KEEPCODING

Práctica de módulo

Aplicación del Machine Learning en la Ciberseguridad

por Javier González Espinoza

Date: November 19, 2023
Módulo: Machine Learning & Cibersecurity
Profesor: Gabriel Valverde

Contents

I. Detalles e indicaciones

1. Tareas a realizar

A la hora de evaluar lo aprendido a lo largo de estas semanas lectivas se plantean una serie de tareas a realizar en las que poner en práctica lo aprendido y demostrarse a uno mismo que la potencia del Machine Learning puede ser determinante en proyectos de ciberseguridad.

Se plantean dos tareas con enfoque dispar:

- Realizar la descripción de un caso de uso de Machine Learning en el ámbito de la ciberseguridad.
- Desarrollar un modelo predictivo siguiendo las pautas aprendidas.

a. Diseño de un casos de uso de ML en Ciber Seguridad

A lo largo del módulo se han ido estudiando distintos métodos y algoritmos útiles en la resolución de distintos problemas de ciberseguridad a partir de datos. También hemos aprendido la metodología de desarrollo y hemos estudiado las distintas claves que permiten evaluar el éxito del proyecto.

Como expertos en ciberseguridad vuestro rol será fundamental a la hora de establecer las bases, evaluar y realizar el seguimiento de distintas soluciones de Machine Learning que se puedan plantear como solución ante un problema de seguridad. Vuestro trabajo en muchos de estos proyectos será definir adecuadamente este documento y aseguraros de que se van cumpliendo las especificaciones haciendo uso del conocimiento adquirido en este módulo.

Por lo tanto el objetivo de la tarea es plantear la defincón de un caso de uso que se pudiera presentar como proyecto potencial a realizar con ML en una compañía siguiendo las intrucciondes descritas en el documento: Conceptualización_CasosDeUso_BigData_v20.pdf

b. Resolver usando python uno de los siguiente problemas de ML planteados

i) El dataset CTU-13

Es un dataset con distintos tipos de tráficos en red: Botnet, Normal y Background traffic. El objetivo será identificar el tipo de tráfico que está habiendo en base a las variables recogidas y otras que podáis generar.

El conjunto de datos utilizado será el que podéis encontrar en el siguiente link, donde además podemos ver su estructura.

Se evaluará el resultado pero también el trabajo realizado en base a la metodología Data Science. Web oficial de la base de datos

De todos los escenarios posibles se ha determinado trabajar con el escenario 7 por contener menos datos y se recomenda utilizar los ficheros csv que se adjuntan. Fuente final de datos: Escenario 7

Encontramos varios ficheros, aunque de todos se puede extraer información valiosa, recomendamos utilizar simplemente: capture20110816-2.binetflow.2.format Fichero csv que contiene información sobre las distintas conexiones con una variable objetivo final en la columna flow que determina el background. Enlace a drive.

Nota: por la complejidad del tratamiento y parseado de datos se plantea como alternativa trabajar en el siguiente problema cuyos datos están pretratados.

ii) Detección de malware en Android

En este caso de uso práctico se pretende resolver un problema de detección de malware en dispositivos Android mediante el análisis del tráfico de red que genera el dispositivo mediante el uso de conjuntos de árboles de decisión.

Descripción: El sofisticado y avanzado malware de Android es capaz de identificar la presencia del emulador utilizado por el analista de malware y, en respuesta, alterar su comportamiento para evadir la detección. Para solucionar este problema, instalamos las aplicaciones de Android en el dispositivo real y capturamos su tráfico de red. El conjunto

de datos de CICAAGM se captura instalando las aplicaciones de Android en los teléfonos inteligentes reales semiautomatizados. El conjunto de datos se genera a partir de 1900 aplicaciones con las siguientes tres categorías:

- **Adware (250 apps)**
  - Airpush: Designed to deliver unsolicited advertisements to the user's systems for information stealing.
  - Dowgin: Designed as an advertisement library that can also steal the user's information.
  - Kemoge: Designed to take over a user's Android device. This adware is a hybrid of botnet and disguises itself as popular apps via repackaging.
  - Mobidash: Designed to display ads and to compromise user's personal information.
  - Shuanet: Similar to Kemoge, Shuanet also is designed to take over a user's device.
- **General Malware (150 apps)**
  - AVpass: Designed to be distributed in the guise of a Clock app.
  - FakeAV: Designed as a scam that tricks user to purchase a full version of the software in order to re-mediate non-existing infections.
  - FakeFlash/FakePlayer: Designed as a fake Flash app in order to direct users to a website (after successfully installed).
  - GGtracker: Designed for SMS fraud (sends SMS messages to a premium-rate number) and information stealing.
  - Penetho: Designed as a fake service (hacktool for Android devices that can be used to crack the WiFi password). The malware is also able to infect the user's computer via infected email attachment, fake updates, external media and infected documents.
- **Benign (1500 apps)**
  - 2015 GooglePlay market (top free popular and top free new)
  - 2016 GooglePlay market (top free popular and top free new)
  - Ficheros de datos pcap files – el tráfico de red del malware y benigno (20% malware y 80% benigno)
    http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/PCAPs/03-11/PCAP-03-11.zip.csv

files - la lista de características de tráfico de red extraídas generadas por el CI-Cflowmeter
http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/CSVs/CSV-03-11.zip.
Se recomienda de nuevo usar csv.

- Descarga de los ficheros de datos. Enlace de descarga Enlace de descarga2 Fichero csv en drive
- Referencias adicionales sobre el conjunto de datos Arash Habibi Lashkari, Andi Fitriah A. Kadir, Hugo Gonzalez, Kenneth Fon Mbah and Ali A. Ghorbani, “Towards a Network-Based Framework for Android Malware Detection and Characterization”, In the proceeding of the 15th International Conference on Privacy, Security and Trust, PST, Calgary, Canada, 2017.

II. Diseño de un caso de uso de ML en Ciberseguridad

1. Descripción del caso de uso

Al día de hoy, una de las herramientas conceptuales empleadas en el ámbito de la ciberseguridad para evaluar y gestionar los riesgos asociados a la seguridad de la información es el *Cubo de McCumber*, creado por *John McCumber* en el año 1991. La hipótesis de este concepto es que la seguridad de la información no puede ser reducida a un solo elemento o componente, sino que debe involucrar diversas nociones las cuales están interconectadas entre si. Estas nociones se representan como un cubo, en donde cada una de sus caras aborda un aspecto específico de la seguridad de los datos.

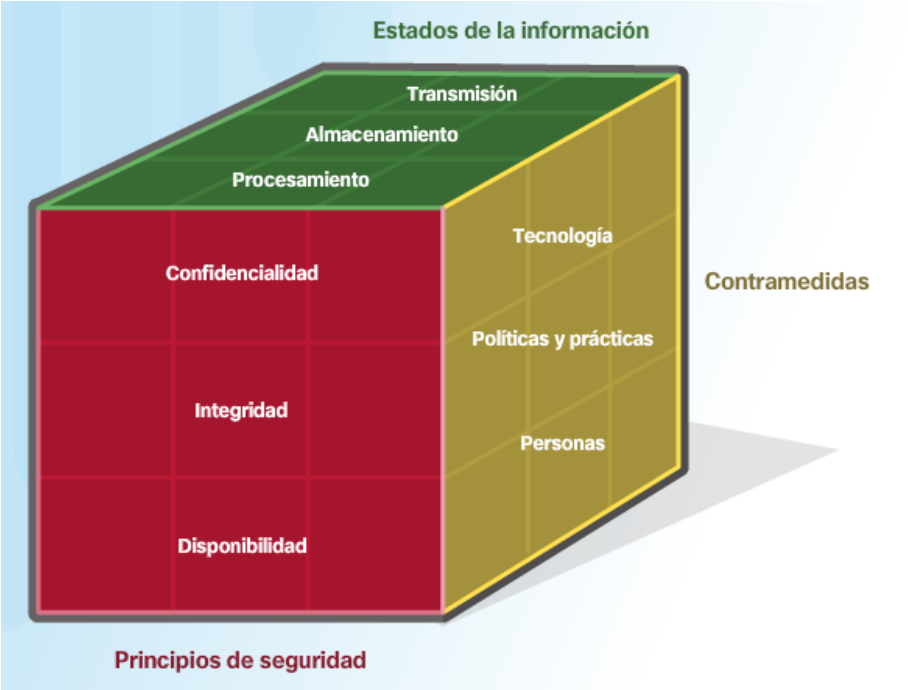

Fig. 1: Cubo de McCumber.

Como se aprecia en la imagen anterior, los aspectos fundamentales de la seguridad de la información planteados por McCumber son:

- **Los principios de seguridad:**
  - *Confidencialidad:* protección de la información contra el acceso no autorizado a esta, garantizando que datos sensibles o clasificados no estén al alcance de usuarios o sistemas no autorizados.
  - *Integridad:* exactitud y completitud de la información, en donde se previene que los datos sean modificados de forma no autorizada.
  - *Disponibilidad:* permite mantener el acceso y uso de la información cuando sea necesario. Gracias a este concepto es que que los sistemas y los datos siempre están disponibles para aquellos usuarios autorizados que lo necesiten, evitando interrupciones no planificadas.

- **Los estados de la información**
  - *Procesamiento:* datos que se utilizan para realizar una operación, como actualización de un registro de base de datos (datos en proceso) o mejoras en un servicio establecido.
  - *Almacenamiento:* el almacenamiento hace referencia a los datos almacenados en memorias o dispositivos de almacenamiento permanente, con la finalidad de generar persistencia de estos para su uso.
  - *Transmisión:* son los datos que se transmiten entre diversos sistemas, a través de los cuales se gestiona el envío y la recepción de información en general.

- **Las contramedidas o medidas de seguridad**
  - *Concientización:* la capacitación y la educación son las medidas que las empresas implementan con sus usuarios y trabajadores con la finalidad de garantizar la correcta información acerca de las ciberamenazas existentes y las acciones que se pueden y deben llevar a cabo para securizar y proteger los sistemas y los datos.
  - *Tecnología:* soluciones basadas en software y/o hardware, las cuales son diseñadas para proteger los sistemas de información.
  - *Políticas y procedimientos:* controles administrativos que proporcionan una base específica para el como una organización implementa el aseguramiento de la información, tales como planes de respuesta ante incidentes y pautas de mejores prácticas.

En función a esto, y con el objetivo de garantizar la protección de la información en todas sus dimensiones, se plantea como caso de uso de las tecnologías implementadas a través del Machine Learning para el ámbito de la ciberseguridad, la ***"detección de trá-***

fico anómalo o inusual en redes internas (red local) empresariales". El planteamiento de dicho caso se ha realizado siguiendo la asignación de datos solicitados en el documento entregado por el profesor del módulo, *Conceptualización_CasosDeUso_BigData_v20.pdf*.

a. ¿Cuál es el problema?

El problema principal es el tráfico anómalo presente en las redes locales de grandes corporaciones, el cual puede ir desde ataques ciberneticos realizados por ciberdelincuentes externos o internos a la corporación con el propósito de comprometer la información en todas las dimensiones (desde problemas en la disponibilidad de distintos servicios, hasta el robo o secuestro de información confidencial), errores humanos en la administración de los diversos sistemas presentes en la infraestructura, corrupción de software, falla en el hardware, entre otros. La detección de este tráfico es importante para proteger la red y sus sistemas de información.

El tráfico anómalo puede tener una serie de efectos negativos en una red, entre los que se incluyen:

- Reducción del rendimiento de la red, generando la saturación en la transmisión de información y provocando la disminución del rendimiento fundamental en esta.
- Pérdida de datos tras la intrusión de ciberdelincuentes, o corrupción de datos debido a una dismunición en los niveles de funcionamiento de la red o problemas en producción del hardware, lo que presenta un impacto negativo en las operaciones de la organización y la disponibilidad de la información.
- Incidentes de seguridad en los cuales se vean comprometidos sistemas y servicios internos de la organización.

Debido a lo anterior es que nace la necesidad de detectar todo el posible tráfico no deseao en una red. Un sistema de detección eficaz puede ayudar a proteger la red y sus sistemas de información de ataques cibernéticos, errores humanos, mal funcionamiento de hardware o cualquier dimensión que genere un nivel de compromiso sobre la información de la entidad.

b. ¿Cómo se está afrontando ahora?

En la actualidad, la detección de tráfico anómalo se realiza de forma manual o mediante

sistemas basados en reglas (firewalls, IDS, IPS, etc). Por su parte, la detección manual es un proceso que lo realiza un analista de seguridad o de redes, en donde se revisan los datos del tráfico recolectados gracias a analizadores de protocolos de red (como Wireshark o Tshark) o gestores de eventos (como logs de sistema almacenados) en busca de patrones inusuales. Este proceso puede ser muy laborioso, costoso y requiere un alto nivel de experiencia, conocimiento, tiempo y esfuerzo humano. Por otro lado, los sistemas basados en reglas son relativamente sencillos de implementar, pero pueden ser poco precisos debido a la generalidad de estas, y no pueden adaptarse a nuevos tipos de ataques sin que se realice la modificación de su configuración y/o parametrización. Estas reglas pueden ser definidas directamente por el administrador del sistema o el equipo de TI Security, o proporcionadas por un proveedor de seguridad externo que gestione esta temática.

En el año 2022, un ataque cibernético a la empresa Colonial Pipeline provocó la interrupción del suministro de combustible a gran parte de la costa este de los Estados Unidos. Los atacantes se infiltraron en la red interna de la corporación tras descubrir una contraseña para conexión por VPN expuesta en internet. El ataque se realizó mediante la implantación de un ransomware sobre los sistemas, el cual mantuvo cautiva la información del 6 al 12 de mayo de ese año. Los atacantes ganaron acceso a los sistemas de la organización ya que no se detectó el tráfico anómalo a tiempo, generando la perdida de mas de 100 gigabytes de información y un pago a los ciberdelincuentes de 75 bitcoins ($4.4 millones de dólares).

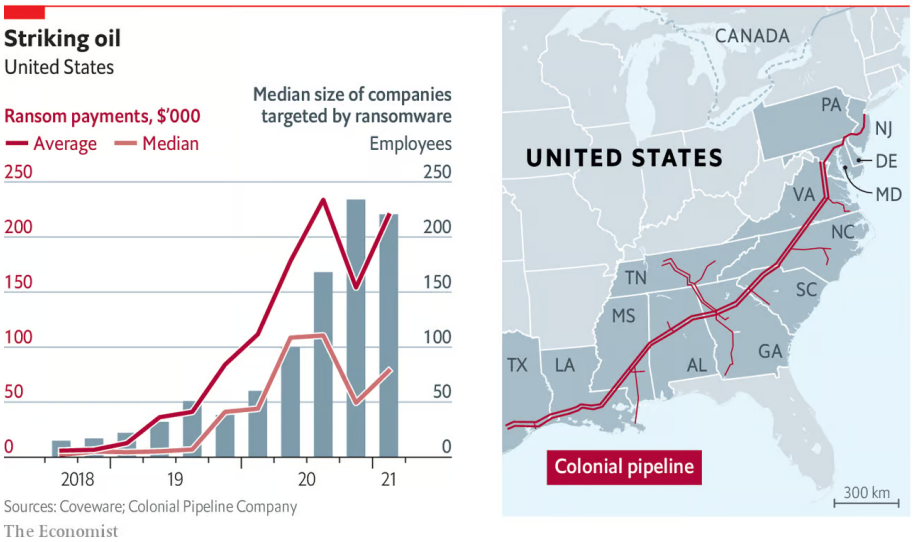

Fig. 2: Ransomeware attack sobre Colonial Pipeline.

En 2023, un ataque cibernético a la empresa Sony Pictures dió como consecuencia la filtración de una gran cantidad de datos confidenciales, incluyendo datos personales de empleados e información sobre material filmográfico en proceso de estreno. El ataque se

realizó mediante la infiltración de un malware a los sistemas internos, el cual logró pasar desapercibido por los sistemas de detección al encontrarse oculto en archivos completamente legítimos.

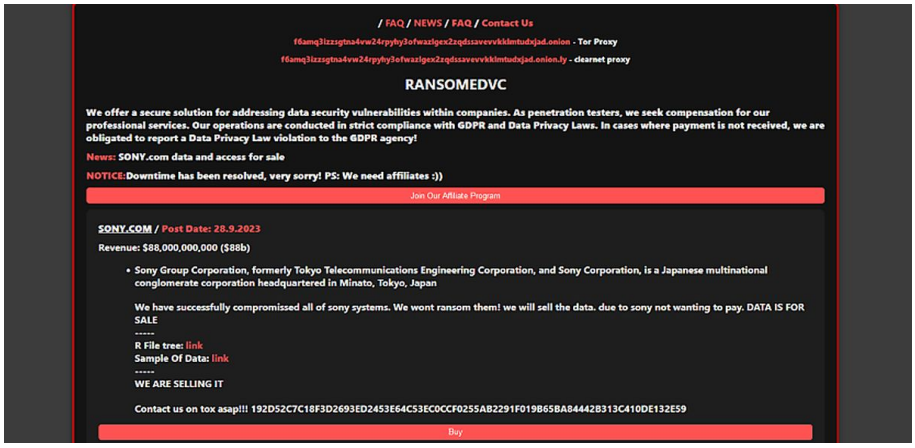

Fig. 3: Ransomeware attack con data breach sobre Sony Pictures.

Estos son solo dos ejemplos de los muchos ataques cibernéticos que se han producido y se producen a diario como resultado de la no o tardía detección de tráfico extraño sobre sistemas internos de una corporación. Es importante que las organizaciones implementen medidas efectivas para proteger sus sistemas de información.

c. ¿Cuál es la acción que buscamos poder hacer para solucionar el problema?

Enfocando la solución hacia la detección de ciberamenazas internas o externas, la implementación de un sistema de detección de tráfico basado en Machine Learning es una solución prometedora, la cual permite abordar el problema de la tardía o nula detección de tráfico anómalo en sistemas empresariales, debido a que ofrece una serie de ventajas sobre los sistemas basados en reglas o por detección manual. Entre estas ventajas se encuentran:

- Mayor precisión: Los sistemas de detección de tráfico basado en Machine Learning pueden aprender patrones de tráfico específicos, los cuales pueden llegar a ser difíciles de detectar con sistemas basados en reglas generales.
- Capacidad de adaptación: Los sistemas de detección de tráfico basados en machine Learning pueden adaptarse a nuevos tipos de ataques desarrollados por ciberdelin-

cuentes o personal de ciberseguridad.

- Eficiencia: La detección de tráfico basada en el Machine Learning entrega una mayor velocidad de detección de tráfico anómalo y es capaz de analizar el tráfico de una red de una forma mas eficiente que el análisis manual de datos. Además se genera la reducción de falsos positivos si el algoritmo es entrenado de manera correcta y constante.

De igual manera, la importancia de poder recopilar y analizar datos de tráfico de una forma eficiente y adecuada, así como la utilización de un conjunto de datos de entrenamiento representativo presenta una mejora a los sistemas actuales. Pero a pesar de estas ventajas, existen algunos desafíos que deben tenerse en cuenta:

- Cantidad de datos: se requiere una gran cantidad de datos de entrenamiento para poder enseñar patrones de tráfico generales y específicos al algoritmo.
- Complejidad: la implementación de un sistema de detección de tráfico basado en ML puede ser complejo de implementar y mantener. Esto requiere un grupo de especialistas que posean los conocimientos adecuados para estar constantemente interactuando con el algoritmo, con el fin de mantener activo y actualizado tanto el sistema de aprendizaje y pruebas en desarrollo como la implementación en el sistema de producción.
- Coste: este tipo de implementaciones basadas en Machine Learning pueden ser costosos de implementar y mantener, lo cual puede suponer un desafío para organizaciones con presupuestos limitados.

d. KPIs – Indicadores de negocio

La selección de los KPIs adecuados para este caso de uso va a depender directamente de las necesidades específicas de la organización. Los principales para este caso de uso son:

- Reducción del tráfico inusual sobre los sistemas internos de una compañía. Este KPI es el mas importante debido a que la no detección de tráfico anómalo genera un impacto negativo en los negocios, entre los cuales se pueden incluir la interrupción del servicio mediante ataques de denegación de servicio, rapto de sistemas, filtración de información confidencial, entre otros.
- Reducción de falsos positivos en donde se detecta tráfico potencialmente peligroso sin serlo. Los falsos positivos pueden generar una gran cantidad de trabajo analítico

de forma innecesaria, provocando la perdida del enfoque hacia los eventos realmente importantes que se presentan sobre la red.

- Reducción del tiempo de respuesta ante un evento con altos niveles de peligrosidad gracias a la rápida respuesta del algoritmo. Gracias a esto se pueden tanto anular la intrusión por parte de ciberdelincuentes detectados a tiempo, como disminuir el impacto generado tras la intrusión debido a la detección y mitigación oportuna del ataque.

e. ¿Cuáles son los mínimos que se esperan de este caso de uso?

Se espera en primera instancia que el sistema sea capaz de detectar el tráfico extraño con un nivel de precisión del 95%, para así garantizar que no se genere una excesiva cantidad de falsos positivos. Para esto se deben recopilar datos de entrenamiento que representen una amplia gama de los tipos de tráfico, incluyendo el normal, el anómalo y posibles falsos positivos. Así, se podrá dar con el mayor detalle posible un buen entrenamiento al algoritmo. Un nivel de precisión del 95% significa que de cada 100 eventos detectados como tráfico inusual, 95 de ellos pertenecerá realmente a este concepto y solo 5 serán falsos positivos. En las primeras instancias se espera una reducción de al menos el 50% del tráfico anómalo sobre la red, porcentaje que con las mejoras y el entrenamiento del algoritmo tome un ascenso por sobre el 75%.

En cuanto a términos financieros, esta reducción podría significar un ahorro de millones de dólares, ya que la mayoría de los ataques cibernéticos tienen un impacto económico negativo en las organizaciones debido al rapto o robo de información, costos de reparación, pérdida de datos, entre otros.

Por otro lado, una disminución en las intrusiones a los sistemas y una baja en la filtración de información debido a la detección inusual de tráfico en redes internas mejora significativamente la reputación de la organización, debido a que plantea una mejor imagen ante la protección de la data particular y la de sus usuarios.

Finalmente, en cuanto al cumplimiento normativo esta reducción podría ayudar a la organización a cumplir las normativas de seguridad internas y plantear una mejora ante los protocolos de respuesta ante incidentes que esta posea.

f. Validación: ¿Qué criterio se va a usar para decidir si la solución es aceptable?

- **Criterios cuantificables**

Para decidir si la solución es aceptable se deben definir criterios cuantificables, los cuales deben tener en cuenta ciertos factores importantes. Algunos ellos son:

- El sistema debe ser capaz de reducir el tráfico anómalo en una cantidad significativa. Por ejemplo, se podría establecer un objetivo de reducción del 50% en primera instancia, para que con entrenamientos y el aprendizaje futuro entregado sobre el algoritmo, este valor vaya en aumento.
- El número de falsos positivos debe ser inferior a un límite máximo aceptable. Por ejemplo, se puede plantear un límite de 10 falsos positivos por cada 100 eventos detectados como tráfico anómalo como máximo aceptable, tomando en cuenta este como el peor de los casos (menor cantidad de falsos positivos implicaría una mayor detección de situaciones no deseadas).
- El sistema debe ser capaz de detectar el tráfico anómalo en un tiempo razonable. Por ejemplo, se podría establecer un objetivo de tiempo de respuesta de 1 minuto, dentro del cual se logre confirmar el tráfico anómalo como evento real y no como falso positivo, que se informe adecuadamente sobre el evento descubierto y que se tomen las acciones pertinentes contra la fuente generadora de manera eficiente y eficaz dentro de este rango de tiempo.

- **Costes y beneficios**

Además de estos criterios también se debe considerar los costes y beneficios de la solución. Por ejemplo, un sistema que tiene un alto porcentaje de reducción del tráfico extraño pero también genera un alto número de falsos positivos puede no ser aceptable si los costes de detección y confirmación de estos falsos positivos son demasiado elevados.

- **Algoritmos de IA**

Para llevar a cabo este caso de uso, se pueden utilizar algunos de los siguientes algoritmos de inteligencia artificial escritos en Python como base para realizar pruebas iniciales y como comparativa de funcionamiento y eficiencia al desarrollar un algoritmo propio enfocado especialmente en las necesidades de una compañía:

- *Isolation Forest:* Este algoritmo es un método de aprendizaje automático no supervisado que se utiliza para detectar anomalías. El algoritmo funciona creando un bosque de árboles de decisión y mide la probabilidad de que un punto de datos pertenezca a un conjunto de datos normal. Los puntos de datos con una probabilidad baja de pertenecer al conjunto de datos normal se consideran anomalías.

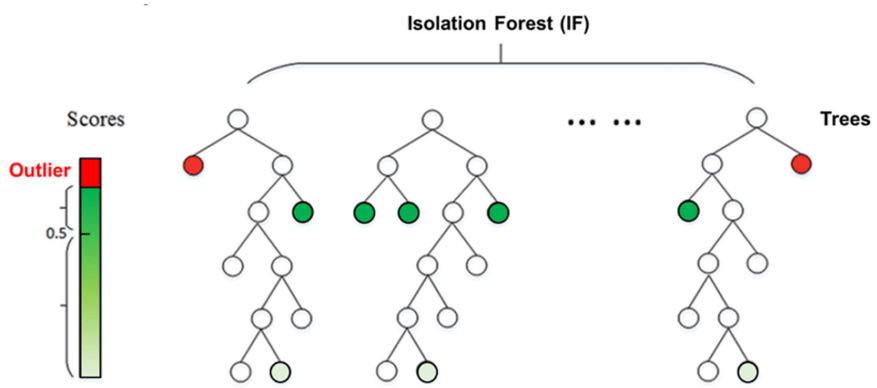

Fig. 4: Lógica para toma de decisiones de algoritmo Isolation Forest.

- *Local Outlier Factore:* Este algoritmo es otro método de aprendizaje automático no supervisado que se utiliza para detectar anomalías. El algoritmo funciona calculando la distancia entre un punto de datos y sus vecinos más cercanos. Los puntos de datos con una distancia significativamente mayor que la de sus vecinos se consideran anomalías.

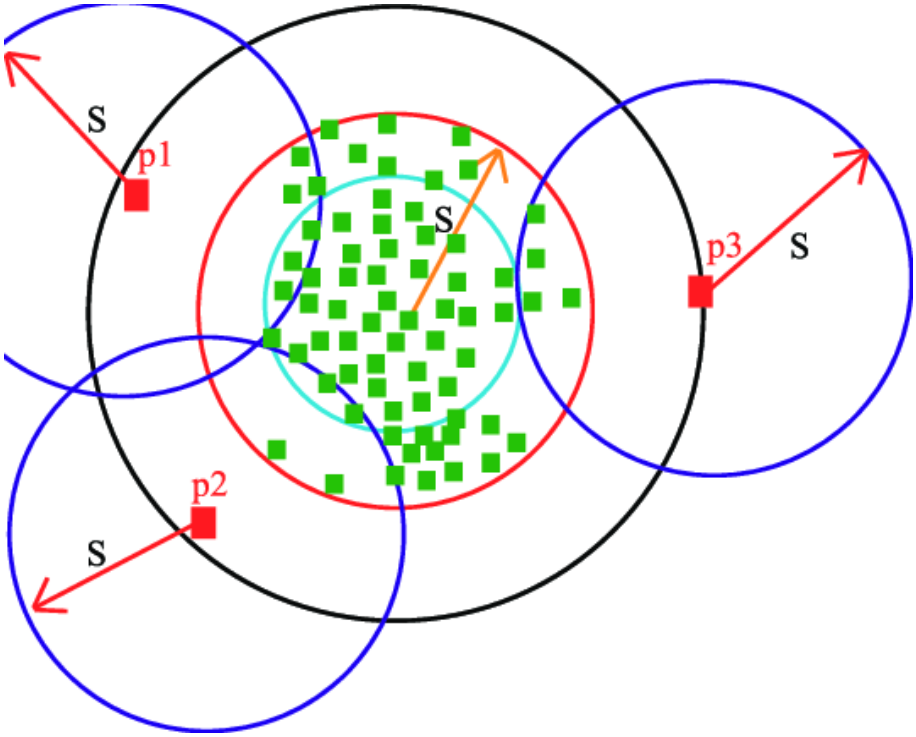

Fig. 5: Lógica para toma de decisiones de algoritmo Local Outlier Factor.

- *One-Class SVM:* Este algoritmo es un método de aprendizaje automático supervisado que se utiliza para aprender un modelo de un solo tipo de datos. El algoritmo se puede utilizar para detectar anomalías calculando la distancia entre un punto de datos y el modelo aprendido. Los puntos de datos con una distancia significativamente mayor que la del modelo se consideran anomalías.

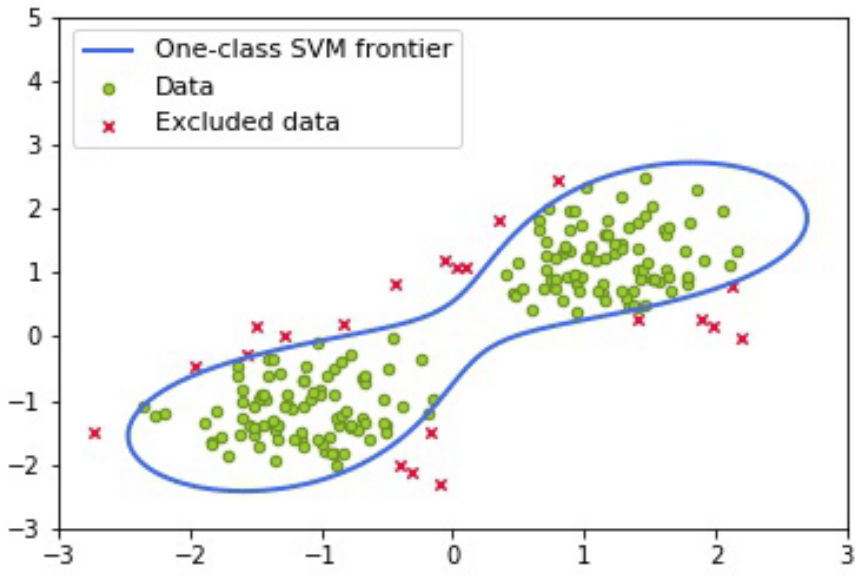

Fig. 6: Lógica para toma de decisiones de algoritmo One-Class SVM.

Estos algoritmos son de fácil implementación gracias a Python y han sido utilizados con éxito en distintas aplicaciones de detección de anomalías. La selección del algoritmo adecuado depende de los datos y del objetivo específico de la aplicación. Si los datos fueran de naturaleza multidimensional, el algoritmo Isolation Forest puede ser una buena opción. Si los datos son de naturaleza temporal, el algoritmo Local Outlier Factor podría resultar mas eficiente. Y por otra parte, teniendo en cuenta los costes y beneficios de cada algoritmo, One-Class SVM puede requerir una cantidad mayor de datos de entrenamiento para lograr un correcto funcionamiento y un alto desempeño sobre el objetivo principal.

g. Experimentación: ¿Cómo vamos a corroborar el funcionamiento?

Para corroborar y validar el correcto funcionamiento del sistema de detección de tráfico anómalo, se deben realizar las siguientes acciones experimentales en un entorno de desarrollo controlado y aislado a producción:

- Pruebas de precisión: utilizando datos de tráfico real y significativos, para así lograr evaluar la capacidad de detección del sistema ante tráfico que es normal para el sistema y el tráfico que debe acusar como no deseado.
- Pruebas de sensibilidad: utilizando conjuntos de datos de tráfico para evaluar la ca-

pacidad del sistema de detectar el porcentaje no permitido.

- Pruebas de tiempo de respuesta: mediante datos de tráfico real, se evalua la capacidad del sistema para detectar el tráfico anómalo dentro de un rango de tiempo indicado.

Estas acciones experimentales se deben realizar de forma periódica, con una frecuencia que dependerá de los resultados de cada prueba anterior implementada. Al principio se recomienda realizar pruebas de precisión, sensibilidad y tiempo de respuesta cada semana sobre el sistema de producción, con la finalidad de gestionar de una mejor forma el aprendizaje inicial del algoritmo, y buscar una mayor robustez ante la respuesta de incidentes. Para esto se deben establecer ciertos umbrales de precisión, sensibilidad y tiempo de respuesta. Con el pasar del tiempo se pueden implementar parches sobre el algoritmo mucho mas avanzados, disminuyendo las pruebas a meses o trimestres, siempre y cuando no se detecte un aumento considerable en la cantidad de tráfico anómalo sobre la red.

En otro tema, el tiempo necesario para verificar el funcionamiento del sistema dependerá directamente de la cantidad de datos que este deba procesar. Las pruebas de precisión y sensibilidad se pueden realizar en un tiempo relativamente corto, de unas pocas horas a unos pocos días. En cambio, las pruebas de tiempo de respuesta pueden tardar más, dependiendo de la cantidad de datos a procesar y de la complejidad del algoritmo de detección utilizado.

h. Productivización: ¿Qué salida debe tener la solución que se desarrolle?

La solución planteada debe tener una salida que sea útil para los agentes interesados, tal como una lista de eventos en donde se detallen las diferencias entre el tráfico normal sobre los sistemas y el tráfico extraño. Esto para que los analistas de seguridad investiguen estos eventos y gestionen las medidas de seguridad necesarias para mitigar situaciones similares.

La lista de eventos de tráfico anómalo puede incluir la siguiente información:

- Fecha y hora del evento
- Origen y destino del evento
- Tipo de evento

En caso de que la experimentación realizada se consolide con resultados satisfactorios, la solución se puede poner a disposición de los analistas de seguridad de una compañía

como un servicio web instaurado en la nube o en un servidor local, como una herramienta integrada en el sistema de seguridad prioritario de la empresa o estipulando un método que cumpla con los estandares de esta, dependiendo de las necesidades específicas de la organización.

2. Equipo de trabajo

a. Identificación de personas colaboradoras

En primera instancia, el equipo de trabajo necesario para llevar a cabo la selección y entrenamiento del algoritmo, las pruebas en entornos controlados de desarrollo y de producción y gestionar el análisis de los porcentajes de efectividad de este luego de la implementación, deberá estar formado por los siguientes miembros:

- Un responsable del proyecto (director de seguridad de la información, ingeniero de seguridad senior, consultor de ciberseguridad). Este estará encargado de coordinar el trabajo del equipo y gestionar los mecanismos propios de escalado que tenga cada una de las áreas. Además debe aportar sobre el grupo liderazgo, experiencia en aplicación del ML en el ámbito de la ciberseguridad y la capacidad de tomar la mejores desiciones hacia con el proyecto.
- Un ingeniero de Machine Learning (ingeniero de datos, científico de datos, ingeniero de aprendizaje automático), el cual será responsable del desarrollo del algoritmo de detección de anomalías y del entrenamiento y mejora de este a medida que se le facilita una mayor cantidad de datos para gestionar un mejor aprendizaje.
- Un administrador de sistemas (ingeniero de software, arquitecto de sistemas, especialista en DevOps), con la finalidad de implentar el nuevo detector de anomalías en la red sobre dispositivos individuales o insertarlos a sistemas en operación y gestionar que este no afectará el correcto y común funcionamiento del sistema o de la red tras su incorporación a entornos de desarrollo o producción.
- Finalmente, un analista de seguridad (analista de seguridad informática, pentester, ingeniero de seguridad) para analizar y evaluar la efectividad y eficiencia del algoritmo al detectar tráfico en la red, y así validar que la detección de tráfico anómalo se realiza de forma correcta.

3. Detalle del caso de uso

a. Detalle funcional

El sistema estará basado en un algoritmo de aprendizaje automático que identifique patrones anómalos en el tráfico total de una red. Este recibirá como entrada todo el tráfico de la red. Estos datos incluirán información sobre el origen y el destino de los paquetes, los puertos utilizados para el envío y recepción de información y el contenido específico de cada dato. El sistema utilizará estos datos de tráfico para entrenar al modelo de aprendizaje automático, el cual será el que se utilice para identificar y discernir entre eventos de tráfico no permitidos y el tráfico convencional.

Primero que todo, en cuanto al conocimiento de negocio del caso de uso se puede destacar como objetivo la *protección de la red corporativa de posibles intrusiones a través de la detección eficaz de tráfico no permitido*. Los beneficios de esta implementación en un entorno de producción bajo un algoritmo estable y eficiente son la reducción ante el riesgo de pérdida de datos, disminución de los niveles de posibles interrupciones de los servicios y una baja en los costes asociados a la recuperación corporativa posterior a la mitigación y erradicación de ataques cibernéticos consolidados.

Por otro lado, la operativa del sistema implementado estará basado principalmente en la gestión de los diferentes tipos de tráfico detectados en la red corporativa, diferenciando el ráfico normal (o tráfico que se espera que circule y se genere de manera normal en una red corporativa, como correos internos y externos, navegación entre servidores internos por usuarios previa y correctamente autorizados, tráfico entrante hacia servicios web puestos sobre internet, gestión de información a través de servidores de gestión de archivos, conexiones VPN externas entrantes,etc), y el tráfico anómalo (o tráfico que no sea común en la red interna, como peticiones desde servicios webs a bases de datos, consultas programadas por servicios webss que permitan la recuperación de información a través de dichas apis, conexiones VPN no autorizadas, acciones de alto nivel realizadas por usuarios de bajos privilegios, correos internos o externos desde fuentes no autorizadas, detección de archivos potencialmente maliciosos que se almacenen en ficheros o directorios de servidores o servicios internos, etc).

En cuanto a los procesos involucrados se incluyen en estos la recopilación constante de información acerca del tráfico generado sobre la red (sea este de cualquier tipo), información antigua acerca de ataques cibernéticos que se hayan concretado sobre la organización y datos sobre planes de respuesta frente a estos que se implementaban entonces. Esto con la finalidad de nutrir el aprendizaje del algoritmo sobre situaciones que en un momento específico se llevaron a cabo y no fueron detectadas a tiempo concretando una intrusión o similar.

Como cumplimiento de normativa se destaca la asignación del estandar ***ISO/IEC 27001***, debido a que el algoritmo constantemente poseerá en circulación información confidencial perteneciente a la empresa, ya sea durante el tiempo de entrenamiento con el tráfico asignado o al estar instaurado en un entorno de producción. Para esto es necesario que el sistema de detección respete las principales secciones de control de la norma:

- Cumplir con las políticas de seguridad de la información internas de la organización
- Seguridad de los recursos humanos propios de la compañía
- Gestionar de la mejor forma los activos entregados. Utilizarlos específica y únicamente para la tarea asignada
- Controles de acceso tanto hacia las funciones del algoritmo y del algoritmo sobre los sistemas internos
- Cifrado y gestión de la información utilizada
- Seguridad física en los dispositivos que gestionen el detector de tráfico anómalo
- Seguridad operacional
- Seguridad en las comunicaciones
- Adquisición, desarrollo y mantenimiento del sistema
- Gestión de incidentes de seguridad de la información

Para gestionar el correcto aprendizaje del algoritmo, se deben entregar al entrenamiento datos verídicos, ya sean de ejemplos generales o propio de la empresa, en el cual se debe incluir información de origen y destino de los paquetes, puertos y protocolos utilizados para la comunicación de distintos servicios externos e internos, tamaño de los distintos tipos de información, frecuencia de transmisión de cierta información, ataques cibernéticos más comunes que ocurran a nivel de internet (como envío de malware, ataques DoS y DDoS, phishing, entre otros muchos), ataques cibernéticos que la cooporación haya sufrido (en caso de sido atacados en ocaciones anteriores), etc.

El mayor detalle funcional del caso de uso debe ser la capacidad de discernir entre los diferentes tipos de tráfico internos y externos, para así diferenciar situaciones reales en donde el tráfico es totalmente seguro y cuando es proveniente de fuentes peligrosas. Tomando en cuenta la alta velocidad del transito de la información, este debe ser capaz de adaptarse a las constantes variaciones del tráfico de datos, en cuanto a cantidad, tamaño, protocolos y tipo. Por su parte, generar o entregar la información necesaria a sistemas que generen alertas en tiempo real para gestionar la respuesta ante tráfico detectado de forma precisa, eficiente y escalable.

La especificación técnica del sistema de detección de tráfico anómalo debe incluir la siguiente información:

- Arquitectura del sistema en el que se instaurará el detector de tráfico.
- El algoritmo de detección de anomalías seleccionado (en primera instancia se plantean las opciones de **IsolationForest**, **One-class SVM** y **Local Outlier Factor**).
- Los criterios con los que se llevará a cabo la detección de anomalías dentro de la red.

Así por su parte, el proceso de desarrollo del sistema debe incluir la recolección y preparación de los datos para el entrenamiento, el entrenamiento como tal del modelo escogido, pruebas realizadas en primera instancia sobre el algoritmo para establecer ciertos criterios de mejora, la implementación del modelo en un entorno de desarrollo para gestionar una evaluación de su comportamiento (en cuanto a su precisión, eficiencia y ecalabilidad) y la instauración de este en una versión mejorada sobre un entorno de producción

A continuación se presentan algunos ejemplos de reglas de detección de anomalías que se pueden utilizar para tráfico no deseado en la red:

- **Regla 1:** El tráfico que se originen o destinen a direcciones IP o puertos no autorizados se considerará anómalo. Los ataques cibernéticos suelen dirigirse a direcciones IP o puertos específicos de un sistema, a través del cual se encuentra en ejecución un servicio específico. Por ejemplo, un ataque de malware podría dirigirse a una dirección IP específica para infectar un dispositivo, desde el cual se genere la transmisión hacia el resto de la red.
- **Regla 2:** El tráfico que sea demasiado grande o demasiado pequeño se considerará anómalo. El tráfico de datos normal suele tener un tamaño y una longitud de paquetes relativamente constantes dependiendo los servicios y protocolos. Por ejemplo, un ataque de denegación de servicio podría generar un gran volumen de tráfico de datos para saturar una red.
- **Regla 3:** El tráfico que se envíe o reciba con una frecuencia inusual se considerará anómalo. Esta regla se basa en la observación de que el tráfico de datos normal suele tener un patrón relativamente constante.
- **Regla 4:** El tráfico que contenga un conjunto de palabras clave específicas se considerará anómalo. En ataques cibernéticos suelen utilizar un conjunto de palabras clave específicas para desencadenar ciertas acciones, como ruptura de sistemas mediante vulnerabilidades web, o infiltración para instaurar backdoors a través de la ruptura de servicios específicos en un sistema expuesto a internet.
- **Regla 5:** El tráfico que tenga un formato incorrecto se considerará anómalo. En ciertos ataques se suelen alterar el formato de los paquetes de datos con la finalidad de

filtrar ficheros maliciosos.

- **Regla 6:** El tráfico que sea inconsistente con el protocolo utilizado se considerará anómalo. Esta regla se basa en la observación de que los ataques cibernéticos suelen utilizar protocolos de manera incorrecta.

b. Identificación de orígenes de datos

Los datos del tráfico utilizados para el entrenamiento y funcionamiento del sistema deben ser obtenidos directamente desde la fuente de activos, la red. Para esto se analizará:

- Registros de tráfico de red, los cuales contienen información sobre todos los paquetes de datos que se han enviado o recibido a través de la red. La información incluye la dirección IP de origen, la dirección IP de destino, el puerto de origen, el puerto de destino, el tamaño del paquete y el protocolo utilizado.
- Sistemas de detección de intrusiones (IDS): Estos sistemas pueden detectar interacciones con la red en base a reglas planteadas, a través de las cuales pueden detectar si una cción realizada sobre la red corresponde acciones normales o maliciosas y genera alertas sobre esto. La información que se podría utilizar incluye el tipo de conexión que se busca realizar, dirección IP de la request e IP del objetivo y data sobre la fecha y hora en que se detecto dicho tráfico.
- Herramientas de análisis de tráfico de red: Estas herramientas pueden proporcionar información adicional sobre el tráfico de red, como el tiempo de respuesta, la pérdida y tamaño de paquetes, protocolos utilizados y el uso de la red durante el tráfico detectado.

En concreto, se considera el realizar el análisis desde las tablas generadas por dispositivos y sistemas, en donde se pueden destacar:

- Tabla de registros de tráfico de red que contengan información sobre todos los paquetes de datos que se han enviado o recibido a través de la red.
- Tabla de alertas de dispositivos IDS/IPS, las cuales contienen información sobre todo el tráfico en primera instancia detectado como anómalo.
- Tabla de análisis de tráfico de red que tengan nformación adicional sobre el tráfico de red, como el tiempo de respuesta, la pérdida de paquetes y la utilización de la red.
- Gestores de registros o logs de sistema, en donde se destaquen interacciones de conectividad realizadas a servicios internos de al red.

Además de esto, se puede utilizar como fuentes de información de datos registros de seguridad pasados de la red, registros de auditorías pasadas, información de terceros como fuentes de datos externas para el entrenamiento, listas negras, entre muchas otras. La combinación de estos orígenes de datos puede proporcionar una visión más completa del tráfico de generado sobre la red corporativa y ayudar a detectar posibles intrusiones no deseadas.

4. Desarrollo del caso de uso

a. Puntos intermedios o seguimiento

En el desarrollo del caso se pueden identificar los siguientes puntos intermedios o seguimiento del proyecto:

- La recolección de datos, en donde se deben recopilar un conjunto de datos de tráfico de red que represente el tráfico normal, en conjunto con tráfico anómalo recopilado de otras fuentes. Este conjunto de datos se utilizará para entrenar el modelo de detección de anomalías.
- Preparación de datos: Los datos recopilados deben prepararse para su análisis. Esto incluye tareas como la limpieza, estandarización y reducción de la dimensionalidad de los datos.
- En el entrenamiento del modelo se debe enseñar un modelo de detección de anomalías utilizando los datos preparados a través de una serie de reglas aplicadas sobre el algoritmo de machine learning.
- El modelo entrenado debe probarse utilizando un conjunto de datos de tráfico de red que no se utilizó para entrenar el modelo. Estas pruebas se utilizarán para evaluar el rendimiento del modelo.
- Finalmente el modelo entrenado debe implementarse en un sistema de detección de tráfico anómalo para analizar una red en tiempo real.

Además de estos puntos intermedios, también se debe considerar lo siguiente:

- Se puede realizar un análisis de los datos recopilados para identificar patrones y tendencias que puedan ser útiles para la detección previa de posibles ataques.
- Se deben desarrollar nuevas reglas de detección con la finalidad de complementar las ya existentes y darle una mayor robustez al funcionamiento del algoritmo.

- El sistema de detección de tráfico anómalo se puede integrar con otros sistemas de seguridad, como IDS o IPS, para así generar una red de protección con una mayor probabilidad de detección y respuesta ante incidentes.

b. Aporte esperado por Big Data

El aporte esperado por Big Data en el caso de uso recientemente explicado abarca la mejora de la precisión y la eficiencia en la detección de ataques cibernéticos, ya que el Big Data puede ayudar a los sistemas de detección a recopilar y analizar grandes cantidades de datos de tráfico de red. Por otra parte, aumenta la capacidad para detectar ataques cibernéticos más sofisticados. Además, una buena implementación del sistema de detección en base al machine learning, aunque con probabilidades de ser costosa en un comienzo, plantea reducir los costes de la detección de ataques cibernéticos gracias a la automatización de muchas de las tareas que actualmente se realizan manualmente en los sistemas de detección de anomalías, evitando posibles perdidas a raíz de intrusiones no deseadas.

En otro aspecto, las limitaciones actuales de la detección de tráfico anómalo incorporan la baja precisión de los sistemas de detección de anomalías ante la identificación específica de patrones y tendencias que sean indicativos de ataques cibernéticos. También, los sistemas de detección pueden generar un gran número de falsas alarmas, dificultando la detección de amenazas reales sobre la red.