1-1-2024

Detección, análisis y caracterización de anomalías en los logs de la Armada Nacional mediante métodos de inteligencia artificial

Detección, análisis y caracterización de anomalías en los logs de la Armada Nacional mediante métodos de inteligencia artificial

Autor:

Yasmin Johanna García Gaviria

Tutor:

PhD. Uram Anibal Sosa

Maestría en Ciencia de Datos

Universidad Icesi

2024

Detección, análisis y caracterización de anomalías en los logs de la Armada Nacional mediante métodos de machine learning

Autor:

Yasmin Johanna García Gaviria

Armada Nacional de Colombia

Comando de Cibernética Naval

2024

Tabla de Contenido

Introducción

El ámbito de la ciberseguridad ha experimentado un incremento notable en la frecuencia y sofisticación de los ciberataques en los últimos años, sobre todo aquellos destinados a infraestructuras críticas cibernéticas. El aumento no solo ha afectado a sectores como la energía, la salud y las finanzas, sino que también ha convertido a entidades militares y de defensa en objetivos prioritarios para estos actores maliciosos. En este contexto, la Armada Nacional se ha transformado en un objetivo constante, representando una amenaza creciente para la seguridad cibernética y la integridad de las operaciones de las Fuerzas Armadas, por lo tanto, para la seguridad nacional.

La infraestructura critica cibernética naval (ICCN), un componente estratégico para la defensa y el funcionamiento efectivo de la Armada se enfrenta a desafíos constantes en términos de detección, prevención y mitigación de estos ataques. La necesidad de abordar este problema de manera integral y proactiva se ve agravada por la sofisticación y frecuencia crecientes de los ciberataques. La seguridad de la infraestructura del COCIB no solo es vital para proteger los activos y datos críticos de la Armada, sino también esencial para garantizar la efectividad de las operaciones navales y la seguridad nacional en general.

La integración de técnicas avanzadas de Ciencia de Datos e Inteligencia Artificial se presenta como una solución prometedora para abordar estos desafíos. Estas tecnologías permiten el análisis de grandes volúmenes de datos, la detección de anomalías, patrones y la predicción de comportamientos maliciosos que podrían pasar desapercibidos con métodos tradicionales. Los algoritmos de aprendizaje automático y otras técnicas de IA mejoran la capacidad de detección y permiten una respuesta más proactiva y eficiente a posibles ataques.

Al automatizar tareas de monitoreo y análisis, se liberan recursos, permitiendo que el personal se concentre en actividades estratégicas, como la elaboración de políticas de seguridad y la planificación de medidas preventivas. Implementar estos métodos en los datos provenientes de herramientas tecnológicas al servicio de la armada, es esencial para fortalecer la postura de ciberseguridad de la Armada de Colombia y garantizar la protección de sus operaciones críticas. La adopción de estas tecnologías avanzadas no solo representa una evolución en las capacidades defensivas del comando, sino que también sienta un precedente para la modernización de la ciberseguridad en otras entidades críticas del país.

El uso de las ciencias de datos y la inteligencia artificial de datos permitirá al COCIB poder aumentar su análisis proactivo para mitigar ataque de naturaleza cibernética

Palabras claves

Anomalía, Ciberseguridad, Ciberespacio, Cibernética Naval, Seguridad informática, Vulnerabilidad, COCIB, Ciencia de Datos, Inteligencia Artificial, Comando y Control (CC2).

Identificación del Problema

“El aumento exponencial de ciberataques a las ICCN con un mayor grado de frecuencia y sofisticación.”

En los últimos años, las FFMM a nivel global han experimentado un aumento significativo en la cantidad y complejidad de los ciberataques contra su infraestructura crítica. Esta tendencia está causando una preocupación por el impacto potencial de estos ataques en la ciberseguridad y la integridad de las operaciones militares, así como sus implicaciones directas para la seguridad nacional. La infraestructura del ICCN, como componente estratégico para la defensa y la operación efectiva de la Armada, enfrenta constantes desafíos en la detección, prevención y mitigación de ciberataques, que van desde pequeñas intrusiones hasta ataques sofisticados de Comando y Control (C&C). La creciente sofisticación y frecuencia de estas ciberamenazas plantean desafíos importantes, lo que pone de relieve la necesidad urgente de abordar este problema de manera integral para permitir una respuesta más rápida y eficaz a los incidentes de seguridad.

La seguridad de las infraestructuras del comando es vital no solo para proteger los activos y datos críticos de la Armada, sino también para garantizar la eficiencia y continuidad de las operaciones navales y, en última instancia, la seguridad nacional en su conjunto. Por tanto, es necesario estudiar en profundidad los factores que contribuyen al aumento de los ciberataques dirigidos contra el ICCN, así como desarrollar estrategias efectivas de detección, prevención y respuesta frente a estas amenazas.

Comprender la naturaleza y el alcance de estos ataques es esencial para implementar medidas de protección adecuadas y fortalecer la postura de ciberseguridad de la Armada de Colombia. En particular, es fundamental que estas acciones sean proactivas y no reactivas, anticipando los riesgos y minimizando su impacto antes de que se comprometa la seguridad y las operaciones navales. La implementación de estrategias avanzadas basadas en ciencia de datos e inteligencia artificial es crucial para mejorar las capacidades de detección y

respuesta, asegurando así la resiliencia y operatividad del comando y, en consecuencia, de las FFMM.

En este sentido, las preguntas que guían este proyecto son:

1. ¿Qué tipos de datos generados por la Armada son relevantes para el análisis de ciberataques, y cómo pueden ser preprocesados y estructurados para su uso en modelos de machine learning?
2. ¿Cómo caracterizar las anomalías detectadas por el análisis combinado de ciencias de datos e inteligencia artificial que sean relevantes para la armada nacional?
3. ¿Cómo implementar soluciones basadas en Inteligencia Artificial para detectar anomalías en los logs de dispositivos críticos, con énfasis en ataques de Comando y Control (C&C), garantizando su eficacia, escalabilidad y adaptabilidad a las necesidades del SOC de la Armada Nacional?

Justificación

En un entorno donde las amenazas cibernéticas dirigidas a infraestructuras críticas se multiplican y se vuelven cada vez más sofisticadas, se subraya la necesidad urgente de fortalecer las capacidades de ciberseguridad del Centro Operaciones de Ciberseguridad SOC**.** La seguridad de esta infraestructura es esencial no solo para proteger los activos y datos críticos de la Armada, sino también para garantizar la efectividad y continuidad de sus operaciones navales y, en última instancia, la seguridad nacional en su conjunto. Por lo tanto, es crucial desarrollar estrategias efectivas para identificar y mitigar posibles riesgos y vulnerabilidades que podrían convertirse en ataques.

El crecimiento exponencial de las amenazas cibernéticas plantea un desafío sin precedentes para la seguridad informática en todas las industrias, especialmente en aquellas relacionadas con la seguridad nacional. Los ataques, cada vez más sofisticados y difíciles de detectar, ponen en riesgo la integridad y confidencialidad de la información crítica, así como la continuidad operativa de las organizaciones, y en particular, de las naciones.

Los ataques de tipo Comando y Control representan una amenaza particularmente grave, ya que permiten a los atacantes establecer canales de comunicación encubiertos dentro de las redes comprometidas, facilitando el robo de información, el sabotaje de sistemas y la coordinación de actividades maliciosas. Para hacer

frente a estas amenazas, es imperativo adoptar un enfoque proactivo que incluya la identificación y caracterización de patrones de comportamiento malicioso en los registros de dispositivos críticos.

La aplicación de técnicas avanzadas de Ciencia de Datos se presenta como una solución prometedora para mejorar la detección temprana de ciber amenazas. Los métodos de Ciencia de Datos ofrecen una capacidad robusta para analizar grandes volúmenes de datos de forma rápida y precisa. Mediante el uso de algoritmos de Machine Learning y otras técnicas derivadas de la Inteligencia Artificial, es posible identificar patrones y anomalías que podrían pasar desapercibidos para los enfoques tradicionales de detección, que por lo general dependen de la capacidad humana. Esta capacidad predictiva y analítica se convierte en un recurso invaluable para anticipar y responder eficazmente a las amenazas cibernéticas emergentes.

Además, la integración de métodos de Ciencia de Datos en la infraestructura de ciberseguridad existente no sólo mejora la capacidad de detección, sino que también permite una respuesta más rápida y eficiente ante posibles ataques. Al automatizar tareas de monitoreo y análisis, los sistemas basados en Ciencia de Datos apoyan y liberan esfuerzos humanos, permitiendo que el personal se concentre en actividades de mayor valor estratégico, como la elaboración de políticas de seguridad y la planificación de medidas preventivas.

En este sentido, el presente trabajo no solo abordará los desafíos actuales de ciberseguridad que enfrenta el COCIB, sino que también contribuirá a una defensa más sólida y resiliente de la Armada nacional.

Objetivos

Objetivo general

Implementar modelos de inteligencia artificial para aportar a la protección de los activos y operaciones esenciales del SOC de la Armada Nacional, mediante la identificación y caracterización de patrones de comportamiento anómalos en los logs de los dispositivos críticos enfocados a ataques de comando y control.

Objetivos específicos

- Realizar un Análisis Exploratorio de Datos (EDA) de los logs de dispositivos críticos del COCIB para comprender la estructura, características y patrones de comportamiento normal y anómalo en los datos, con el fin de identificar y clasificar los patrones de comportamiento malicioso más comunes asociados a ataques de Comando y Control (C&C).
- Identificar y caracterizar anomalías en los logs de los dispositivos críticos del SOC de la Armada Nacional que indiquen comportamientos maliciosos y que puedan comprometer la seguridad de los activos y operaciones
- Desarrollar e implementar soluciones basadas Inteligencia Artificial para la detección de anomalías en los logs generados por dispositivos críticos del SOC relacionado con ataques de comando y control

Marco Teórico

El incremento y la sofisticación de los ataques en el ciberespacio han puesto en alerta a las agendas nacionales, destacando la necesidad urgente de fortalecer las defensas cibernéticas. La Agencia Europea de Ciberseguridad (ENISA) ha revelado un análisis exhaustivo de más de 2.500 ciberataques mayores entre julio de 2022 y junio de 2023, donde el 19% se dirigieron contra administraciones públicas, subrayando la vulnerabilidad en este sector. Este aumento en la sofisticación de las herramientas disponibles para perpetrar ciberataques se acompaña de una creciente tensión geopolítica, mientras que la continua innovación en la industria digital plantea nuevos desafíos para la ciberseguridad. Además, la evolución tecnológica exige el desarrollo de nuevos marcos regulatorios para garantizar una protección adecuada, desde la fijación de reglas para la implementación de Inteligencia Artificial hasta el establecimiento de estándares de ciberseguridad.

Se observa, en general, que tanto en el ámbito corporativo como en el sector público existen algunas insuficiencias en cuanto a la cultura de la ciberseguridad. Esto evidencia una falta generalizada de conciencia y participación en las prácticas recomendadas para protegerse en el entorno digital, así como la carencia de protocolos adecuados para enfrentar los incidentes y recuperarse de los mismos,

especialmente preocupante en entornos con alta rotación de personal. Estos desafíos, considerados críticos, incluyen la dificultad para cubrir los perfiles necesarios para la gestión de la ciberseguridad, la concentración del poder digital, la escasa implementación de estándares internacionales, la ausencia de una estructura nacional de gobierno ejecutivo, la falta de arquitectura resiliente en las organizaciones y la dificultad para perseguir a los ciberdelincuentes. Ante este panorama, es esencial que tanto las organizaciones corporativas como las entidades del sector público aumenten su atención y dedicación a la gestión de estos riesgos para garantizar su seguridad digital y proteger sus activos críticos.

En este contexto, el COCIB ha reconocido la urgencia de implementar modelos de Ciencia de Datos e Inteligencia Artificial centrados en la detección de amenazas cibernéticas. Estos modelos tienen como objetivo mejorar la predicción, análisis, procesamiento, visualización y extracción de características de eventos potencialmente peligrosos, utilizando registros provenientes de diversas herramientas de monitoreo cibernético y de inteligencia de amenazas. Este macroproyecto se dividirá en varios subproyectos interconectados que abordarán diferentes aspectos de la ciberseguridad. Se aprovecharán las capacidades del aprendizaje automático y del aprendizaje profundo para generar conocimientos más precisos y efectivos en la monitorización de eventos que representen un riesgo real.

Ciberseguridad

La ciberseguridad abarca el conjunto de medidas, tecnologías y prácticas diseñadas para proteger los sistemas, redes y datos de ataques maliciosos en el ciberespacio. Esta busca garantizar la confidencialidad, integridad y disponibilidad de la información, salvaguardando tanto a individuos como a organizaciones de amenazas como el robo y secuestro de datos, el fraude, el espionaje y la denegación de servicios. Si bien no se tiene un momento exacto donde se haya hablado de ciberseguridad por primera vez, Thoma Creeper (1975) marcó el inicio de la preocupación por la seguridad en los sistemas informático, si bien diversos autores han marcado internacionalmente las pautas de la ciberseguridad informática, algunos autores destacados en este campo incluyen a Bruce Schneier, reconocido por su enfoque en la criptografía y la seguridad informática, y Dan Kaminsky, conocido por sus importantes contribuciones en la detección y resolución de vulnerabilidades en internet, experto en seguridad de redes y descubridor de vulnerabilidades críticas.

Análisis de la ciberseguridad en Colombia

Colombia enfrenta desafíos significativos en materia de ciberseguridad, como lo evidencian los numerosos ataques cibernéticos reportados en los últimos años. Según la revista Forbes, para el año 2024, Colombia ha ocupado por segunda vez el segundo puesto en ser el país de América latina con más ciberataques[1]. La falta de conciencia y capacitación en seguridad digital, junto con la rápida adopción de tecnologías, aumentan la vulnerabilidad del país. Sin embargo, se han realizado avances importantes, como la creación del Centro de Respuesta a Incidentes Cibernéticos de Colombia (ColCERT) y la implementación de políticas y estrategias para fortalecer la ciberseguridad a nivel nacional.

En este sentido, este proyecto se enfoca en fortalecer la ciberseguridad del Comando de Cibernética Naval, desarrollando estrategias y herramientas para prevenir y mitigar los ataques cibernéticos más comunes que amenazan la integridad de sus sistemas y datos. Se centrará en identificar vulnerabilidades, implementar medidas de protección y establecer protocolos de respuesta ante incidentes para garantizar la seguridad de la información crítica y la continuidad de las operaciones navales.

Marco legal

La Armada Nacional de la República de Colombia (ARC) hace parte de una de las FFMM, su principal objetivo es contribuir y garantizar la defensa del Estado y las zonas marítimas de su jurisdicción mediante la aplicación de su poder naval.[2] Es la fuerza que deficiente los ríos, lagos, mares y océanos de Colombia.

La ARC se extiende a lo largo del territorio colombiano mediante una red de fuerzas Navales estratégicamente distribuidas. Esta red incluye un número de bases navales y fluviales en ambos litorales del país, además de múltiples bases fluviales que se encuentran ubicadas estratégicamente a lo largo del territorio nacional.

La Armada de la República de Colombia ejerce su presencia a lo largo y ancho del territorio nacional a través de una red estratégica de Fuerzas Navales y bases. Esta compleja red se compone de:

[1]Forbes. (2024). *Colombia es el país con más ataques de ciberseguridad en Latinoamérica*. Recuperado de https://forbes.co/2024/02/28/tecnologia/colombia-es-el-pais-con-mas-ataques-de-ciberseguridad-en-latinoamerica Recuperado el 7 de junio de 2024

[2]Armada Nacional de Colombia. (2021). Armada de Colombia. Recuperado el 20 de Mayo de 24 de https://www.armada.mil.co/node/64757

Fuerzas Navales:

- **Fuerza Naval del Caribe (FNC):** Con jurisdicción en el mar Caribe colombiano.
- **Fuerza Naval del Pacífico (FNP):** Con jurisdicción en el océano Pacífico colombiano.
- **Fuerza Naval del Oriente (FNO):** Con jurisdicción en los ríos y afluentes de la Orinoquía colombiana.
- **Fuerza Naval del Amazonas (FNA):** Con jurisdicción en los ríos y afluentes de la Amazonía colombiana.

Asi mismo, la red se complementa con 6 bases navales distribuidas asi:

1. **Base Naval ARC "Bolívar" (BN1):** Ubicada en Cartagena, principal base naval en el Caribe. Base Naval ARC Bahía Málaga - BN2, cerca de Buenaventura
2. **Base Naval ARC "Bahía Málaga" (BN2):** Ubicada cerca de Buenaventura, principal base naval en el Pacífico.
3. **Base Naval ARC "Leguízamo" (BN3):** Ubicada cerca de Puerto Leguízamo, en el departamento del Putumayo, con presencia en el sur del país.
4. **Base Naval ARC "San Andrés" (BN4):** Ubicada en la isla de San Andrés, con jurisdicción en el archipiélago de San Andrés, Providencia y Santa Catalina
5. **Base Naval ARC "Orinoquía" (BN5):** Ubicada en Puerto Carreño, en el departamento del Vichada, con presencia en la región de la Orinoquía
6. **Base Naval ARC "Bogotá" (BN6):** Ubicada en Bogotá, sede del Comando de la Armada

El Comando de Cibernética Naval

El comando de cibernética naval o COCIB, recibe su denominación, sigla y estructura en la resolución 127 del año 2021[3]:

Artículo 165° Cambiar denominación, sigla y modificar la estructura interna de la Dirección de Cibernética Naval, sigla (DICIB), por Comando de Cibernética Naval, sigla (COCIB), orgánico de la Jefatura de Inteligencia Naval y su organización se sujetará a lo establecido en la Tabla de Organización y Equipo TOE (No. 3-03-06-07-00-21).

1. [3]Ministerio de Defensa Nacional. (2021). Resolución No. 127 de 2021. Recuperado el día 20 de mayo. de 24 de http://marinanet.armada.mil.co

El COCIB se fundamenta en dar respuesta a la necesidad de fortalecer y desarrollar operaciones para la defensa y seguridad nacional particularmente en el ciberespacio. Centrando su atención en aquellas que salvaguarden la seguridad cibernética de la Armada Nacional, protegiendo sus activos digitales y garantizando la operación continua. Sus funciones principales se describen a continuación.

1. **Monitoreo y Análisis:** Supervisa de manera continua los sistemas de información y comunicaciones de la Armada para detectar y analizar posibles amenazas cibernéticas.
2. **Respuesta a Incidentes:** Coordina la respuesta ante incidentes cibernéticos, proporcionando asistencia técnica y operativa para mitigar los impactos y recuperar la normalidad en caso de ataques o intrusiones.
3. **Protección de la Infraestructura:** Implementa medidas de seguridad cibernética para proteger la infraestructura crítica de la Armada, incluyendo sistemas de Comando y Control, comunicaciones, y otros activos digitales.
4. **Gestión de Riesgos:** Evalúa y gestiona los riesgos cibernéticos, identificando vulnerabilidades y estableciendo medidas preventivas para garantizar la integridad, confidencialidad y disponibilidad de la información

Debido a la necesidad de una comprensión más profunda, el equipo COCIB y los autores del proyecto han establecido una colaboración bilateral para enfocar y desarrollar su entendimiento.

Analítica de datos e inteligencia artificial en ciberseguridad

La ciencia de datos es un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas científicos para extraer conocimiento de datos estructurados y no estructurados. Se basa en principios y técnicas de diversas disciplinas, como matemáticas, estadística, informática y disciplinas específicas del dominio. La ciencia de datos es fundamental para abordar problemas complejos en diversas industrias, incluyendo la ciberseguridad, donde se utiliza para detectar y mitigar amenazas.

En el actual panorama de amenazas cibernéticas, la ciencia de datos y la inteligencia artificial (IA) se han vuelto esenciales para la ciberseguridad por varias razones. En primer lugar, el volumen, la velocidad y la variedad de datos generados por los sistemas o aplicativos modernos podrían ser considerados como big data por lo abrumadores para los métodos tradicionales de análisis. La ciencia de datos proporciona las herramientas y técnicas para extraer información valiosa de estos datos, identificando patrones, anomalías y tendencias que pueden indicar comportamientos maliciosos. Asi mismo, la IA permite automatizar la detección y respuesta a amenazas, liberando a los analistas de seguridad de tareas repetitivas

y permitiéndoles enfocarse en amenazas más complejas. Los algoritmos de IA pueden aprender de datos históricos y en tiempo real para identificar y clasificar amenazas con mayor precisión y rapidez que los métodos tradicionales. Finalmente, la IA permite la detección proactiva de amenazas, anticipándose a los ataques en lugar de simplemente reaccionar a ellos. Mediante el análisis de patrones y comportamientos, la IA puede identificar amenazas emergentes y predecir futuros ataques.

Los lineamientos para aplicar la Ciencia de Datos (CD) y la Inteligencia Artificial (IA) en el enfoque de Ciberseguridad se enfocan en la detección y prevención de ataques comunes, tales como:

Detección de Intrusos: Se emplean modelos predictivos, entrenados con datos históricos y análisis de comportamiento, para identificar accesos no autorizados y actividades sospechosas en tiempo real. Esto permite una respuesta rápida ante posibles amenazas y minimiza el impacto de intrusiones.

Análisis de Malware: Mediante técnicas de aprendizaje automático, como el análisis de clústeres y la clasificación, se analiza el código y el comportamiento de archivos para detectar malware de forma automatizada. Este análisis se basa en características como firmas, patrones de comportamiento y técnicas de ofuscación, lo que permite identificar nuevas variantes de malware de forma más eficiente.

Monitoreo de Redes: El análisis de datos en tiempo real del tráfico de red permite identificar anomalías y patrones sospechosos, como picos de tráfico inusuales o conexiones desde ubicaciones inesperadas. Esto facilita la detección temprana de ataques DDoS y otras amenazas a la red, permitiendo la implementación de medidas de mitigación de forma oportuna.

Análisis de Riesgos: Se evalúa la vulnerabilidad de sistemas y redes mediante el análisis de datos de seguridad, como registros de eventos, configuraciones de seguridad y resultados de pruebas de penetración. La IA puede predecir posibles vectores de ataque basándose en datos históricos y patrones de ataques previos, lo que permite priorizar la implementación de medidas de seguridad y optimizar la asignación de recursos

Algoritmos de Aprendizaje Automático sugeridos

En el contexto de la ciberseguridad, varios algoritmos de aprendizaje automático pueden ser aplicados para la detección y mitigación de amenazas:

1. **Detección de anomalías no supervisadas:** es un campo del aprendizaje automático que se enfoca en identificar patrones inusuales o "anómalos" en

un conjunto de datos sin la ayuda de etiquetas o ejemplos predefinidos de lo que constituye una anomalía. Este enfoque es comúnmente utilizado cuando no se tiene información previa sobre qué puntos son normales y cuáles son anómalos, y se aplica en muchos contextos como la detección de fraudes, seguridad cibernética, monitoreo de redes, y calidad de manufactura, entre otros.

Ventajas:

- No requiere etiquetas: Puedes trabajar con conjuntos de datos grandes sin la necesidad de tener datos etiquetados, lo cual es especialmente útil en muchos problemas reales donde etiquetar los datos es difícil y costoso.
- Versatilidad: Puede detectar anomalías desconocidas. Esto es especialmente importante en áreas como la ciberseguridad, donde los tipos de ataques cambian constantemente.

Desventajas:

- Falsos Positivos: Los modelos de detección de anomalías no supervisada tienden a tener una alta tasa de falsos positivos porque no tienen ejemplos claros de lo que es "normal" y lo que es "anómalo".
- Ajuste Difícil: Puede ser difícil ajustar los parámetros, como el umbral de detección, para reducir la cantidad de falsos positivos y falsos negativos

Existen varios métodos y técnicas para realizar la detección de anomalías no supervisada

Métodos Basados en la Densidad:

- **Local Outlier Factor (LOF)**: Este algoritmo mide la densidad local de cada punto y lo compara con la densidad de sus vecinos. Si un punto tiene una densidad significativamente menor que sus vecinos, se considera una anomalía.
- Estos métodos son útiles para detectar puntos que están en áreas de baja densidad en comparación con el resto de los datos.

Modelos Basados en Árboles:

- **Isolation Forest**: Es un método que utiliza árboles de decisión para dividir el espacio de los datos aleatoriamente. Los puntos que se aíslan rápidamente (con menos particiones) se consideran anómalos. Es eficiente en cuanto a

memoria y procesamiento, por lo que es adecuado para grandes volúmenes de datos.

Modelos de Reducción de Dimensionalidad:

- **PCA (Principal Component Analysis)**: Se puede utilizar para reducir la dimensionalidad de los datos y encontrar las componentes principales. Luego, se examinan los puntos que no se pueden representar bien en esas componentes principales, y se consideran anomalías.

2. **Redes Neuronales:** Son modelos inspirados en la estructura del cerebro humano, compuestos por **neuronas artificiales** organizadas en capas. Las redes neuronales son particularmente útiles en tareas de **detección de patrones complejos**, clasificación y regresión, debido a su capacidad para aprender **representaciones no lineales** de los datos.

   Las **redes neuronales profundas** (Deep Learning) cuentan con **múltiples capas ocultas**, lo que les permite capturar **relaciones sutiles y de alto nivel** en grandes volúmenes de datos. En el contexto de la **detección de anomalías**, se pueden utilizar técnicas como los **Autoencoders**, que son redes neuronales diseñadas para reconstruir sus entradas. Estas redes son entrenadas con datos normales, de modo que cuando reciben **datos anómalos**, su error de reconstrucción aumenta significativamente, lo que facilita la **detección de comportamientos inusuales**.

   Algunas de las librerías más utilizadas para implementar redes neuronales en **Python** son:

   - **TensorFlow**: Una librería desarrollada por Google que permite construir y entrenar redes neuronales tanto simples como profundas. Ofrece flexibilidad y escalabilidad, ideal para la implementación de modelos complejos.
   - **Keras**: Una API de alto nivel que funciona sobre TensorFlow y simplifica el desarrollo de redes neuronales mediante un enfoque modular y fácil de usar.
   - **PyTorch**: Una librería de código abierto desarrollada por Facebook que facilita la creación de redes neuronales dinámicas y es muy popular para **investigación** y **prototipado rápido**.

- **Scikit-learn**: Aunque no está específicamente diseñada para Deep Learning, se puede utilizar para tareas de preprocesamiento, validación y evaluación de modelos en combinación con las librerías mencionadas. Por ejemplo, para un **Autoencoder** simple en Keras se pueden definir dos componentes: un **codificador** que reduce la dimensionalidad de los datos y un **decodificador**

que intenta reconstruirlos. La diferencia entre la entrada original y la reconstrucción (error de reconstrucción) es una medida clave para identificar anomalías.

En este sentido se hace necesario tomar decisiones de la mano de los expertos para identificar cómo la combinación de algoritmos de machine learning e inteligencia artificial nos permitirá cumplir el objetivo del proyecto.

Estado del arte

¿Cómo va la ciberseguridad en la región?

La ciberseguridad y la ciberdefensa se han convertido en aspectos cruciales para la seguridad nacional en muchos países, y Colombia no es una excepción. Ante la creciente sofisticación de los ataques cibernéticos y la necesidad de proteger al Estado y su infraestructura crítica, Colombia busca establecer una política nacional integral de ciberseguridad y ciberdefensa para contrarrestar el aumento de amenazas informáticas. Los documentos CONPES 3701, 3854 y 3995, junto con el Decreto 338 de 2022, establecen el marco de políticas y lineamientos para fortalecer la seguridad digital en Colombia. A continuación, se presenta una síntesis de cada uno:

CONPES 3701 (2011): Lineamientos de Política para Ciberseguridad y Ciberdefensa

Este documento tiene como objetivo fortalecer la capacidad del Estado para enfrentar amenazas en el ámbito cibernético. Propone la creación de instancias para prevenir, coordinar y atender incidentes cibernéticos, así como conformar organismos con capacidad técnica y operativa en seguridad digital.

CONPES 3854 (2016): Política Nacional de Seguridad Digital

Busca crear condiciones para que diversas partes interesadas gestionen riesgos de seguridad digital en sus actividades socioeconómicas, fomentando la confianza en el entorno digital. Introduce un enfoque de gestión de riesgos y promueve la cooperación nacional e internacional en seguridad digital.

CONPES 3995 (2020): Política Nacional de Confianza y Seguridad Digital

Establece medidas para desarrollar la confianza digital mediante la mejora de la seguridad digital, con el fin de que Colombia sea una sociedad incluyente y competitiva en el ámbito digital. Fortalece capacidades y actualiza el marco de gobernanza en seguridad digital, adoptando modelos con énfasis en nuevas tecnologías.

Decreto 338 de 2022

Este decreto adiciona el Título 21 a la Parte 2 del Libro 2 del Decreto Único 1078 de 2015, estableciendo lineamientos generales para fortalecer la gobernanza de la seguridad digital en Colombia. Crea el Modelo y las instancias de Gobernanza de Seguridad Digital, y define responsabilidades para entidades públicas en la implementación de políticas de seguridad digital.

En este sentido, se hace un diagnóstico de la situación actual, identificando debilidades en la capacidad del Estado para enfrentar estas amenazas, la falta de coordinación interinstitucional y la necesidad de fortalecer la legislación y la cooperación internacional en esta materia.

Para abordar estos problemas, se propone la creación de varias instancias, incluyendo una Comisión Intersectorial para establecer políticas, un Grupo de Respuesta a Emergencias Cibernéticas (colCERT) para coordinar acciones, un Comando Conjunto Cibernético (CCOC) para la defensa y un Centro Cibernético Policial (CCP) para la seguridad.

Además, se hace hincapié en la importancia de la capacitación especializada en ciberseguridad y ciberdefensa, así como en el fortalecimiento de la legislación y la cooperación internacional en estas áreas. El documento también incluye un plan de acción detallado y un presupuesto para la implementación de estas políticas

El artículo "Ciberseguridad, un desafío para las Fuerzas Militares colombianas en la era digital" (Peña Suárez, 2023) analiza los retos que enfrentan las Fuerzas Militares de Colombia en el ámbito de la ciberseguridad. El autor destaca la importancia de este tema debido a la creciente dependencia de las tecnologías digitales y al aumento de las amenazas en el ciberespacio, haciendo especial enfasis es fundamental que las Fuerzas Militares colombianas desarrollen estrategias y líneas de acción para aumentar sus capacidades defensivas y ofensivas en el ciberespacio, y que establezcan alianzas estratégicas con otros actores para proteger la infraestructura crítica del país.

Por otro lado, autores nacionales como Cujabante Villamil, Bahamón Jara, Prieto Venegas y Quiroga Aguilar (2020) profundizan en el desarrollo institucional de la ciberseguridad y la ciberdefensa en Colombia, haciendo énfasis en cómo este desarrollo ha impactado las relaciones cívico-militares. Su análisis abarca la

evolución de la política de seguridad digital en el país y la creciente participación de diversos actores, tanto civiles como militares, en la gestión del riesgo cibernético. Destacan especialmente la redefinición de las relaciones entre civiles y militares en el ámbito de la ciberdefensa y la ciberseguridad.

Otro autor que aborda este tema es Cortés Borrero (2015), quien realiza un análisis detallado del estado actual de la política pública de ciberseguridad y ciberdefensa en Colombia. En su artículo, examina los desafíos emergentes en la era digital y la implementación de políticas públicas para hacer frente a estos desafíos, involucrando a diversos sectores de la sociedad.

Además, Urcuqui López, Navarro Cadavid, Osorio Quintero y García Peña (2018) exploran la aplicación de la ciencia de datos en el campo de la ciberseguridad en su libro "Ciberseguridad: un enfoque desde la ciencia de datos". Estos autores investigan la viabilidad de utilizar la ciencia de datos para desarrollar soluciones a problemas en ciberseguridad, utilizando proyectos de investigación específicos centrados en la detección de malware en dispositivos móviles y el control de defensa en páginas web como ejemplos ilustrativos.

Por otro lado, en un artículo de reflexión, Cortés Borrero (2015) analiza el estado actual de la política pública de ciberseguridad y ciberdefensa en Colombia. El autor examina los desafíos emergentes en la era digital y la implementación de políticas públicas para hacer frente a estos desafíos, involucrando a diversos sectores de la sociedad.

Por su parte E. Mayer et all (2014) discute los desafíos de analizar grandes volúmenes de datos de registro en entornos empresariales, donde la inconsistencia y la complejidad de los registros dificultan la detección de amenazas y menciona a Beehive, el cual utiliza un enfoque de aprendizaje no supervisado para identificar incidentes de comportamiento anómalo del host y agrupar hosts con comportamientos similares.

Por otra parte, Alazab, M., Khraisat, A., & Kumar, R. (2020) en Machine Learning-Based Cyber Threat Detection: A Comprehensive Review. Este artículo proporciona una revisión exhaustiva de los enfoques de detección de amenazas cibernéticas basados en aprendizaje automático, evaluando su eficacia, desafíos y tendencias emergentes.

Choudhary, A., Saini, J. K., & Singh, M. (2020) realizan un estudio exhaustivo que revisa varios algoritmos de aprendizaje automático utilizados para la detección de amenazas cibernéticas, destacando su efectividad en la identificación y mitigación de riesgos. En el mismo sentido Swami (2020) presenta un análisis detallado de las

técnicas de aprendizaje automático aplicadas en el campo de la ciberseguridad, identificando sus aplicaciones prácticas y desafíos inherentes.

Finalmente, Ahmed, M., Mahmood, A. N., & Hu, J. (2021) presentan una investigación que examina el uso de técnicas de aprendizaje profundo en la detección de amenazas cibernéticas, destacando sus ventajas, limitaciones y áreas de aplicación.

MITRE y Categorización de ataques comunes

MITRE es una organización sin fines de lucro que opera centros de investigación y desarrollo financiados por el gobierno de los Estados Unidos, conocidos como Centros de Investigación y Desarrollo Financiados Federalmente (FFRDC). Su misión es abordar problemas críticos de interés nacional mediante la aplicación de tecnologías avanzadas y ciencias emergentes. MITRE colabora estrechamente con diversas agencias gubernamentales para ofrecer soluciones innovadoras en áreas clave como ciberseguridad, salud, defensa y seguridad nacional.

Una de sus contribuciones más destacadas en el ámbito de la ciberseguridad es el desarrollo del marco MITRE ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge), una herramienta ampliamente adoptada que proporciona un conocimiento detallado de las tácticas, técnicas y procedimientos (TTPs) utilizados por actores de amenazas. Este marco permite a las organizaciones mejorar su postura de seguridad cibernética al identificar y mitigar efectivamente las amenazas avanzadas.

En este sentido, el comando se enfoca en aquellos ataques que ya han sido categorizados por MITRE ATT&CK.

1. Phishing (T1566)

El phishing es una técnica común en la que los atacantes envían correos electrónicos fraudulentos para engañar a los usuarios y hacerles revelar información confidencial o instalar malware.

2. Command and Control (C2 o C&C)

Las técnicas de comando y control permiten a los atacantes mantener comunicación con sistemas comprometidos y ejecutar comandos. Algunas técnicas comunes de C2 incluyen:

1. **Application Layer Protocol (T1071):** Utilización de protocolos de capa de aplicación como HTTP, HTTPS, DNS para establecer comunicación C2.

1. **DNS (T1071.004):** El uso de DNS para comunicaciones C2 puede incluir técnicas como la exfiltración de datos a través de consultas DNS.

2. **Web Service (T1102):** Los atacantes utilizan servicios web legítimos para C2**.**

3. **Remote Access Software (T1219):** Uso de software de acceso remoto como TeamViewer, VNC, RDP para mantener el control de los sistemas.

En general, existen diversos tipos de ataques que, según MITRE, ponen en riesgo la seguridad informática y los activos de las organizaciones. En este análisis, el enfoque será en aquellos que afectan directamente a los mandos medios y altos, como los administradores, clasificados como ataques de comando y control.

Metodología

La metodología CRISP-DM fue elegida debido a sus ventajas en el desarrollo de proyectos de Ciencia de datos. En primer lugar, CRISP-DM proporciona una estructura clara y organizada en fases secuenciales, lo que facilita la gestión del proyecto y asegura que se aborden todos los aspectos cruciales, desde la comprensión del problema hasta la implementación de posibles soluciones. En segundo lugar, el enfoque iterativo de CRISP-DM permite la flexibilidad de adaptarse a los descubrimientos y desafíos que surjan durante el análisis, lo cual es esencial para el proyecto. Además, CRISP-DM orienta los objetivos específicos, en este caso, la detección de anomalías y la mejora de la ciberseguridad del SOC de la Armada Nacional. Finalmente, CRISP-DM facilita la comunicación con el tutor y otras partes interesadas al proporcionar un marco común para discutir el progreso y los resultados. Su amplia adopción en la industria también permite adquirir experiencia en un proceso estándar de ciencia de datos, lo cual es valioso para el futuro profesional.

Fases del proyecto

1. Entendimiento del negocio:

Esta primera se centra en la definición precisa de los objetivos del trabajo, que buscan fortalecer la ciberseguridad del SOC de la Armada Nacional mediante la detección proactiva de anomalías. Además, entender el entorno de la Armada Nacional, leyes y decretos que rigen la ciberseguridad en el ámbito militar para salvaguardar la seguridad nacional y finalmente entender las necesidades del SOC

2. Comprensión de los Datos:

En esta fase, se recopilan los logs de los dispositivos críticos del SOC de la Armada Nacional. Se lleva a cabo un análisis exploratorio de datos (EDA) para comprender la estructura, características y patrones de los datos, identificando variables relevantes, valores atípicos y posibles relaciones. Se verifica la calidad de los datos, realizando tareas de limpieza y preprocesamiento para asegurar su consistencia e integridad.

3. Preparación de los Datos:

Los datos recopilados se preparan para el modelado mediante la selección de variables relevantes, el tratamiento de valores faltantes, la transformación de variables (como la normalización o la codificación de variables categóricas) y la creación de nuevas variables si es necesario. El objetivo es asegurar que los datos estén en un formato adecuado para los algoritmos de Machine Learning.

4. Modelado:

Se seleccionan y aplican técnicas de Machine Learning apropiadas para la detección de anomalías en los logs, como algoritmos de clasificación, clustering o modelos específicos de detección de anomalías. Se entrenan los modelos con los datos preparados, se ajustan los hiperparámetros para optimizar su rendimiento y se evalúa su eficacia mediante métricas apropiadas como la precisión y el recall.

5. Evaluación:

Se evalúan los resultados obtenidos por los modelos, interpretando los patrones de comportamiento anómalos identificados. Se analiza si el proyecto ha cumplido con los objetivos y requisitos establecidos en la fase de comprensión del negocio. Se identifican las fortalezas y debilidades del enfoque y se formulan conclusiones sobre la efectividad de la solución propuesta.

6. Implementación:

Los resultados del proyecto se comunican a las partes interesadas, incluyendo al personal del SOC de la Armada Nacional, mediante informes, presentaciones o tableros de control. Si es factible, se implementa la solución desarrollada en el entorno del SOC, integrando los modelos de detección de anomalías en los

sistemas de seguridad existentes. Se define un plan de mantenimiento para la solución, que incluye la actualización de los modelos con nuevos datos y la adaptación a nuevas amenazas.

La aplicación de la metodología CRISP-DM permite un desarrollo riguroso y estructurado del proyecto, asegurando que se aborden todas las etapas necesarias para alcanzar los objetivos de forma eficiente. Esto permite demostrar las habilidades en ciencia de datos e inteligencia artificial aplicadas a la ciberseguridad y contribuir al fortalecimiento de la protección de los activos críticos de la Armada Nacional.

Capítulo 1

Análisis exploratorio de datos de los activos críticos de la Armada Nacional

El análisis de datos exploratorio o simplemente EDA es un enfoque analítico que se utiliza para comprender y analizar los datos de manera inicial y profunda. Es un enfoque flexible, que además debe ser iterativo para dialogar con los datos, en todo el proceso del ciclo de vida del proyecto.

A continuación, se presentan cada una de las fases del EDA aplicadas a los datos entregados para llevar a cabo el proyecto.

Recolección de Datos

La muestra entregada por la Armada proviene de distintos sensores ubicados de manera estratégica y que son registrados y almacenados por el SIEM o plataformas de monitoreo de seguridad; esta muestra consiste en dos grupos de archivos de distintas fechas, con los respectivos logs de cada una de ellas.

Se tiene un total de 4 muestras discriminadas por fechas del año 2024:

- Muestra 1: 24 al 30 abril
- Muestra 2: 7 al 13 marzo
- Muestra 3: 29 de mayo
- Muestra 4: 15 al 24 Julio

Las cuales contienen las siguientes tablas:

File Name	Size
Active Directory	1896 KB
Asset Authentication	3128173 KB
Audit Logs	322 KB
DNS Query	773055 KB
Endpoint Activity	8013366 KB
File Access Activity	2709269 KB
File Modification Activity	336999 KB
Firewall Activity	2716502 KB
Host To IP Observations	1310100 KB
IDS Alert	325997 KB
Ingress Authentication	4019 KB
Internal Logs	7 KB
Network Flow	1249659 KB
Raw Logs	1570448 KB
Third Party Alert	54 KB
Unparsed Data	76010 KB
Virus Alert	102 KB
Web Proxy Activity	3126848 KB

Preprocesamiento de Datos

Una de las primeras etapas del preprocesamiento se llevó a cabo cuando se eligió, de todas las tablas entregadas por el comando, aquellas que aportaran de manera significativa directa o indirectamente al ataque elegido C2.

DataFrames para C2

En este sentido, con el fin de llevar a cabo un análisis exhaustivo de los ataques que se ajusten al perfil de Comando y Control, se dará prioridad al estudio de las siguientes tablas y sus relaciones y correlaciones:

Active Directory:

Es una base de datos y un conjunto de servicios que conectan a los usuarios con los recursos de red de la Armada que necesitan para realizar su trabajo. La base de datos (o directorio activo) contiene información crítica sobre su entorno,

incluidos los usuarios y las computadoras que hay, además lo que el usuario identificado puede o no hacer.

Asset Authenticator:

El Asset Authenticathor es una herramienta que se utiliza para realizar autenticación y autorización de activos (dispositivos, sistemas, aplicaciones, etc.) en un entorno empresarial. Algunas de las principales funcionalidades y características del Asset Authenticathor son:

- Descubrimiento de activos: Permite identificar y catalogar todos los activos conectados a la red, incluyendo dispositivos, servidores, aplicaciones, etc.
- Autenticación de activos: Verifica la identidad de los activos a través de diversos métodos de autenticación, como credenciales, certificados digitales, etc.
- Gestión de accesos: Controla y administra los permisos y privilegios de acceso a los diferentes activos, asegurando que solo los usuarios y procesos autorizados puedan interactuar con ellos.
- Monitoreo y alertas: Realiza un seguimiento continuo de la actividad de los activos, generando alertas en caso de detección de comportamientos sospechosos o actividades no autorizadas.

DNS Query:

El DNS Query es el conjunto de solicitudes de información enviada a un servidor DNS (Sistema de Nombres de Dominio) para traducir un nombre de dominio en una dirección IP. El SIEM puede analizar estas consultas para detectar actividad sospechosa, como ataques cibernéticos o comunicación con servidores comprometidos

IDS Alert:

El archivo IDS Alert o sistema de detección de instrucciones es un componente relacionado con la detección y respuesta a intrusiones en la infraestructura de TI de una organización, es una herramienta de ciberseguridad que monitorea el tráfico de red o actividades del sistema en busca de comportamientos sospechosos o maliciosos.

Algunas de las principales funcionalidades y características del IDS Alert son:

- Registrar Actividades Sospechosas: Almacena registros detallados de eventos o actividades que el sistema de detección considera potencialmente maliciosas o anómalas. Estos eventos pueden incluir

intentos de acceso no autorizados, escaneos de red, intentos de explotación de vulnerabilidades, entre otros.

- Análisis de Incidentes: Los datos almacenados en el archivo IDS Alert se utilizan para analizar incidentes de seguridad. Esto ayuda a los equipos de seguridad a investigar eventos sospechosos, determinar la naturaleza de las amenazas y tomar acciones correctivas.
- Generación de Alertas: Basándose en los eventos registrados, el sistema puede generar alertas en tiempo real para notificar a los administradores de seguridad sobre posibles incidentes. Estas alertas son fundamentales para una respuesta rápida y eficaz.
- Cumplimiento y Auditoría: Mantener un registro de todas las alertas de IDS es también importante para cumplir con diversas regulaciones de seguridad y auditorías. Proporciona una pista de auditoría que puede ser revisada en caso de un incidente de seguridad significativo.

File Access Activity:

El File Access Activity (Actividad de Acceso a Archivos) se refiere al monitoreo y registro de acciones relacionadas con archivos dentro de un entorno de red o sistema informático. Esta actividad incluye cualquier tipo de acceso que implique acceder, leer, escribir, modificar, eliminar o mover a los diferentes tipos de archivo en una red o en sistemas locales, registrando datos como el usuario, dominio e IP desde el cual se realizó el acceso junto con los horarios en que se realizaron dichas actividades. Este log es crucial para la seguridad de la información y la detección de amenazas, ya que puede ayudar a identificar actividades sospechosas (atípicas) o maliciosas, como intentos de acceso no autorizado, cambios no autorizados en archivos críticos, intentos de exfiltración de datos o actividades en horarios irregulares.

File Modification Activity:

El File Modification Activity (Actividad de Modificación de Archivos) se refiere al monitoreo y registro de cambios realizados en archivos dentro de un entorno de red o sistema informático. Esta actividad incluye cualquier acción que implique la modificación, actualización, creación o eliminación de archivos en una red o en sistemas locales. La monitorización de la actividad de modificación de archivos es crucial para la seguridad de la información y la detección de amenazas, ya que puede ayudar a identificar actividades sospechosas o maliciosas, como cambios no autorizados en archivos críticos, manipulación de archivos por parte de usuarios no autorizados, o intentos de alterar o borrar datos importantes.

Firewall Activity:

El Firewall Activity (Actividad de Firewall) se refiere al monitoreo y registro de eventos y acciones relacionadas con el firewall de red. Esto incluye cualquier actividad que implique el tráfico de red que pasa a través del firewall, como conexiones entrantes y salientes, reglas de firewall aplicadas, intentos de conexión bloqueados o permitidos, y otros eventos relacionados con la seguridad de la red. El firewall es una parte fundamental de la infraestructura de seguridad de una red, ya que actúa como una barrera entre la red interna y externa, controlando y filtrando el tráfico de red según las reglas predefinidas. Monitorear la actividad del firewall es esencial para detectar y prevenir posibles amenazas de seguridad, como intentos de intrusión, tráfico malicioso, o comportamientos anómalos que podrían indicar un ataque.

Limpieza de los Datos

El proceso de limpieza de datos implica identificar y corregir errores, valores faltantes, duplicados y otros problemas que puedan eventualmente afectar la calidad de los resultados.

En este sentido las tablas seleccionadas para C2 se sometieron al Datapipeline. Esta limpieza se llevó a cabo para todas las tablas e incluye aplicar dos criterios generales:

Criterio 1: En todas las tablas, se eliminarán aquellas columnas cuyos registros presenta un 70% o más de valores faltantes o nulos. El método .drop() aplicado en el pipeline permitió automatizar el proceso

Criterio 2: En algunas columnas de las tablas se presentan un alto registros duplicados lo que genera desbalanceo en las decisiones, lo cual puede afectar el algoritmo, en algunos casos incluso es del 100%, lo cual afecta el análisis de datos. En este sentido aquellos cuyos registros tengan el 70% de datos iguales se eliminan, tomando atenta y cuidadosa nota de su significado en las fases finales.

Una vez se lleva a cabo la limpieza, procedemos a hacer las transformaciones en atributos y registros.

Transformaciones y selección de atributos

Duplicidad de columnas: Se hizo evidente, en la mayoría de las tablas, que algunas de sus columnas se comportaban de manera similar, teniendo una especie de duplicidad entre ellas, si bien en algunos casos no era del 100% de registros, se elige el criterio que aquellas columnas que tengan el 80% o más de datos iguales, se eliminen dejando una, que actuará como representante de las demás. Este

proceso busca además de evadir la correlación, reducir la dimensionalidad de las tablas. Para esto se tuvo en cuenta el estadístico Xhi cuadrado.

Datatime: En todas las tablas se hace la transformación del formato timestamp a datatime. Esto se debe a que todas las fechas de las tablas en el cual se hizo el registro del evento o logs están en formato ISO 8601. Estos datos tienen la Zona horaria zulú: La Z al final indica que las marcas de tiempo están en Tiempo Universal Coordinado (UTC). Para convertirla a la hora local de Colombia (GMT-5), se debe restar 5 horas y hacer la transformación a un formato de YYYY-MM-DD HH: MM.

Manejo de Datos Faltantes: Para el manejo de datos faltantes, se aplicarán de manera simultánea dos de los métodos más usados

- **Listwise deletion:** En algunas tablas se optará por eliminar las filas, esto se debe a que, por la naturaleza de los datos, se hace difícil anticipar el dato faltante, por lo que no se puede aplicar ningún tipo de imputación
- **Imputación**: usando métodos basados en estadísticas y modelado, se buscará imputar los datos en algunas de las tablas cuya naturaleza permita aplicar el método.

Una vez los Dataframe se someten a esta limpieza y transformación general, se lleva a cabo un proceso adicional de selección de variables (Feauture Selection) a cada una de las tablas, entendiendo que cada tabla tiene sus características propias, este proceso se hizo combinando estadísticos de prueba, como xhi cuadrado con niveles de confiablidad del 0.95 y un nivel de significancia de 0.05 además de la visión de expertos en el negocio. Esto se hace con el objetivo de buscar reducir dimensionalidad (PCA)

Capítulo 2

Detección de anomalías

Anomalía

En el contexto de la ciberseguridad, El instituto nacional de estándares y tecnología NIST el cual ofrece marcos utilizados para gestionar riesgos, amenazas y vulnerabilidad define una anomalía como cualquier evento, actividad o patrón que **se desvíe significativamente de lo que se considera normal** o esperado dentro de un sistema informático Para ello es importante, a partir de los datos obtenidos del comando, identificar cómo y cuáles son sus comportamientos considerados normales. Con ello se busca identificar de manera temprana un posible vector de riesgo, muchas ciberamenazas, como ataques, malware o intrusiones, se manifiestan inicialmente como anomalías.

Identificar las anomalías a tiempo permite responder de manera proactiva, rápida y efectiva. Además, se usa para la prevención de pérdidas; Las anomalías pueden indicar brechas de seguridad que, si no se abordan, podrían llevar a pérdidas de información, robo de datos o daños a la reputación. Finalmente se busca una mejora de la seguridad, al analizar las anomalías, la Armada nacional puede identificar vulnerabilidades en sus sistemas y tomar medidas para fortalecer su seguridad.

Algunos ejemplos de anomalías en ciberseguridad son:

- Tráfico de red inusual: Un aumento repentino en el tráfico de red a horas no habituales o entre y hacia direcciones IP desconocidas puede indicar un ataque en curso.
- Accesos no autorizados: Intentos de acceso a sistemas o datos por parte de usuarios no autorizados o desde ubicaciones geográficas inesperadas.
- Cambios inesperados en los archivos: Modificaciones no autorizadas en archivos críticos del sistema o datos confidenciales.

- Consumo anormal de recursos: Un aumento repentino en el uso de la CPU, memoria o ancho de banda puede ser señal de un malware o un ataque de denegación de servicio.
- Comportamiento inusual de los usuarios: Actividades fuera de lo común de los usuarios, como descargas de archivos sospechosos o visitas a sitios web maliciosos.

Técnicas para detectar anomalías

Algunas de las técnicas más utilizadas para verificar posibles comportamientos anormales en los datos son:

Sistemas de detección de intrusiones (IDS): son técnicas y procesos manuales o automatizados que monitorean la red y el tráfico entre ellas en busca de patrones de actividad inusual

Análisis de registros: son técnicas que examinan los registros del sistema para identificar eventos inusuales. En este caso cruzamos información de los dos aplicativos usados por el SOC.

Aprendizaje automático: consiste en usar algoritmos de estadística descriptiva, analítica y prescriptiva, además de algoritmos de aprendizaje automatizado para identificar patrones en grandes conjuntos de datos y detectar desviaciones de la norma.

Marco de Mitre para la detección de anomalías

MITRE proporciona varias herramientas y marcos que pueden ser extremadamente útiles para la **detección de anomalías**, particularmente en el contexto de la **ciberseguridad**. Si bien MITRE no proporciona una solución directa como un software específico de detección de anomalías, ofrece marcos de conocimiento y metodologías que son fundamentales para identificar y mitigar comportamientos anómalos en una red.

Los elementos más destacados que MITRE ofrece son los siguientes:

MITRE ATT&CK Framework

El **MITRE ATT&CK** (Adversarial Tactics, Techniques & Common Knowledge) es un marco muy conocido que ayuda a las organizaciones a entender cómo operan los atacantes, cuáles son sus tácticas y técnicas, y qué métodos usan para comprometer la seguridad. En el contexto de la **detección de anomalías**, ATT&CK ofrece lo siguiente:

Tácticas y Técnicas para la Detección de Amenazas:

ATT&CK proporciona una lista exhaustiva de tácticas y técnicas usadas por los atacantes, que pueden ayudarte a entender qué tipos de anomalías buscar. Por ejemplo, si tienes registros de autenticación y de actividad en dispositivos, puedes buscar patrones que correspondan a las técnicas documentadas en ATT&CK.

Mapeo de Anomalías a Técnicas:

Al analizar anomalías en los registros (como intentos fallidos de autenticación desde un solo dispositivo), se puede mapear estas anomalías a técnicas documentadas en ATT&CK. Esto permite clasificar los eventos detectados y priorizar las respuestas en función del riesgo asociado con cada técnica.

MITRE CAR (Cyber Analytics Repository)

El **Cyber Analytics Repository (CAR)** es otra herramienta proporcionada por MITRE que puede ser útil para la detección de anomalías:

Casos de Uso Analíticos:

CAR contiene una colección de análisis basados en técnicas de ATT&CK y ejemplos de cómo implementar estos análisis en tus datos de registros.

Puedes utilizar estos análisis como referencia para implementar tus propias detecciones. Por ejemplo, MITRE CAR puede ofrecer un análisis para detectar intentos repetidos de autenticación fallida o el uso de herramientas de ataque específicas.

Especificaciones de Análisis:

Cada análisis dentro de CAR tiene especificaciones claras de cómo implementar la detección, qué datos utilizar (como registros de eventos de Windows, autenticación, etc.), y qué lógica seguir para identificar patrones anómalos.

Relaciones entre las tablas y caracterización de Anomalías

Teniendo en cuenta las tablas elegidas, se llevará a cabo los siguientes merged join para la caracterización de algunas anomalías relevantes en el proceso

1. Active Directory + Asset Authentication:

Relación: Unir ambas tablas por el nombre de usuario o ID de empleado.

Posibles Anomalías:

1.1 Intentos de autenticación fallidos repetidos desde un mismo dispositivo (Asset Authentication) con diferentes cuentas (Active Directory). Esto podría indicar un ataque de fuerza bruta.
1.2 Autenticaciones en horarios o ubicaciones geográficas inusuales para un usuario.
1.3 Usuarios con privilegios excesivos o acceso a recursos sensibles sin justificación.

2. Active Directory + DNS Query:

Relación: Se tiene información del equipo en las consultas DNS, se debe relacionar con Active Directory a través del nombre de usuario o ID de la máquina.

Posibles Anomalías:

2.1 Consultas a dominios sospechosos o conocidos por ser maliciosos por parte de usuarios o equipos específicos.
2.2 Patrones inusuales de consultas DNS, como un aumento repentino en el volumen de consultas o consultas a dominios nuevos y desconocidos. Esto podría indicar actividad de malware o command-and-control.

3. Active Directory + File Access/Modification Activity:

Relación: Unir por el nombre de usuario o ID de empleado.

Posibles Anomalías:

3.1 Acceso o modificación de archivos sensibles por parte de usuarios no autorizados.

4. Asset Authentication + Firewall:

Relación: Unir por la dirección IP del dispositivo.

Posibles Anomalías:

4.1 Intentos de conexión desde dispositivos (Asset Authentication) a direcciones IP externas bloqueadas por el firewall.

5. Detección de anomalías por tráfico inusual de red analizando el Firewall

Relación: Análisis de Firewall

Posibles Anomalía

5.1 El tráfico no habitual se caracteriza por desviarse de los patrones normales de tráfico en una red normal. Esto puede incluir un aumento significativo en el volumen de tráfico, conexiones a hosts desconocidos o el uso de protocolos inusuales

En el ejercicio de caracterizar posibles anomalías, la información siempre se contrasta con el experto del SOC que nos va direccionando sobre la pertinencia e interpretabilidad de los resultados obtenidos por el despliegue de los scripts en Python.

Modelos

La construcción de los modelos en Python para la detección de anomalías se llevó a cabo mediante un enfoque estructurado y por etapas, combinando el análisis de datos, el diseño de algoritmos y la colaboración interdisciplinaria.

El proceso comenzó con el diseño de los scripts personalizados en Python utilizando bibliotecas de procesamiento de datos como Pandas y NumPy para la integración y transformación de los registros en un formato estandarizado. Este paso incluyó técnicas como la normalización, codificación de datos categóricos y creación de nuevas variables que capturan características críticas, como patrones de tráfico inusual, intentos fallidos de autenticación o accesos no autorizados a archivos sensibles.

En la etapa de modelado, se utilizaron algoritmos de aprendizaje automático implementados con bibliotecas como Scikit-learn, TensorFlow y PyDeequ. Los modelos incluyeron técnicas no supervisadas, como clustering, detección de outliers y algoritmos específicos para la caracterización de anomalías, adaptados a las particularidades de los registros analizados. Por ejemplo, se aplicaron modelos de agrupamiento para detectar patrones de comportamiento fuera de lo común y técnicas basadas en árboles de decisión para identificar características clave de las anomalías detectadas.

El desarrollo de estos modelos fue enriquecido con el uso de repositorios especializados que proporcionaron ejemplos y mejores prácticas, lo que facilitó la implementación eficiente de técnicas avanzadas. Además, en cada etapa del proceso, se contó con mi asesoramiento técnico para garantizar que las herramientas y enfoques utilizados fueran los más adecuados para abordar los desafíos específicos de la tarea.

Una vez construidos, los modelos fueron validados utilizando subconjuntos de los datos históricos, contrastando los resultados obtenidos con el conocimiento experto del SOC. Este paso permitió ajustar los parámetros de los modelos y mejorar su

capacidad predictiva y de detección. Además, se diseñaron pruebas específicas para asegurar que las anomalías identificadas correspondieran a eventos relevantes desde el punto de vista operativo, evitando falsos positivos que podrían desviar la atención de las amenazas reales.

Finalmente, los resultados se documentaron de manera detallada, destacando los patrones anómalos detectados, las métricas de evaluación del desempeño de los modelos y las recomendaciones para su implementación en entornos reales.

Este proceso no solo fortaleció la capacidad del SOC para anticipar y mitigar riesgos cibernéticos, sino que también estableció una metodología replicable que puede ser escalada a otras áreas críticas de la Armada Nacional. La colaboración con los expertos del SOC fue crucial en todas las etapas, garantizando que los modelos fueran interpretables, operativos y alineados con las necesidades estratégicas de la organización

Capítulo 3

Análisis de rendimientos

En este capítulo se presenta un análisis detallado de las anomalías detectadas, explicando cada una de ellas a partir de los resultados obtenidos por los modelos implementados. Se incluirá el código utilizado y una descripción de los parámetros ajustados, con el fin de proporcionar una visión clara del proceso de detección y su impacto en los datos. Además, se discutirán los hallazgos más relevantes y su interpretación en el contexto de la ciberseguridad operativa de la Armada Nacional.

Detección de intentos fallidos a través del Isolation Forest

Una de las primeras anomalías caracterizadas que se presenta en los merged join de nuestros primeros dos data set es el de intentos fallidos, en este sentido usando el primer modelo obtuvimos que no se presentaba patrones anormales de comportamientos de user/asset pero cuando implementamos un Isolation forest ajustando parámetros, se obtuvo el siguiente resultado

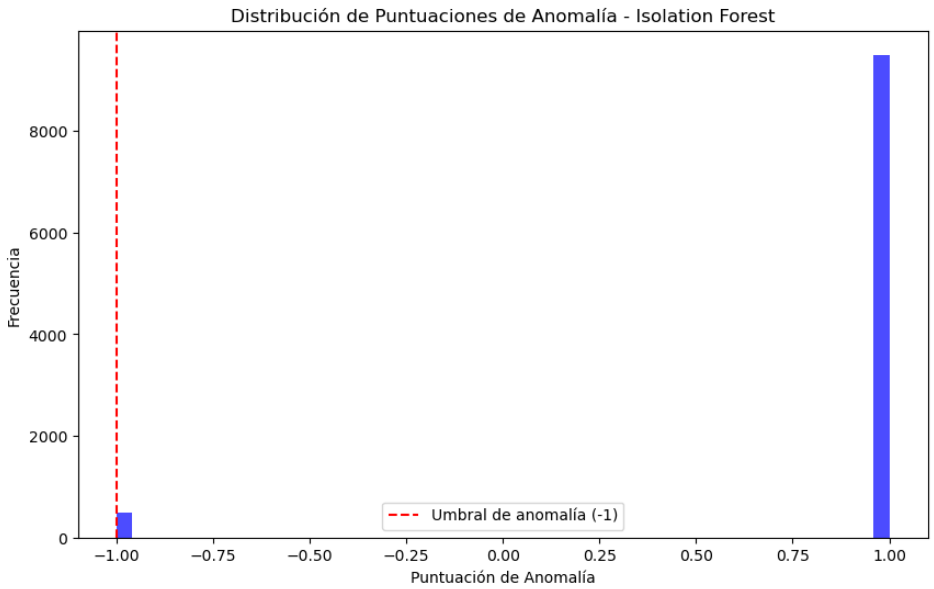


Ilustración 1: Anomalías usando Isolation Forest

Cuando se ajustaron nuevamente los parámetros en el árbol como los pasos (steps) y la profundidad, el cambio no fue significativo en los resultados; por lo tanto podemos concluir:

Distribución de puntuaciones de anomalía:

El gráfico muestra que las puntuaciones generadas por el modelo están mayormente concentradas en -1, lo que indica que el modelo ha detectado un número significativo de anomalías.

Anomalías detectadas:

Se detectaron **499 anomalías** en el subconjunto de datos procesado.

Estas anomalías se identificaron basándose en la estructura del modelo Isolation Forest, que evalúa desviaciones significativas en los datos y que se ha considerado normal.

En este proceso se intenta aplicar etiquetas Label encoder y se intenta utilizar un Random Forest tradicional para clasificar, pero los resultados no son positivos.

---

*Esto ocurre porque el modelo **Random Forest Classifier** no puede calcular probabilidades para múltiples clases si y (las etiquetas) tiene una sola clase (en este caso, todos los valores de y son 0).*

Resumen de anomalías:

Las columnas relevantes del conjunto de anomalías incluyen:

- **Usuarios (json.destination_user):** Usuarios con comportamiento fuera de lo esperado.
- **Dispositivos (json.source_json.computerName):** Dispositivos que pueden haber registrado actividad anómala.
- **Puntos relevantes como "IpPort" y "EventCode".** Los cuales se contrastan con los primeros resultados obteniendo patrones en las Ip que se analizan después con el firewall
- The 'json.result' column was not found in the merged DataFrame. Unable to filter failed authentication attempts.

  Frecuencia de Códigos de Evento Repetidos:

index	EventCode	Count
0	4624.0	2769654
1	4769.0	26454
2	4768.0	3869
3	4625.0	57

Interpretación:

- La mayoría de los registros clasificados como anomalías presentan características que se desvían significativamente del comportamiento normal en el conjunto de datos, de usuarios o activos, tráfico o conexiones, todos desde usuarios autorizados y cuyo resultado fue “Succes”
- Estos puntos requieren análisis adicional para determinar si las anomalías detectadas son verdaderas amenazas o falsos positivos, sin embargo, con las dos primeras tablas no es suficiente para concluir el RGB de la anomalía.
- Las actividades involucran una variedad de usuarios, dominios y dispositivos, pero muestran ciertas concentraciones en valores específicos (e.g., puertos altos y eventos con códigos en el rango 4624–4769).
- Los resultados uniformes en json.result y anomaly_score sugieren que estos eventos tienen características muy diferentes del comportamiento normal en el dataset completo.

Análisis de horarios habituales

Para el análisis de anomalías de usuarios y activos fuera del horario considerado normal, se establece que los horarios habituales del personal de la Armada son de 7:00 a 18:00. Sin embargo, este rango no es una regla estricta, ya que puede variar para algunos usuarios debido a pruebas, revisiones o actividades específicas. Por ello, es esencial que cualquier patrón anómalo identificado por el agente sea verificado y validado por el personal del SOC

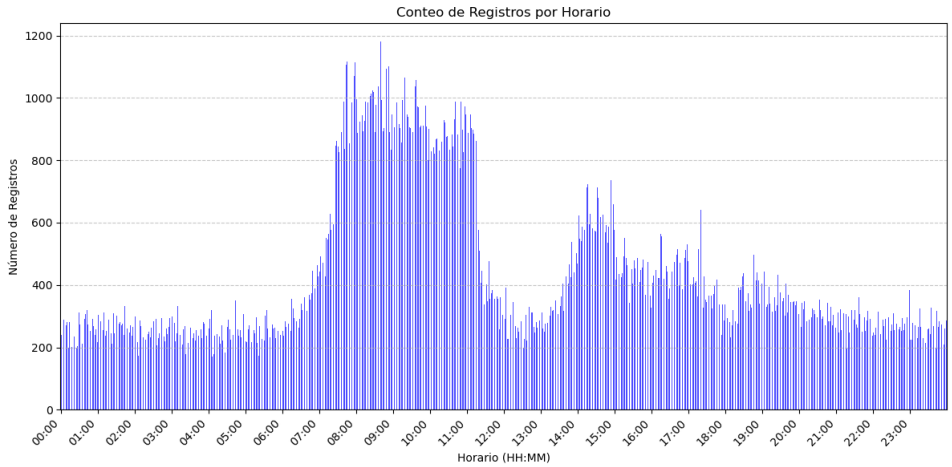


Ilustración 2: Users en diferentes horarios

Utilizando como baseline el horario habitual, analicemos los logs generados por el asset y el Authenticathor

La tabla muestra los **primerousuarios (json.destination_user)** más frecuentes con actividad fuera del horario considerado normal (07:30 a 19:00). Los valores representan el número de eventos registrados por usuario durante esos horarios no habituales.

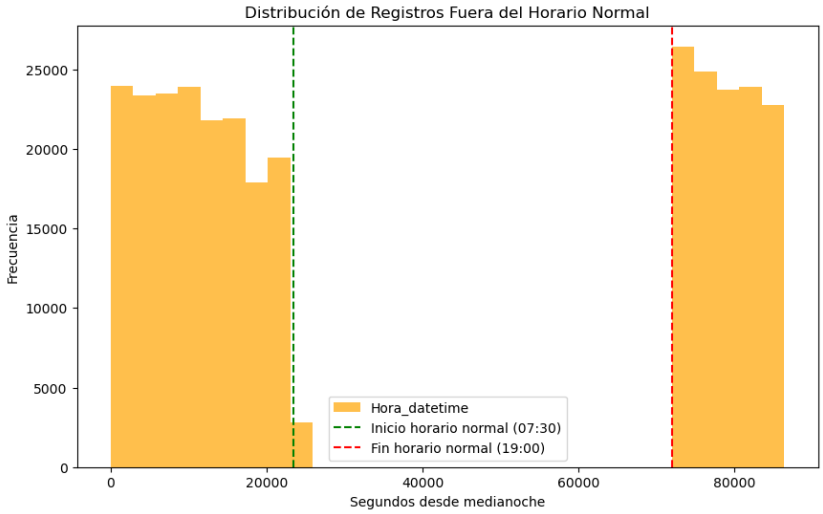
json.destination_user	count
User_1	29418
User_2	25422
User_3	22538
User_4	22168
User_5	20940
User_6	3013
User_7	1791
User_8	1696
User_9	1168

User_10	1124
User_11	1099
User_12	1021
User_13	1021
User_14	972
User_15	958

Observaciones:

1. **Usuarios con mayor actividad fuera de horario:**
    - **User_1** lidera la lista con **29,418 registros**, seguido por **User_2** con **25,422 registros** y **User_3** con **22,538 registros**.
    - Esto sugiere que estos usuarios tienen actividades recurrentes fuera del horario normal.
2. **Patrones de actividad específica:**
    - Algunos usuarios, podrían estar asociados a funciones operativas o tareas de mantenimiento que ocurren fuera del horario laboral estándar.
    - Otros usuarios podrían realizar actividades vinculadas a roles de guardia o turnos nocturnos.
3. **Usuarios con menor actividad fuera de horario:**
    - Los últimos en la lista muestran menos eventos, pero siguen estando fuera del rango normal.
4. **Posibles interpretaciones:**
    - **Actividades legítimas:** Algunos usuarios pueden tener horarios específicos debido a la naturaleza de sus tareas (e.g., mantenimiento, soporte, guardias).
    - **Potenciales anomalías:** Un análisis adicional es necesario para determinar si estas actividades están justificadas o representan comportamientos atípicos que requieren atención.

Relación con la gráfica:



La gráfica de distribución muestra un número significativo de registros fuera del horario normal (antes de las 07:30 y después de las 19:00), respaldando los datos de la tabla. La validación con el equipo SOC sería clave para interpretar correctamente los patrones de actividad.

Actividades fuera de horario pueden ser esperadas para ciertos roles (e.g., soporte, mantenimiento, guardias) o circunstancias excepcionales (e.g., pruebas, actualizaciones del sistema). Sin embargo, la validación con el SOC es clave para confirmar la legitimidad de estas actividades.

Hay que considerar que:

- **Riesgo de comportamiento anómalo:**
  - Usuarios con actividades regulares fuera del horario pueden ser más vulnerables a ser blanco de amenazas de seguridad, ya que sus comportamientos pueden no ser monitoreados con el mismo rigor.
  - Algunos registros también pueden representar anomalías o accesos inusuales que necesitan revisión detallada.
- **Validación requerida:**
  - Es crucial que los registros sean revisados por el personal del SOC para identificar si las actividades están justificadas o si corresponden a patrones anómalos que podrían indicar riesgos de seguridad.

Recomendaciones:

1. **Clasificación de usuarios y roles:**

- Identificar roles específicos asociados con estos usuarios para determinar si sus actividades fuera de horario son esperadas.

2. **Validación con el SOC:**
   - Revisar los usuarios con actividad más frecuente fuera del horario y confirmar con el equipo de seguridad si estas actividades están justificadas.
3. **Refuerzo de monitoreo:**
   - Implementar reglas de alerta para monitorear actividades fuera del horario y evaluar su legitimidad en tiempo real.
4. **Análisis de registros históricos:**
   - Comparar patrones actuales con datos históricos para determinar si estos comportamientos fuera de horario son nuevos o recurrentes.

Dominios sospechosos

Siguiendo con las anomalías, otra que suele ser un caso común es la identificación de dominios sospechosos, para esto hemos hecho un merged con el DNS y el active directory

Para construir una lista de dominios sospechosos, podemos buscar en varias fuentes de información conocidas por rastrear dominios maliciosos o que tienen un historial de uso para actividades sospechosas.

Aquí se presentan algunas de ellas:

1. Fuentes de Amenazas Públicas (Threat Intelligence)

Existen varias organizaciones y proyectos que publican listas de dominios sospechosos y maliciosos:

* PhishTank: Proporciona una lista de dominios de phishing y es mantenida por OpenDNS (https://phishtank.com/).

* VirusTotal: Permite analizar URLs y archivos en busca de amenazas. Puedes enviar solicitudes a la API para obtener información de dominios (https://www.virustotal.com/).

* Spamhaus: Publica listas de dominios sospechosos y dominios asociados con actividades maliciosas como spam (https://www.spamhaus.org/).

* URLhaus: Mantenida por abuse.ch, proporciona listas de dominios y URLs asociadas con malware (https://urlhaus.abuse.ch/).

2. Extensiones de Dominio Comúnmente usadas

Algunas extensiones de dominio suelen estar asociadas a actividades de baja reputación debido a que son públicas de fácil acceso o permiten registros anónimos.

Entre estas están las extensiones de alto riesgo: .ru, .cn, .xyz, .top, .biz, .tk, .pw, .cc, .work, .support

Una vez definido los dominios de alto riesgo, se obtiene

Index	json.destination_user	json.query	json.query_type	Fecha	Hora
1	gxx. Sxx	safeurl.maxthon.cn	A	6/04/2024	20:24:21
2	gxx. Sxx	safeurl.maxthon.cn	A	6/04/2024	20:27:21
3	gxx. Sxx	safeurl.maxthon.cn	A	6/04/2024	20:30:21
4	gxx. Sxx	safeurl.maxthon.cn	A	6/04/2024	20:33:21
5	gxx. Sxx	safeurl.maxthon.cn	A	6/04/2024	20:36:21

Los registros muestran consultas repetidas al dominio safeurl.maxthon.cn realizadas por el usuario gxx.Sxx desde la dirección IP interna. Estas actividades ocurrieron fuera del horario normal (entre las 20:24 y las 20:36). A continuación, se detalla el análisis:

Contexto Relevante:

Usuario Crítico

- El SOC ha indicado que este usuario pertenece a un sistema crítico. Por lo tanto, cualquier actividad anómala asociada debe investigarse con prioridad.

Uso de herramientas no permitidas (safeurl.maxthon.cn):

- El dominio pertenece a un servicio relacionado con navegadores o herramientas de VPN/proxy.
- El SOC ha declarado explícitamente que estas herramientas están prohibidas debido al riesgo de exfiltración de datos o compromisos de seguridad.

Frecuencia de consultas DNS:

- Las consultas al dominio son recurrentes con intervalos de 3 minutos. Este comportamiento puede indicar una configuración automática o un posible uso malintencionado.

Riesgos Asociados:

- **Exfiltración de datos**: Las VPN/proxys pueden usarse para desviar tráfico hacia servidores externos, lo que podría representar un riesgo de pérdida de datos.
- **Compromiso de la red**: Si el dominio es parte de un servicio malicioso o comprometido, podría facilitar la entrada de amenazas externas a la red crítica.

Recomendaciones

1. **Validación de Actividad**:
   - Verificar con el usuario si las consultas fueron realizadas de forma legítima y aceptada por el soc
   - Confirmar si la dirección IP corresponde a un equipo autorizado para esta función.
2. **Bloqueo del Dominio**:
   - Implementar reglas en el firewall y las políticas DNS para bloquear safeurl.maxthon.cn y otros dominios relacionados con servicios de VPN o proxy no autorizados.
3. **Auditoría de Seguridad**:
   - Revisar los registros históricos del usuario gxx.Sxx y la IP involucrada para detectar patrones similares o accesos no autorizados.
   - Analizar si esta actividad pudiera estar relacionada con un posible compromiso del dispositivo.
4. **Refuerzo de Políticas**:
   - Reiterar a todos los usuarios las políticas sobre el uso prohibido de VPN/proxys.
   - Asegurarse de que los sistemas críticos no tengan configuraciones que permitan el uso de herramientas no autorizadas.
5. **Monitoreo Continuo**:
   - Configurar alertas específicas para detectar consultas futuras a dominios de alto riesgo desde sistemas críticos.
   - Implementar herramientas de análisis avanzado para identificar automáticamente actividades similares.

Modificación o eliminación de archivos y el impacto de las redes neuronales

Un ejemplo común de anomalías relacionadas con comportamientos de comando y control es la modificación o eliminación de archivos por parte de usuarios, tanto internos como externos a las organizaciones. En particular, se consideran sospechosas aquellas actividades realizadas por usuarios que, aunque estén autorizados en el Directorio Activo, sean identificadas como inusuales o potencialmente maliciosas dentro del SOC (Centro de Operaciones de Seguridad).

El código desarrollado tuvo como objetivo identificar accesos o modificaciones sospechosas a archivos sensibles por parte de usuarios, combinando información proveniente de diferentes tablas de registros. Inicialmente, se analizaron los datos del archivo de accesos (file_acces) para detectar usuarios con alta frecuencia de accesos mediante la columna json.account. Posteriormente, se inspeccionan las modificaciones de archivos a través del archivo de modificaciones (file_modification), utilizando la columna json.r7_context.asset.name para identificar los dispositivos o activos involucrados. Finalmente, los datos se combinan con la tabla de usuarios activos (active), vinculando los accesos y modificaciones con los usuarios responsables, empleando las columnas json.account y json.destination_user.

El resultado final incluye información detallada de los accesos y modificaciones, tales como el nombre del archivo, el usuario involucrado, la marca de tiempo y el dispositivo relacionado. Además, se generan estadísticas sobre los archivos más accedidos o modificados, y se guarda un resumen de los registros detectados en un archivo CSV para su posterior análisis. Este análisis permite identificar patrones inusuales de comportamiento que podrían representar riesgos de seguridad en el entorno monitoreado.

En esta tabla se puede observar la actividad de aquellos usuarios internos con la alta frecuencia de actividad en el file activity

json.account	count
account_xx1	127347
account_xx2	106184
account_xx3	100456
account_xx4	80600

account_xx5	73415
account_xx6	67599
account_xx7	62703
account_xx8	49165
account_xx9	45089
account_xx10	43894

Además de caracterizar los equipos con actividad y modificaciones frecuentes

json.r7_context.asset.name	count
asset_xx1	32407
asset_xx2	18004
asset_xx3	15366
asset_xx4	11571
asset_xx5	7250
asset_xx6	7133
asset_xx7	5439
asset_xx8	4920
asset_xx9	4902
asset_xx10	4886

El análisis mediante redes neuronales con un modelo Autoencoder permitió identificar registros anómalos dentro de los datos combinados de accesos y modificaciones de archivos. El modelo fue entrenado para aprender los patrones normales de interacción entre usuarios, direcciones de origen y extensiones de archivos. Posteriormente, se calcularon los errores de reconstrucción para cada registro, considerando como anomalías aquellos cuyo error superó el percentil 95. Estas anomalías representan accesos o modificaciones que se desvían significativamente de los patrones esperados, lo que podría indicar comportamientos inusuales, configuraciones erróneas o posibles amenazas de seguridad. Los resultados se consultarán en un trabajo futuro con los expertos en el dominio.

Análisis del Firewall

Una manera efectiva de identificar anomalías a través del firewall es mediante el monitoreo del tráfico de red. Cuando se detecta un aumento significativo en el tráfico asociado a las direcciones IP internas, esto podría ser indicativo de actividades sospechosas, como las siguientes:

- Ataques de Denegación de Servicio (DoS): Un volumen anormalmente alto de bytes entrantes o salientes puede ser un signo de un ataque DoS, en el que un atacante busca saturar un sistema mediante un flujo masivo de tráfico, dificultando o impidiendo su funcionamiento normal.
- Escaneo de Puertos: Un número inusual de conexiones dirigidas a diferentes puertos de destino, especialmente a puertos que no suelen estar en uso, podría indicar que alguien está escaneando el sistema en busca de posibles vulnerabilidades para explotar.

Estos comportamientos, si son identificados oportunamente, pueden ayudar a mitigar riesgos y proteger los activos críticos de la organización.

Trafico de Red y método Z

El conjunto de datos proporcionado contiene información sobre el tráfico de la red, procedente del registro de firewall.

Para identificar anomalías en este conjunto de datos, nos centraremos en las columnas numéricas que son relevantes para el análisis del tráfico de red: json.incoming_bytes, json.outgoing_bytes, json.destination_port y json.source_port.

A continuación, se continua con las puntuaciones Z para cada una de estas columnas para identificar cualquier valor inusual. Las puntuaciones Z representan cuántas desviaciones estándar tiene un punto de datos de la media. Una puntuación Z absoluta alta indica un valor atípico o una anomalía.

El filtrado basado en puntajes Z se realiza utilizando la línea

filtered_df = df[(df['z_json.incoming_bytes'].abs() > 3) | ...], que crea un nuevo DataFrame llamado filtered_df aplicando un filtro al DataFrame original df. Este filtro selecciona filas donde el valor absoluto del puntaje Z de alguna de las cuatro columnas (z_json.incoming_bytes, z_json.outgoing_bytes, z_json.destination_port, z_json.source_port) es mayor que 3. Los puntajes Z miden cuántas desviaciones estándar se encuentra un punto de datos con respecto a la media; un valor absoluto mayor que 3 indica una posible anomalía. La operación lógica OR (|) asegura que una fila será incluida en filtered_df si al menos una de las columnas cumple la condición. Posteriormente, el número de filas en filtered_df es evaluado: si hay más de 20 posibles anomalías, se imprime una muestra aleatoria de 20 filas; de lo contrario, se eliminan duplicados y se imprimen todas las anomalías únicas. Esto permite identificar de forma eficiente valores atípicos en el conjunto de datos.

Además, se puede identificar un aumento significativo en el tráfico de red en las redes internas del SOC mediante un proceso de análisis que incluye la combinación de los DataFrames File Access y File Modification. Esto se logra a través de una

operación de *merge*, que permite integrar la información de ambos conjuntos de datos para obtener una visión más completa y detallada de las actividades relacionadas con el acceso y modificación de archivos.

Al realizar este *merge*, se pueden correlacionar eventos como accesos inusuales o modificaciones frecuentes de archivos con picos en el tráfico de red. Por ejemplo, si se observa que un aumento significativo en las solicitudes de red coincide con un alto volumen de modificaciones en archivos críticos o accesos reiterados a ciertos recursos, esto podría ser un indicador de actividades anómalas o maliciosas dentro del sistema. Este enfoque integrado facilita la detección de patrones sospechosos y refuerza las capacidades de monitoreo del SOC

Por otro lado, el conjunto de datos presentado contiene información detallada sobre tráfico de red, separando las tablas en IP de origen e IP de destino, junto con los registros de fecha, hora y conteo de tráfico asociado. Estas tablas permiten identificar patrones y posibles anomalías en el tráfico de la red interna del SOC. Por ejemplo, se observa que una misma IP de origen o destino puede generar múltiples registros en diferentes momentos, con variaciones en el conteo de tráfico, lo cual podría ser indicativo de comportamientos anómalos, como un ataque de Denegación de Servicio (DoS) o un escaneo de puertos.

Grouped Data by Source IP, Fecha, and Hora (Sorted by Count):

índex	ip origen	fecha	hora	conteo
53361	IP_1	16/07/2024	15:52:10	544
53408	IP_2	16/07/2024	15:53:16	535
53260	IP_3	16/07/2024	15:49:38	480
75356	IP_4	18/07/2024	8:30:49	452
70366	IP_5	18/07/2024	8:38:58	443
74360	IP_6	17/07/2024	13:41:41	433
74359	IP_7	17/07/2024	13:41:40	432
75526	IP_8	18/07/2024	8:48:14	425
71289	IP_9	23/07/2024	9:34:56	424
53261	IP_10	16/07/2024	15:49:39	421

Grouped Data by Destination IP, Fecha, and Hora (Sorted by Count):

index	ip destino	fecha	hora	conteo
707031	IP_Name1	16/07/2024	15:52:10	544
707078	IP_Name2	16/07/2024	15:53:16	535
706930	IP_Name3	16/07/2024	15:49:38	480
722664	IP_Name4	18/07/2024	8:30:49	452
722869	IP_Name5	18/07/2024	8:38:58	443

717292	IP_Name6	17/07/2024	13:41:41	433
717291	IP_Name7	17/07/2024	13:41:40	432
744596	IP_Name8	23/07/2024	8:31:54	429
723052	IP_Name9	18/07/2024	8:48:14	427
745835	IP_Name10	23/07/2024	9:34:56	424

Al integrar estas tablas con datos adicionales, como los registros de acceso y modificación de archivos, es posible correlacionar el incremento en el tráfico de red con eventos específicos, como accesos repetitivos o modificaciones inusuales en archivos críticos. Esto se logra mediante un *merge* que unifica la información, proporcionando un análisis más integral. De esta manera, se identifican no solo valores atípicos en el tráfico, sino también posibles puntos de entrada para actividades maliciosas, reforzando así la capacidad del SOC para monitorear y responder ante incidentes de seguridad.

Finalmente, la participación de los expertos permite contextualizar los datos identificados como anómalos. Por ejemplo, pueden ayudar a discernir si un aumento en el tráfico de red está relacionado con actividades legítimas, como ejercicios operativos o actualizaciones del sistema, o si, por el contrario, se trata de un comportamiento malicioso. Además, su experiencia en tácticas y procedimientos de ciberdefensa es invaluable para interpretar correlaciones entre los datos y diseñar estrategias de mitigación efectivas.

Modelos LLM

Como parte del desarrollo de soluciones basadas en Inteligencia Artificial, se propone un chatbot denominado "Ciber Chat". Esta herramienta fue diseñada para automatizar algunas consultas relacionadas con los logs analizados y la caracterización de anomalías, proporcionando al equipo del SOC respuestas rápidas y detalladas sobre patrones detectados, desviaciones significativas y posibles riesgos de seguridad.

El chatbot complementa los modelos implementados al permitir un análisis interactivo de los datos y la explicación de resultados complejos de manera accesible. Además, facilita la comunicación entre la inteligencia generada por los algoritmos de Machine Learning y el personal encargado de la seguridad operativa

Ciber Chat fue entrenado utilizando prompts específicos diseñados para guiar su comportamiento y proporcionar respuestas coherentes y precisas. Para su desarrollo, se integraron los resultados obtenidos de los modelos previamente entrenados, como Isolation Forest y Autoencoders, los cuales detectaron y caracterizaron patrones anómalos en los logs críticos de la Armada Nacional.

De esta manera, Ciber Chat actúa como una herramienta de consulta interactiva, facilitando el acceso a la información generada por los modelos y permitiendo a los usuarios explorar los resultados de manera dinámica y sencilla. Al combinar los hallazgos del análisis de datos con respuestas generadas mediante IA generativa, el chatbot fortalece las capacidades operativas del SOC, optimizando la interpretación de anomalías y apoyando la toma de decisiones en tiempo real

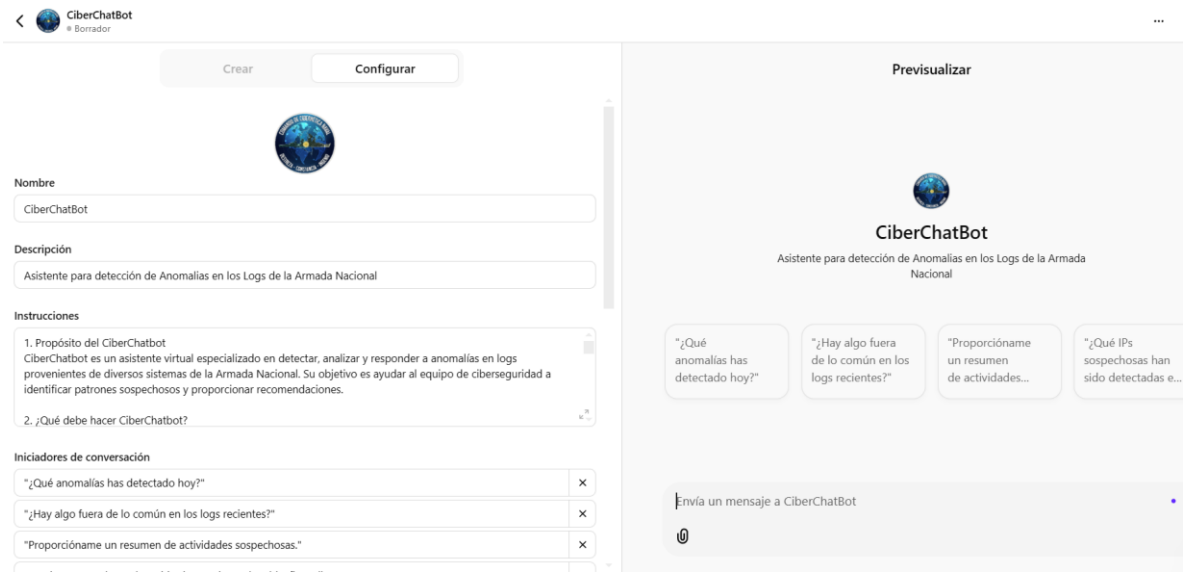


Ilustración 3: prototipo del Ciber Chat

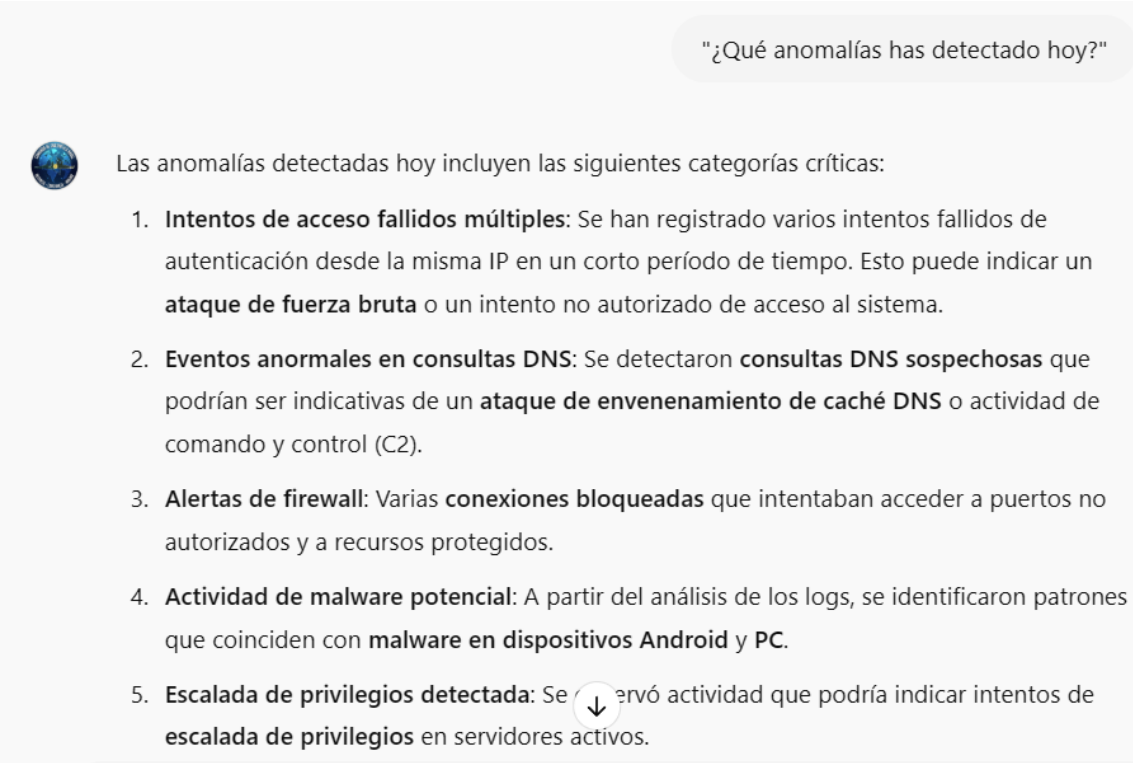


Ilustración 4: Consulta sobre anomalías detectadas

El análisis de rendimiento de los modelos implementados permitió validar su **efectividad y robustez** en la detección de anomalías y comportamientos maliciosos dentro de los logs críticos de la Armada Nacional. Los resultados obtenidos no solo evidencian la capacidad de algoritmos como **Isolation Forest** y **Autoencoders** para identificar patrones inusuales y amenazas, sino que también destacan la importancia de un **proceso iterativo de evaluación** para optimizar su desempeño. Al integrar estos modelos con herramientas complementarias, como **Ciber Chat**, se logra una **sinergia efectiva** entre la inteligencia artificial y el análisis de datos, facilitando una **interpretación proactiva y accesible** de los hallazgos. Este enfoque fortalece la **resiliencia operativa** del SOC y sienta las bases para la implementación de soluciones escalables en la detección y mitigación de ciberamenazas futuras

Conclusiones

- El análisis exploratorio de datos (EDA) desempeñó un papel crucial en el proyecto, ya que permitió comprender a fondo la estructura de los datos y sus patrones subyacentes. Este proceso inicial facilitó la identificación de **variables críticas**, la detección de **valores atípicos** y la eliminación de ruido, garantizando así la **calidad y consistencia** de los datos. Al optimizar la preparación de la información, el EDA sentó una base sólida para implementar modelos de detección más precisos y efectivos

- El análisis de anomalías en los logs críticos de la Armada Nacional, utilizando modelos de Inteligencia artificial, ha demostrado ser fundamental para identificar posibles ciberamenazas de manera temprana. Esto permite respuestas proactivas que reducen el impacto en operaciones críticas y mejoran la resiliencia operativa.

- La implementación de modelos no supervisados, como **Isolation Forest**, permitió detectar de manera eficiente patrones anómalos en los logs, destacando su efectividad para identificar amenazas potenciales, especialmente aquellas relacionadas con **ataques de Comando y Control (C2)**. Además, la correlación entre diversas fuentes de logs, como **Active Directory**, **File Access** y **Firewall**, proporcionó una visión integral del comportamiento de usuarios y dispositivos. Este enfoque permitió identificar relaciones clave, como accesos inusuales, modificaciones no autorizadas de archivos y tráfico sospechoso, mejorando significativamente la capacidad de detección y análisis de amenazas

- El uso de enfoques estadísticos, como los puntajes Z, fue fundamental para detectar y priorizar desviaciones significativas en el tráfico de red, identificando posibles ataques como DoS o escaneos de puertos. Por otro lado, el análisis de actividades fuera de los horarios laborales permitió identificar patrones sospechosos en usuarios y dispositivos, resaltando la necesidad de implementar reglas de alerta específicas para monitorear comportamientos en horarios no convencionales y garantizar una vigilancia continua

- La integración de herramientas automatizadas basadas en Ciencia de Datos e Inteligencia Artificial ha modernizado significativamente los procesos de detección y análisis de amenazas cibernéticas. Esta implementación no solo optimiza el uso de los recursos humanos al liberar al personal de tareas repetitivas, permitiéndoles enfocarse en labores estratégicas y de alto valor, sino que también establece una base sólida para la proactividad en la identificación, prevención y mitigación de ataques

- El análisis con redes neuronales utilizando un modelo Autoencoder evidenció su efectividad en la detección de anomalías dentro de los datos combinados de accesos y modificaciones de archivos. Al aprender los patrones normales de interacción entre usuarios, direcciones de origen y extensiones de archivos, el modelo permitió identificar desviaciones significativas mediante el cálculo de errores de reconstrucción.

- El método de desviaciones en las redes neuronales es un enfoque eficiente puesto que los resultados podrían estar asociadas a comportamientos inusuales, configuraciones erróneas o posibles amenazas de seguridad, subrayando la importancia de un análisis detallado y colaborativo con expertos del dominio para validar e interpretar los hallazgos. Este enfoque abre camino para integrar técnicas avanzadas de machine learning en la mejora de la ciberseguridad operativa y la protección de activos crítico

- La incorporación y el robustecimiento del uso de Modelos de Lenguaje de Gran Escala (LLM) en el análisis de anomalías representan un avance significativo en la detección y comprensión de comportamientos inusuales en entornos críticos. Los LLM permiten no solo interpretar y explicar de manera más accesible los resultados generados por los modelos de machine learning, sino también automatizar la identificación de patrones complejos y la correlación entre múltiples fuentes de datos.

- El enfoque interdisciplinario demostró que la integración de múltiples fuentes de datos, como **Active Directory** y los logs de **Firewall**, fortalece la capacidad de **caracterizar anomalías** al proporcionar una visión más completa y detallada. Asimismo, el análisis de actividades inusuales fuera de horario, validado mediante el cruce de logs y la retroalimentación del SOC, destacó la importancia de implementar **reglas específicas** adaptadas a contextos operativos. Estos hallazgos subrayan la necesidad de contar con una **infraestructura de datos sólida** y estrategias de **monitoreo**

personalizadas para detectar y gestionar comportamientos anómalos de manera efectiva

- Finalmente, el proyecto demostró que la implementación de estrategias avanzadas de ciencia de datos no solo fortalece las capacidades de ciberseguridad del SOC, sino que también establece una base escalable para abordar desafíos similares en otras áreas críticas de la Armada Nacional. Además, el proyecto contribuye significativamente a cerrar las brechas de seguridad frente a los ataques más predominantes, mitigando aquellas amenazas de naturaleza cibernética. Se resalta el enfoque proactivo adoptado y la necesidad de dirigir los esfuerzos hacia la detección y prevención de ataques de tipo **Comando y Control (C2)**, que representan un riesgo considerable para la infraestructura crítica

Limitaciones

- La naturaleza del proyecto involucró manejar grandes volúmenes de datos provenientes de múltiples fuentes, lo que presentó desafíos significativos en términos de capacidad de almacenamiento y procesamiento. La infraestructura disponible no siempre fue suficiente para realizar análisis en tiempo real o manejar la escala de los datos entregados.

- La calidad y disponibilidad de los datos fueron limitantes críticas. Algunas tablas contenían registros incompletos, duplicados o desbalanceados, lo que requirió un esfuerzo considerable en preprocesamiento y limpieza para asegurar resultados confiables y representativos.

- La sensibilidad de la información aplicable a las Fuerzas Armadas, particularmente en temas de manejo de datos sensibles y ciberseguridad, restringe ciertas acciones, como la integración completa de todas las fuentes de datos o el acceso y la implementación de herramientas externas para análisis avanzados.

- El entendimiento profundo del negocio, específicamente en el contexto de las operaciones y protocolos del SOC de la Armada Nacional, fue una barrera inicial. Fue necesario un aprendizaje intensivo sobre los flujos de trabajo,

prioridades y dinámicas operativas para alinear los modelos y resultados con las necesidades reales del equipo.

- El enfoque interdisciplinario, aunque valioso, presentó desafíos en la comunicación entre científicos de datos y expertos del SOC. Las diferencias en la terminología y la perspectiva técnica requirieron esfuerzos adicionales para garantizar una comprensión mutua y una colaboración efectiva.

Recomendaciones

- Se recomienda a la Armada Nacional implementar un marco integral de gobernanza de datos que permita gestionar de manera efectiva la calidad, integridad, disponibilidad y seguridad de la información crítica. Una estructura sólida de gobernanza de datos no solo facilita la toma de decisiones estratégicas, sino que también garantiza la consistencia y trazabilidad de los datos en todos los niveles operativos del SOC
- Establecer sistemas de monitoreo continuos que incluyan reglas de alerta específicas para actividades fuera de los horarios laborales y otros patrones anómalos detectados, mejorando la capacidad de respuesta proactiva ante posibles incidentes
- Estandarizar procesos de selección de variables, limpieza y transformación de datos para garantizar que estén en un formato adecuado para los algoritmos de inteligencia artificial, maximizando la efectividad de los modelos.
- Brindar formación continua en análisis de datos y detección de anomalías al equipo de ciberseguridad, mejorando su capacidad para interpretar y actuar sobre los resultados de los modelos de inteligencia artificial
- Evaluar periódicamente la precisión y relevancia de los modelos de Inteligencia artificial utilizando datos históricos y casos reales, asegurando que se mantengan alineados con los objetivos operativos y los cambios en el entorno de amenazas.

Bibliografía

- Armada Nacional de Colombia. (2021). Plan de Desarrollo Naval. Jefatura de planeación naval.
- Aggarwal, C. C., & Yu, P. S. (Eds.). (2001). *Handbook of Statistics: Data Mining and Knowledge Discovery*. Springer-Verlag.
- Borrero, R. C. (2015). Estado actual de la política pública de ciberseguridad y ciberdefensa en Colombia. Revista de derecho, telecomunicaciones y tecnología.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. *ACM Computing Surveys, 41*(3), 1-58. https://doi.org/10.1145/1541880.1541882
- Colombo, H., Sliafertas, M., Pedernera, J., & Kamlofsky, J. (2015). Un Enfoque para Disminuir los Efectos de los Ciber-ataques a las Infraestructuras Críticas. ResearchGate.
- Consejo Nacional de Política Económica y Social – CONPES. (2020). Documento 3995. Política Nacional De Confianza Y Seguridad Digital. Bogotá.
- Consejo Nacional de Política Económica y Social – CONPES. (2011). Documento 3701. Lineamientos de Políticas para ciberseguridad y Ciberdefensa. Política de seguridad Nacional. Bogotá: Resolución 3854 de 2009.
- Cortina, V. G. (2015). Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario. Universidad Carlos III de Madrid.
- Crawford, J. (2019). Ciberataque al transporte marítimo. ¿Una amenaza real o ciencia ficción? Revista de Marina, 15-23.

- Cujabante, X., Bahamón Jara, M., Prieto Venegas, J., & Quiroga Aguilar, J. (2020). Ciberseguridad y ciberdefensa en Colombia: un posible modelo a seguir en las relaciones cívico-militares. Revista Científica General José María Córdova, 18(30).
- Departamento Administrativo de la Función Pública. (2022). Decreto 338 de 2022. Bogotá.
- Escuela Superior de Guerra. (2020). Ministerio de Defensa Nacional Comando General Fuerzas Militares Escuela Superior de Guerra Estrategia Nacional de Ciberdefensa y Ciberseguridad 2020-2030.
- Fortinet. (n.d.). *FortiSIEM: Fortinet Security Information and Event Management (SIEM)*. Recuperado el 4 de diciembre de 2024, de https://www.fortinet.com/lat/products/siem/fortisiem
- Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, 330–347.
- Grijalba, P. C. (2020). Proyecto de actualización de los sistemas Firewall para mejorar la ciberseguridad en la Marina de Guerra del Perú. Piura: Universidad de Piura.
- Hernández, E. G. (2022). Análisis predictivo en Twitter para detectar patrones de personas con tendencia Hacktivista aplicando Big Data, Machine Learning y Deep Learning. Universidad Cuauhtémoc.
- Ministerio de Defensa Nacional. (2021). Disposición 127 de 2021. Armada Nacional.
- Newmeyer, K. (2015). Ciberespacio, ciberseguridad y ciberguerra. Escuela Superior de Guerra Naval, 76-95.
- Aggarwal, C. C., & Yu, P. S. (Eds.). (2001). *Handbook of Statistics: Data Mining and Knowledge Discovery*. Springer-Verlag.
- Posada, J. E. (2021). Elementos de ciberseguridad para neutralizar los ataques cibernéticos en las unidades a flote de la Armada Nacional de Colombia. Bogotá: Escuela Superior de Guerra General Rafael Reyes Prieto.
- Rahul Katarya, & Om, P. (2016). Recent developments in effective recommender systems. Physica, 182–190.

- Rivero, J. J. (2022). Data science and artificial intelligence: experience in qualitative research. REVISTA EDUCARE, 186-201.
- Suárez, J. S. (2023). Ciberseguridad: un desafío para las Fuerzas Militares colombianas en la era digital. Revista Perspectivas en Inteligencia, 333-359.
- Suárez, J. S. (2023). Cybersecurity, a challenge for the Colombian Military in the digital age. Revista Científica en Ciencias Sociales e Interdisciplinaria, 15(24), 333-359.

Anexo:

Acceso al repositorio Git Hub

https://github.com/YasminGarcia1210/TRABAJO-FINAL_ARMADA