual

podrían significar que una computadora pirateada está enviando datos confidenciales a un destino no autorizado (Gadal & Mokhtar, 2017).

Diferentes comportamientos en los datos de transacciones con tarjetas de crédito podrían indicar el robo de identidad o de la tarjeta de crédito. Las lecturas de comportamientos inusuales de un sensor de nave espacial podrían significar un error en algún componente de la nave. Incluso al trabajar con imágenes médicas, un cambio abrupto en la intensidad de los píxeles en lugares inesperados, puede indicar la presencia de tumores malignos. Todos estos patrones que no siguen el funcionamiento esperado son conocidos como anomalías y su detección permite la prevención de nuevos ataques, malos funcionamientos, así como la detección a tiempo de tumores (Parmar & Patel, 2017).

En el contexto de la ciberseguridad, una anomalía se refiere a cualquier desviación inusual o atípica del comportamiento normal en sistemas, redes o datos. Estas anomalías pueden indicar posibles amenazas o actividades maliciosas. Por ejemplo:

- Anomalías en el tráfico de red: Flujos de datos inusuales, conexiones no autorizadas o patrones de tráfico sospechosos.
- Comportamiento del usuario: Accesos no autorizados, intentos de autenticación fallidos o cambios inesperados en los patrones de uso.
- Uso de recursos anómalo: Consumo inusual de CPU, memoria o almacenamiento.

La detección de anomalías desempeña un papel crucial en el campo de la ciberseguridad, ya que permite identificar y mitigar una amplia gama de actividades maliciosas y amenazas emergentes (Yaseen, 2023). A medida que los ciberataques se vuelven cada vez más sofisticados, la capacidad de detectar y responder rápidamente a patrones inusuales de comportamiento se ha convertido en una prioridad fundamental para las organizaciones.

El uso de modelos de machine learning representa una notable ventaja, especialmente en el análisis de grandes volúmenes de datos, gracias a su capacidad para la toma de decisiones. Estos modelos presentan características que los hacen especialmente valiosos en el entorno empresarial, como su flexibilidad y su habilidad para ajustarse a cambios en los datos a medida que se incorporan al sistema y aprenden de las acciones propias del modelo (Yadav & Souza, 2024).

El Aprendizaje Profundo o Deep Learning (DL) es un nuevo campo del aprendizaje automático que se basa en redes neuronales artificiales. Se ha aplicado en muchas áreas, como el reconocimiento de voz y de imágenes, el procesamiento del lenguaje natural, el descubrimiento de fármacos y los sistemas recomendados. En los últimos años, el aprendizaje profundo ha demostrado su eficacia en el campo de la detección de intrusiones y en el área de la

ciberseguridad en general. Estas técnicas tienden a ser más eficientes que el ML tradicional debido a su estructura profunda y su capacidad para aprender las características importantes (Sarker et al., 2020).

Particularmente, en la detección de anomalías, las soluciones generalmente se clasifican en tres aspectos según la naturaleza del aprendizaje, incluida la detección de anomalías supervisada, no supervisada y parcialmente supervisada (Liu & Lang, 2019).

Detección de anomalías de manera supervisada: Este método utiliza técnicas de clasificación supervisada una vez que se conocen las anomalías existentes en los datos. Se etiquetan los datos en función de si son una anomalía o un dato normal dentro del umbral seleccionado. Aunque este método puede enfrentar problemas como la falta de disponibilidad de etiquetas de datos y el desequilibrio de las etiquetas de datos, estos problemas se pueden solucionar a través de métodos basados en instancias y basados en algoritmos.

Detección de anomalías semi-supervisadas: En este enfoque, se conoce la existencia de las anomalías, pero estas no se contemplan en el conjunto de datos. Con el conjunto de datos, se llega a definir la normalidad del ambiente. Existen varias técnicas para abordar este problema, incluyendo técnicas basadas en clasificación, en reglas de asociación y en máquinas de soporte vectorial.

Detección de anomalías no supervisadas: Este método se emplea normalmente en una situación en la que no se conoce ningún conocimiento previo del conjunto de datos, por lo tanto, no se presenta información de la etiqueta. Existen varias técnicas alternativas como: aproximaciones gráficas, aproximaciones paramétricas, aproximaciones basadas en vecinos cercanos y aproximaciones basadas en clustering.

Las técnicas de detección de anomalías, especialmente aquellas que emplean aprendizaje profundo, han demostrado ser eficaces en la identificación de una variedad de amenazas cibernéticas. Algunas de las principales aplicaciones incluyen:

- *Threat Hunting*: Se utilizan para detectar y mitigar amenazas avanzadas y persistentes (*Advanced Persistent Threats*, APT) mediante el análisis de patrones de comportamiento anómalos en la red (Mahboubi et al., 2024).
- Gestión de vulnerabilidades (*Vulnerability Management*): Pueden analizar grandes volúmenes de datos sobre vulnerabilidades de sistemas y aplicaciones, identificando aquellas más críticas y priorizando su remediación (Ahmed et al., 2016).
- Data centers: Se emplean para monitorizar y optimizar el rendimiento de los data centers, detectando patrones anómalos que puedan indicar actividad maliciosa.

- Seguridad en las redes: Se utilizan para la detección de intrusiones y ataques en las redes, analizando el tráfico de red en busca de comportamientos sospechosos.
- Identificación segura de usuarios (*Securing Authentication*): Pueden implementarse en sistemas de autenticación biométrica, como reconocimiento facial o de huellas dactilares, para mejorar la seguridad de los procesos de identificación (Fernandes et al., 2019).
- Privacidad de la información y compliance: Ayudan a detectar fugas de información confidencial y a garantizar el cumplimiento normativo en materia de protección de datos.
- Bloqueo de bots a partir de su comportamiento: Pueden identificar patrones de comportamiento característicos de bots maliciosos y automatizar su bloqueo en los sistemas.

Al adoptar técnicas de detección de anomalías, las organizaciones pueden mejorar significativamente su capacidad de respuesta y mitigación ante una amplia gama de amenazas cibernéticas, lo que les permite proteger de manera más eficaz sus sistemas, datos y activos críticos. La implementación de estos enfoques avanzados de detección de anomalías se ha convertido en una estrategia esencial para fortalecer la ciberseguridad y mantener la confianza de los usuarios y clientes.

Los algoritmos basados en la técnica de aprendizaje profundo están motivados por el campo de la inteligencia artificial, y tratan de emular la habilidad cognitiva del cerebro humano. Comúnmente estos algoritmos hacen uso de la estructura de datos conocida como red neuronal, a la cual se le han realizado modificaciones creando nuevos tipos de redes destinadas a trabajar con diferentes tipos de datos o funcionalidades específicas. Entre estas nuevas estructuras podemos mencionar:

AutoEncoders (AEs): Consiste una red neuronal artificial entrenada para entregar en su salida el mismo dato que se le introduce. Los codificadores automáticos han sido utilizados por empresas como PayPal para crear sistemas de detección de fraude mediante la extracción de características clave que determinan si una transacción es fraudulenta. Por otro lado, aquellos que se usan para la detección de anomalías son de gran utilidad en la industria bancaria para ayudar a automatizar la generación de algoritmos de recomendación de préstamos. Por ejemplo, si un banco tiene una gran cantidad de datos sobre sus clientes y préstamos, entonces estos datos se pueden usar para caracterizar cómo son los mejores préstamos (Mavikumbure et al., 2022).

Deep Neural Networks (DNN): Las Deep Neural Networks (DNN) son un tipo avanzado de redes neuronales artificiales que se caracterizan por tener múltiples capas ocultas entre la capa de entrada y la capa de salida. Esta profundidad en la arquitectura de la red les permite aprender representaciones cada vez más abstractas y complejas de los datos de entrada, lo que las convierte en herramientas poderosas para resolver problemas de gran complejidad. En

el ámbito de la ciberseguridad, las DNN han demostrado ser muy efectivas en diversas aplicaciones. Por ejemplo, pueden utilizarse para la detección de intrusiones, analizando patrones en el tráfico de red y la actividad del sistema con el fin de identificar comportamientos anómalos que puedan indicar la presencia de amenazas (Ahmad et al., 2021).

Restricted Boltzmann Machines (RBM): Las Restricted Boltzmann Machines (RBM) son un tipo de red neuronal no supervisada que se ha destacado por su utilidad en el campo de la ciberseguridad. Estos modelos emplean los datos de entrada, como registros de eventos de seguridad o tráfico de red, para generar estados probabilísticos que permiten identificar características latentes en el conjunto de datos. Esta capacidad de extraer patrones y relaciones ocultas en los datos es fundamental para la detección de amenazas avanzadas, como ataques de día cero o actividad maliciosa sofisticada. Las RBM pueden aprender representaciones compactas de los datos de entrada, lo que facilita la identificación de anomalías y comportamientos sospechosos. Además, estas redes neuronales se pueden utilizar como una etapa de pre-entrenamiento para modelos supervisados más complejos, mejorando así la eficacia de los sistemas de detección y respuesta a incidentes de seguridad.

Deep Belief Networks (DBN): Se caracterizan por estar modeladas a través de una composición de RBM. Esta arquitectura le confiere a las DBN la capacidad de extraer características relevantes de los datos de entrada de manera eficiente, lo que las hace idóneas para tareas de pre-entrenamiento en aplicaciones de ciberseguridad. A diferencia de otros modelos de aprendizaje profundo, las DBN requieren una fase de entrenamiento sin etiquetas, lo que les permite trabajar efectivamente con conjuntos de datos no supervisados, como registros de eventos de seguridad o tráfico de red. Esta característica las convierte en una herramienta valiosa para la detección de anomalías y la identificación de patrones de comportamiento malicioso dentro de los sistemas de información. Gracias a su capacidad de extraer características de alto nivel a partir de los datos, las DBN pueden ser utilizadas como un paso previo al entrenamiento de modelos supervisados más complejos, mejorando así el rendimiento general de los sistemas de detección y respuesta a incidentes de seguridad (Sohn, 2021).

Convolutional Neural Networks (CNN): Son un tipo de red neuronal profunda que se ha destacado por su excelente desempeño en el procesamiento y análisis de datos con estructura espacial o temporal, como imágenes y secuencias de video. En el campo de la ciberseguridad, las CNN han demostrado ser una herramienta muy valiosa para diversas aplicaciones (Kwon et al., 2018).

Una de las principales aplicaciones de las CNN en ciberseguridad es el análisis de malware. Estas redes neuronales pueden procesar muestras de malware, ya sea en formato de imagen (como capturas de pantalla de la interfaz del malware) o como secuencias de código, para identificar patrones y características que permitan clasificar y detectar

nuevas variantes de software malicioso. Gracias a su capacidad de aprender representaciones jerárquicas de los datos, las CNN pueden extraer automáticamente rasgos relevantes que faciliten la detección de amenazas.

Además, las CNN también se han utilizado con éxito en la detección de intrusiones en redes, analizando el tráfico de red en busca de patrones que puedan indicar actividad maliciosa. Al procesar los paquetes de datos como secuencias temporales, las CNN pueden identificar comportamientos sospechosos que puedan revelar intentos de ataques o intrusiones.

Otra aplicación interesante de las CNN en ciberseguridad es la clasificación de imágenes relacionadas con la seguridad, como capturas de pantalla de aplicaciones, interfaces de usuario o elementos de infraestructura. Esto puede ser útil para la detección de vulnerabilidades visuales o la identificación de amenazas basadas en el análisis de la interfaz gráfica.

Recurrent Neural Networks (RNN): Son un tipo de red neuronal que se caracteriza por su estructura dinámica y la capacidad de manejar información secuencial. A diferencia de las redes neuronales feedforward tradicionales, las RNN no tienen una estructura de capas definida, sino que permiten conexiones aleatorias entre las neuronas, incluso con la posibilidad de crear ciclos, lo que les confiere la habilidad de mantener una memoria a corto plazo. Esta característica las convierte en una herramienta muy valiosa para una amplia variedad de aplicaciones informáticas, incluyendo la predicción y el modelado de sistemas dinámicos. En el ámbito de la ciberseguridad, las RNN se pueden utilizar para reconocer o reproducir secuencias de eventos, lo que las hace especialmente adecuadas para problemas donde el factor temporal o el contexto son relevantes, como en la detección de intrusiones o el análisis de tráfico de red. Gracias a su capacidad de aprender patrones complejos a partir de datos secuenciales, las Recurrent Neural Networks han demostrado su efectividad en tareas como la identificación de comportamientos sospechosos, la predicción de ataques y la clasificación de malware (Ullah & Mahmoud, 2022).

Redes Generativas Adversariales (GAN): Son un tipo de modelo de inteligencia artificial que consta de dos redes neuronales, el generador y el discriminador, que se entrenan de manera adversarial. El generador crea datos sintéticos, como imágenes, textos o sonidos, que intentan imitar los datos reales de entrenamiento, mientras que el discriminador intenta distinguir entre los datos reales y los generados. A medida que ambas redes se entrenan simultáneamente, el generador mejora su capacidad para generar datos más realistas, mientras que el discriminador mejora su capacidad para diferenciar entre datos reales y generados. Este proceso de entrenamiento iterativo resulta en la generación de datos sintéticos de alta calidad que son indistinguibles de los datos reales para el discriminador. Las GANs han demostrado tener un excelente rendimiento, por este motivo, en la última década se han propuesto numerosos algoritmos de detección de anomalías basados en este tipo de red neuronal (Kimura, 2024).

Estas redes neuronales ofrecen una amplia gama de enfoques para la detección de anomalías en ciberseguridad, desde la identificación de patrones en datos estructurados hasta el análisis de imágenes y secuencias temporales. Su aplicación demuestra el potencial de la inteligencia artificial en la protección de sistemas y datos críticos contra amenazas cibernéticas.

Investigaciones relacionadas:

Modelado Probabilístico Basado en Aprendizaje Profundo para la Detección de Anomalías en el Tráfico de Red (Eguren et al., 2019)

Este trabajo se centra en la detección de anomalías en el tráfico de red utilizando aprendizaje profundo. Se exploran dos estrategias: detección por mal uso, que identifica patrones de comportamiento malicioso conocidos, y detección por anomalías, que modela el tráfico normal y considera cualquier desviación como comportamiento malicioso. El proyecto busca generar modelos probabilísticos para la detección de anomalías, utilizando redes neuronales profundas. Se desarrolla en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat y se divide en tres etapas: análisis preliminar del problema, desarrollo de un algoritmo para el reconocimiento de anomalías y experimentación. Al final, se espera fortalecer la línea de investigación en modelos probabilísticos relacionados con el tráfico de red, obtener una implementación funcional del modelo y aumentar la experiencia para aplicar modelos probabilísticos basados en aprendizaje profundo a nuevas líneas de investigación.

Diseño de un sistema de detección de intrusos (IDS) basada en técnicas supervisadas de anomalías mediante la aplicación de aprendizaje profundo (Villalba & Varón, 2023)

El artículo se centra en el diseño de un sistema de detección de intrusiones (IDS) utilizando técnicas supervisadas de anomalías y aprendizaje profundo. El sistema IDS propuesto utiliza el aprendizaje profundo para detectar una amplia gama de anomalías en los sistemas informáticos. Este enfoque ha permitido la modernización de los sistemas de detección de intrusiones, permitiendo a los algoritmos reconocer patrones y comportamientos novedosos. El proyecto se llevó a cabo utilizando la metodología Scrum, con varios incrementos o “sprints”. Estos incluyeron la recopilación de datos de tráfico de red en tiempo real, la limpieza y preprocesamiento de los datos, y la generación de un modelo de red neuronal para evaluar el rendimiento en un ambiente controlado. El sistema IDS propuesto se basa en firmas para identificar ataques conocidos y alertar sobre posibles intrusiones en tiempo real. También se aplicaron técnicas de aprendizaje automático al IDS para mejorar su capacidad de detección, permitiendo la identificación de patrones de ataques desconocidos y el análisis de anomalías en el tráfico de red.

Diseño de un sistema multiagentes híbrido basado en aprendizaje profundo para la detección y contención de ciberataques (Santiago & Allende, 2016)

En este artículo se diseña un sistema multiagentes híbrido basado en aprendizaje profundo para la detección y contención de ciberataques, expresa que debido al avance de tecnología la sociedad se expone a hackers maliciosos y fugas de información en las empresas. Las limitaciones a las soluciones de ciberseguridad provienen de tener un único punto de fallo, recursos de procesamiento limitados y la imposibilidad de adaptarse a los cambios de estrategia evidenciados en los ataques de última generación. Las empresas para mitigar los ataques de última generación han optado por utilizar cada vez más tecnología de diferentes fabricantes al mismo tiempo, lo que hace más complejo la gestión de la ciberdefensa.

Dentro del artículo los autores proponen un sistema de ciberseguridad distribuida multicapas y escalable compuesto por un conjunto de agentes con capacidad adaptativa que se apoyan en un sistema de aprendizaje automático avanzado, los agentes que dan solución se dividen en 3 capas:

- Capa de monitorización; tiene como objetivo obtener los parámetros necesarios para facilitar la identificación de acciones maliciosas que podrían considerarse ataques contra los activos de información de cualquier empresa.
- Capa de análisis: lleva a cabo las actividades de procesamiento de los parámetros obtenidos por los agentes de monitorización que son necesarios para determinar si los ordenadores en evaluación están enfrentados una situación de riesgo.
- Capa de supervisión: encargada de realizar el registro de actividades de los agentes de las capas anteriores y su posterior almacenamiento persistente en una base de datos de la cual pueden extraerse reportes referentes a los eventos e incidentes de seguridad detectados.

Las ventajas de las organizaciones frente a esta propuesta son: aprendizaje adaptativo que le permite evolucionar ante los cambios de estrategia de los ataques informáticos., mayor confiabilidad en la detección, capacidad distribuida., recuperación rápida de errores, tolerancia a fallos, cubrimiento de nuevos tipos de ataque.

A partir de lo antes expuesto se examinan algunas áreas clave de mejora y nuevas direcciones de investigación en la detección de anomalías en ciberseguridad usando aprendizaje profundo:

Interpretación de Modelos de Aprendizaje Profundo: Aunque los modelos de aprendizaje profundo han demostrado ser muy eficaces en la detección de anomalías, la interpretación de cómo y por qué estos modelos toman decisiones sigue siendo un desafío. Explorar técnicas para hacer que los modelos de aprendizaje profundo sean más

interpretables y comprensibles para los expertos en seguridad podría ser una dirección de investigación interesante. Esto podría incluir métodos para visualizar y explicar la lógica subyacente detrás de las decisiones del modelo.

Integración de Contexto y Conocimiento del Dominio: La detección de anomalías puede beneficiarse enormemente de la integración de contexto y conocimiento del dominio. Explorar métodos para incorporar información contextual, como la topología de la red, el comportamiento normal del usuario y las características específicas del sistema, en los modelos de detección de anomalías podría mejorar su precisión y capacidad de generalización.

Privacidad y Ética en la Detección de Anomalías: Con el aumento del uso de datos sensibles en la detección de anomalías, también es importante abordar preocupaciones relacionadas con la privacidad y la ética. Explorar técnicas para garantizar la privacidad de los datos durante el proceso de detección de anomalías y desarrollar marcos éticos para el uso de modelos de detección de anomalías en entornos sensibles podría ser una dirección de investigación relevante.

Conclusiones

Las conclusiones extraídas del trabajo abordan de manera integral la detección de anomalías en el ámbito de la ciberseguridad, con un enfoque particular en la aplicación de técnicas de aprendizaje profundo. La revisión realizada sobre los principales métodos y tecnologías empleadas en este campo ha permitido afirmar que la detección de anomalías desempeña un papel fundamental en la protección de sistemas y datos críticos frente a una amplia gama de amenazas cibernéticas. Esta técnica no solo permite identificar y mitigar actividades maliciosas y emergentes, sino que también se ha vuelto esencial en un entorno donde los ciberataques son cada vez más sofisticados.

El uso de modelos de aprendizaje profundo representa un avance significativo en la detección de anomalías, especialmente en el análisis de grandes volúmenes de datos. La flexibilidad y capacidad de adaptación de estos modelos los hacen especialmente valiosos en el entorno empresarial, donde la velocidad y precisión en la toma de decisiones son cruciales. Las diversas aplicaciones de las técnicas de aprendizaje profundo en la detección de anomalías, desde la identificación de patrones en datos estructurados hasta el análisis de imágenes y secuencias temporales, demuestran el potencial de la inteligencia artificial en la protección de sistemas y datos críticos contra amenazas cibernéticas.

Además, se destaca la importancia de la interpretación de los modelos de aprendizaje profundo y la integración de contexto y conocimiento del dominio para mejorar la precisión y capacidad de generalización de los modelos de detección de anomalías. Es fundamental abordar preocupaciones relacionadas con la privacidad y la ética en la detección de anomalías, especialmente con el aumento del uso de datos sensibles en este campo. Explorar técnicas

para garantizar la privacidad de los datos y desarrollar marcos éticos para el uso de modelos de detección de anomalías en entornos sensibles son áreas clave de mejora y futuras direcciones de investigación.

Conflictos de intereses

Los autores no poseen conflictos de intereses.

Contribución de los autores

1. Conceptualización: Mónica Peña Casanova
2. Curación de datos: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
3. Análisis formal: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
4. Investigación: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
5. Metodología: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
6. Administración del proyecto: Mónica Peña Casanova
7. Software: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
8. Supervisión: Mónica Peña Casanova
9. Validación: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
10. Visualización: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez
11. Redacción – borrador original: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez, Mónica Peña Casanova
12. Redacción – revisión y edición: Adiane Cueto Portuondo, Mónica Delgado Hernández, Heidy Rodríguez Malvarez, Mónica Peña Casanova

Financiamiento

La investigación no requirió fuente de financiamiento externa.

Referencias

Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, *60*, 708-713. https://www.sciencedirect.com/science/article/pii/S1877050915023479

Ahmad, Z., Shahid Khan, A., Nisar, K., Haider, I., Hassan, R., Haque, M. R., Tarmizi, S., & Rodrigues, J. J. (2021). Anomaly detection using deep neural network for IoT architecture. *applied sciences*, *11*(15), 7050. https://www.mdpi.com/2076-3417/11/15/7050

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, *60*, 19-31. https://www.sciencedirect.com/science/article/pii/S1084804515002891

Behrad, F., & Abadeh, M. S. (2022). An overview of deep learning methods for multimodal medical data mining. *Expert Systems with Applications*, *200*, 117006. https://www.sciencedirect.com/science/article/pii/S0957417422004249

Eguren, S., Catania, C., & Guerra, J. (2019). Modelado probabilístico basado en aprendizaje profundo para la detección de anomalías en el tráfico de red. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan).

Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., & Proença, M. L. (2019). A comprehensive survey on network anomaly detection. *Telecommunication Systems*, *70*, 447-489. https://link.springer.com/article/10.1007/s11235-018-0475-8

Gadal, S. M. A. M., & Mokhtar, R. A. (2017). Anomaly detection approach using hybrid algorithm of data mining technique. 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE),

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. *Complexity*, *2019*(1), 1306039. https://onlinelibrary.wiley.com/doi/abs/10.1155/2019/1306039

Kimura, T. (2024). Exploring the Frontier: Generative AI Applications in Online Consumer Behavior Analytics. *Cuadernos de Gestión*, 1-14. https://ojs.ehu.eus/index.php/CG/article/view/26905

Kwon, D., Natarajan, K., Suh, S. C., Kim, H., & Kim, J. (2018). An empirical study on network anomaly detection using convolutional neural networks. 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS),

Lan, K., Wang, D.-t., Fong, S., Liu, L.-s., Wong, K. K., & Dey, N. (2018). A survey of data mining and deep learning in bioinformatics. *Journal of medical systems*, *42*, 1-20. https://link.springer.com/article/10.1007/s10916-018-1003-9

Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *applied sciences*, *9*(20), 4396. https://www.mdpi.com/2076-3417/9/20/4396

Mahboubi, A., Luong, K., Aboutorab, H., Bui, H. T., Jarrad, G., Bahutair, M., Camtepe, S., Pogrebna, G., Ahmed, E., & Barry, B. (2024). Evolving techniques in cyber threat hunting: A systematic review. *Journal of Network and Computer Applications*, 104004. https://www.sciencedirect.com/science/article/pii/S1084804524001814

Mavikumbure, H. S., Wickramasinghe, C. S., Marino, D. L., Cobilean, V., & Manic, M. (2022). Anomaly detection in critical-infrastructures using autoencoders: A survey. IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society,

Narayana, M. S., Prasad, B., Srividhya, A., & Reddy, K. P. R. (2011). Data mining machine learning techniques–A study on abnormal anomaly detection system. *International Journal of Computer Science and Telecommunications*, *2*(6). https://www.academia.edu/download/85279382/p3_2_6.pdf

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., Malík, P., & Hluchý, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, *52*, 77-124. https://link.springer.com/article/10.1007/s10462-018-09679-z

Pang, G., Cao, L., & Aggarwal, C. (2021). Deep learning for anomaly detection: Challenges, methods, and opportunities. Proceedings of the 14th ACM international conference on web search and data mining,

Parmar, J. D., & Patel, J. T. (2017). Anomaly detection in data mining: a review. *International Journal*, *7*(4), 32-40. https://www.academia.edu/download/102599305/V7I4-0142.pdf

Santiago, E. J., & Allende, J. S. (2016). Diseño de un sistema multiagentes híbrido basado en aprendizaje profundo para la detección y contención de ciberataques. *REVISTA COLOMBIANA DE TECNOLOGIAS DE AVANZADA (RCTA)*, *2*(28), 115-123. https://ojs.unipamplona.edu.co/index.php/rcta/article/view/298

Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. I. (2020). Intrudtree: a machine learning based cyber security intrusion detection model. *Symmetry*, *12*(5), 754. https://www.mdpi.com/2073-8994/12/5/754

Sohn, I. (2021). Deep belief network based intrusion detection techniques: A survey. *Expert Systems with Applications*, *167*, 114170. https://www.sciencedirect.com/science/article/pii/S0957417420309088

Ullah, I., & Mahmoud, Q. H. (2022). Design and development of RNN anomaly detection model for IoT networks. *IEEE Access*, *10*, 62722-62750. https://ieeexplore.ieee.org/abstract/document/9777970/

Villalba, D. A. M., & Varón, D. F. M. (2023). Diseño de un sistema de detección de intrusos (IDS) basada en técnicas supervisadas de anomalías mediante la aplicación de aprendizaje profundo. *Encuentro Internacional de Educación en Ingeniería*. https://acofipapers.org/index.php/eiei/article/view/2877

Wang, S., Cao, J., & Philip, S. Y. (2020). Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*, *34*(8), 3681-3700. https://ieeexplore.ieee.org/abstract/document/9204396/

Yadav, N., & Souza, M. D. (2024). Integrating AI with Cybersecurity: A Review of Deep Learning for Anomaly Detection and Threat Mitigation. *Nanotechnology Perceptions*, 1756-1785. http://nano-ntp.com/index.php/nano/article/view/3007

Yaseen, A. (2023). The role of machine learning in network anomaly detection for cybersecurity. *Sage Science Review of Applied Machine Learning*, *6*(8), 16-34. https://journals.sagescience.org/index.php/ssraml/article/view/126