Detección de anomalías en tráfico de red de Sistemas de Control Industrial soportada en algoritmos de machine learning

Presentado para obtener el título de

MAGÍSTER EN MATEMÁTICAS APLICADAS Y CIENCIAS DE LA COMPUTACIÓN

Miguel Angel Tristancho Muñoz

Dirección:

Daniel Orlando Díaz López

Universidad del Rosario

Escuela de Ingeniería, Ciencia y Tecnología

DEDICATORIA

Dedico este trabajo a Ruddy Lhay, Juan José y Martín,
porque gracias a su amor y apoyo incondicional fue más fácil
convertir este logro personal en un logro familiar.

AGRADECIMIENTOS

Agradecimientos a Milton Alexander Velasquez, compañero y amigo, quien me brindó el apoyo para que el inicio de esta experiencia tuviese un final exitoso.

Agradecimiento a ADECO y a la Refinería de Barrancabermeja quienes brindaron esta oportunidad para adquirir conocimiento de la Maestría en Matemáticas Aplicadas y Ciencias de la Computación como parte del nuevo esquema de Transformación Digital y SosTecnibilidad de Ecopetrol hacia el 2040.

ABSTRACT

El creciente desarrollo de las redes computacionales asociadas a los sistemas industriales y su integración con las redes corporativas (Internet) han convertido a este grupo en un blanco apetecido por los delincuentes cibernéticos a nivel mundial. Mitigar este tipo de riesgo es una de las mayores prioridades por parte de los integradores, fabricantes y usuarios de los sistemas de control debido al gran impacto que se puede presentar sobre la economía, el ambiente y las personas en una organización cuando se presenta la materialización de un intento de ataque o sabotaje a los procesos industriales. Cada vez resulta más importante que las organizaciones industriales tomen conciencia de la debilidad de estos sistemas y busquen estructuras organizacionales para la gestión de la seguridad que les ayuden a optimizar su protección contra las amenazas externas desde todos los puntos de vista para detectar y abordar los incidentes relacionados con la seguridad antes de que se conviertan en un problema importante.

El dominio de las tecnologías de la información (IT *Information Technology*) está basado en estándares y metodologías para la administración de riesgos que no estan implementadas en su totalidad sobre el dominio de las tecnologías de la operación (OT *Operation Technology*) dado el impacto que puedan ocasionar al ambiente, personas, infraestructura y economía de una empresa una falla del sistema. Teniendo en cuenta que los Sistemas de Control Industrial (ICS) en el dominio OT son el corazón y el alma de la infraestructura crítica, es importante monitorear las tecnologías que se encuentran interconectadas (servidores, routers, switches, etc) en su funcionamiento, así que para el presente trabajo se analizan entre los diferentes vectores de ataques, el caracterizar el tráfico de la red de control.

Los modelos de aprendizaje automático son herramientas matemáticas que nos permiten representar eventos externos, con el fin de obtener una mejor comprensión y predecir el comportamiento futuro de una o más variables de un problema. Mediante el presente trabajo se desarrollará un modelo de aprendizaje automático de confianza en la caracterización del tráfico de las redes industriales, basado en entrenamiento y evaluación de algoritmos del aprendizaje automático para detectar condiciones "*normales*" y "*no normales*" que representan detección de anomalías en el tráfico y así estar preparados ante comportamientos que puedan causar indisponibilidad sobre los ICS.

The growing development of computer networks associated with industrial systems and their integration with corporate networks (Internet) have made this group a desired target for cybercriminals worldwide. Mitigating this type of risk is one of the highest priorities for integrators, manufacturers, and users of control systems due to the great impact that can occur on the economy, the environment and the people in an organization when materialization occurs. of an attempted attack or sabotage of industrial processes. It is becoming increasingly important for industrial organizations to become aware of the weakness of these systems and seek organizational structures for security management that help them optimize their protection against external threats from all points of view to detect and address incidents. security-related issues before they become a major problem.

The domain of information technology (IT Information Technology) is based on standards and methodologies for risk management that are not fully implemented on the domain of operation technology (OT Operation Technology) given the impact they can Cause a system failure to the environment, people, infrastructure and economy of a company. Considering that Industrial Control Systems (ICS) in the OT domain are the heart and soul of critical infrastructure, it is important to monitor the technologies that are interconnected (servers, routers, switches, etc.) in their operation, so for the present work, the characterization of the traffic of the control network is analyzed between the different vectors of attacks.

Machine learning models are mathematical tools that allow us to represent external events, to gain a better understanding and predict the future behavior of one or more variables in a problem. Through this work, a reliable machine learning model will be developed in the characterization of industrial network traffic, based on training and evaluation of machine learning algorithms to detect "normal" and "non-normal" conditions that represent detection of anomalies in traffic and thus be prepared for behaviors that may cause unavailability on the ICS.

TABLA DE CONTENIDO

Contenido

LISTA DE FIGURAS

LISTA DE TABLAS

Capítulo 1
INTRODUCCIÓN

En el pasado, las redes de control industrial estaban completamente aisladas de ataques externos con el uso de protocolos de control propietarios que se ejecutaban en hardware y software especializados. Con el continuo desarrollo de la automatización industrial, los sistemas basados en TCP/IP se han abierto camino en el sistema de control industrial. Las redes de control basadas en TCP/IP brindan una mejor conectividad y capacidades de acceso remoto, lo que hace que el sistema de control industrial sea vulnerable a amenazas cibernéticas como virus informáticos, gusanos de Internet, manipulación de datos operativos que pueden provocar la interrupción de todo el sistema de control.

Los ICS juegan un papel esencial en muchos aspectos, especialmente en campos industriales e infraestructura crítica en todo el mundo. Con el auge de esta tecnología, se puede facilitar a los operadores la gestión, seguimiento y control de procesos industriales. Sin embargo, una falla de ICS por ataques cibernéticos puede causar impactos negativos en servicios críticos y personas. Los ataques a los sistemas de control industrial pueden no solo causar pérdidas monetarias, sino también dañar los equipos, el medio ambiente y lastimar al personal. Para atacar el ICS aislado, los atacantes tendrían que obtener acceso físico a la red del ICS, conectar un nuevo nodo a la red del ICS o penetrar y usar el hardware del ICS para transmitir comandos maliciosos. Este nivel de seguridad permitió a los diseñadores de protocolos ICS concentrarse en la seguridad y disponibilidad de operación en sistemas físicos, disminuyendo la necesidad de implementaciones de seguridad cibernética. De este modo, los protocolos utilizados por los ICS casi no tienen características de seguridad y son vulnerables a varios ataques. Sin embargo, conectar redes ICS a Internet beneficia a las empresas ya los ingenieros que las utilizan, y con el uso generalizado de Internet en el mundo, conectar redes ICS a otras redes se ha convertido en una tendencia inevitable.

Los sistemas de control industrial ICS se han conectado globalmente a las redes informáticas abiertas para fines de gestión y control descentralizados. La mayoría de estos sistemas de control en red que no están diseñados con protección de seguridad pueden ser vulnerables a los ataques de red en la actualidad, por lo que existe una demanda creciente de sistemas de detección

de intrusos (IDS) eficientes y escalables en la infraestructura de red de las plantas industriales [8]. Con el rápido aumento de la conectividad de los sistemas de control a las redes abiertas para la gestión descentralizada y el control remoto, la mayoría de estos sistemas de control en red que no están diseñados con restricciones de seguridad pueden ser vulnerables a los ataques de red en la actualidad. Si bien estos sistemas ciberfísicos permiten nuevos caminos hacia una mayor eficiencia y disponibilidad de las unidades de procesos industriales, las oportunidades de uso indebido y abuso aumentan debido a las interconexiones entre los dispositivos que afectan el mundo real y las redes de comunicación convencionales. En consecuencia, las técnicas de análisis de intrusiones y vulnerabilidades que son relativamente comunes en las redes de tecnología de la información se están desarrollando para ICS. Inicialmente, los ICS no se asemejaban a los sistemas de TI en el sentido de que éstos eran sistemas aislados que ejecutaban protocolos de control patentados por fabricantes utilizando hardware y software especializados.

A medida que los equipos TO fueron mejorando su rendimiento computacional, empezaron a adoptar soluciones propias de TI para promover características de esta tecnología, lo que permitía entre otras posibilidades la conectividad corporativa y las capacidades de acceso remoto para el soporte técnico especializado. Los dispositivos Ethernet y de Protocolo de Internet (IP) disponibles en todas las comunicaciones de hoy día y de bajo costo, ahora están reemplazando las tecnologías patentadas más antiguas, lo que aumenta la posibilidad de explotación de vulnerabilidades e incidentes de ciberseguridad. Todo esto permite que el conocimiento a un mayor nivel de detalle sobre los protocolos implicados en procesos industriales sea un importante hito en el entendimiento de los posibles puntos débiles, vectores de ataque y posibles medidas de defensa deban ser barajadas a la hora de implementar o fortificar un sistema de control industrial [4].

Este nuevo proceso de integración admite nuevas capacidades y oportunidades en los dispositivos de TI, pero proporciona un aislamiento significativamente menor para ICS del mundo exterior que los sistemas predecesores, lo que crea una mayor necesidad de proteger estos sistemas. Si bien se han diseñado soluciones de seguridad para tratar estos problemas de seguridad en los sistemas de TI típicos, se deben tomar precauciones especiales al introducir estas mismas soluciones en entornos ICS. En algunos casos, se necesitan nuevas soluciones de seguridad que se adapten al entorno de ICS.

Ninguna red de control es absolutamente segura. Por lo tanto, cualquier arquitectura de sistemas ICS debe ilustrar una serie de controles ante ataques que probablemente infrinjan la postura defensiva que se está considerando. En cualquier conjunto de ataques no derrotados de manera confiable, siempre hay un ataque o conjunto de ataques menos sofisticados o más simples con consecuencias graves. Cuando se diseña una arquitectura de red, es importante que se definan límites de administración evidentes entre los diferentes segmentos conectados. Permitir la separación de las redes en distintas funciones y objetivos es permite implementar medidas de seguridad y evitar flujos de información innecesaria. Siguiendo esta recomendación, el segmento de red de ICS debe separarse del segmento de red corporativa, puesto que la naturaleza del tráfico de las distintas áreas está perfectamente diferenciada. [5]

Este trabajo explora la viabilidad de los métodos de aprendizaje automático para detectar los nuevos escenarios de amenazas ante un tráfico anómalo. De manera similar a los sistemas de detección de intrusos en la red en el dominio de la seguridad cibernética, las comunicaciones de comando y control en una configuración de infraestructura crítica se monitorean y examinan contra ejemplos de tráfico de comando benigno y malicioso, para identificar posibles eventos de ataque. Evaluamos un conjunto de algoritmos de aprendizaje automático en términos de su capacidad para identificar tráfico anómalo al analizar las comunicaciones de una red de control industrial real en términos de tráfico benigno.

Mediante el presente; exploramos la interacción de los algoritmos de aprendizaje automático como medio para discriminar las perturbaciones que pueden llegar a afectar los sistemas industriales generando una no disponibilidad de la plataforma de control [9].

Capítulo 2
OBJETIVOS

2.1 Objetivo general

Establecer un sistema de análisis de tráfico de red basado en algoritmos de machine learning (ML), orientado a sistemas de control industrial que permita: la identificación de comportamientos anormales para evitar la explotación de vulnerabilidades que afecten la seguridad de procesos industriales reduciendo riesgos de disponibilidad y soporte la continuidad del negocio.

2.2 Objetivos específicos

- Caracterizar el tráfico de red existente en una arquitectura ICS para construir un dataset que permita identificar tráfico normal y anormal en una arquitectura conectada a la intranet corporativa
- Construir un modelo de detección de anomalías que permita un análisis de comportamientos en el tráfico de un ICS mediante algoritmos de ML.
- Evaluar el modelo en un escenario controlado similar al tomado en el dataset del proyecto mediante el cual se despliega en un ICS un ataque propio de las redes de control.

Capítulo 3
PROBLEMA Y JUSTIFICACIÓN

En los últimos años, los ICS tradicionales se ha integrado gradualmente con las denominadas tecnología de la información TI, que, al integrar el desarrollo de la información y las características compartidas de Internet, forman una red de control industrial, y es una tendencia dominante que el proceso de producción industrial se convierte en inteligente y en red. Desde una perspectiva de seguridad, definitivamente tendría sentido mantener los ICS de IT lo más separados posible. Sin embargo, desde una perspectiva comercial, tener datos precisos, sobre la marcha y relevantes provenientes del entorno OT tiene mucho sentido. Dicha información permite una programación de producción más estricta, puede disminuir la cantidad de inventario que debe mantenerse en el sitio, ayuda al cálculo de costos y brinda muchas más ventajas logísticas. Los sistemas ERP y MES modernos se basan en entradas e información tanto de la producción como del lado empresarial de una empresa. Esas razones, y muchas más, han impulsado la convergencia de los sistemas de TI y OT [22].

El ICS fue originalmente construido con, y alrededor de, dispositivos, equipos y medios y protocolos de red patentados, sin tener en cuenta la seguridad. Casi todos los proveedores tenían su propia forma de hacer las cosas, y cada uno de ellos tenía una forma diferente de configurar, operar y mantener su instalación. Este comportamiento patentado no funcionó bien con toda la demanda de convergencia de TI/TO y, poco a poco, los proveedores de equipos ICS comenzaron a adherirse a un conjunto común de estándares, a saber, los protocolos de red ampliamente utilizados Ethernet, Protocolo de Internet (IP), Protocolo de control de transporte (TCP) y Protocolo de datagramas de usuario (UDP) para ejecutar sus controles de manera que el uso de los protocolos de automatización inició un retroceso. Para no tener que reinventar o adaptar los protocolos industriales, muchos proveedores colocaron sus bien establecidos controles y protocolos de automatización sobre el protocolo TCP o UDP. De esta manera, todo lo que era necesario para subirse al tren de convergencia de TI/TO era colocar un módulo de comunicaciones compatible con IP/TCP/UDP, y estaban listos para funcionar.

Ahora, los clientes podían adaptarse más fácilmente a un estándar que permitía un conjunto común de tecnologías para cablear toda la instalación de producción con el mismo tipo de cables,

interruptores e incluso el mismo conjunto de habilidades de la persona que realizaba la instalación, sin embargo, esta nueva integración fácil para obtener datos, solucionar problemas y obtener accesibilidad, también abrió estos dispositivos y equipos previamente ocultos, a una red que en su contexto de ciberseguridad presenta el tráfico en texto claro, lo que origina una gran brecha de seguridad pues la información en esta implementación permitiría en un caso particular de ataque, un acceso no autorizado a la red lo que ocasionaría que un atacante inspeccionara y manipulara el tráfico y controlara la infraestructura crítica. Es importante entender, que para los ICS el protocolo Ethernet/IP es susceptible de verse afectado por todas las posibles vulnerabilidades del internet, como puede ser la suplantación de identidad o la captura de tráfico. Además, como utiliza UDP para sus mensajes y este carece de control de la transmisión, es posible la inyección de tráfico malicioso, por ello que las reglas deberán tener en cuenta este tipo de elementos de juicio a la hora de presentar el esquema "normal" y "no normal".

3.1 Los primeros ataques a la infraestructura crítica a nivel mundial

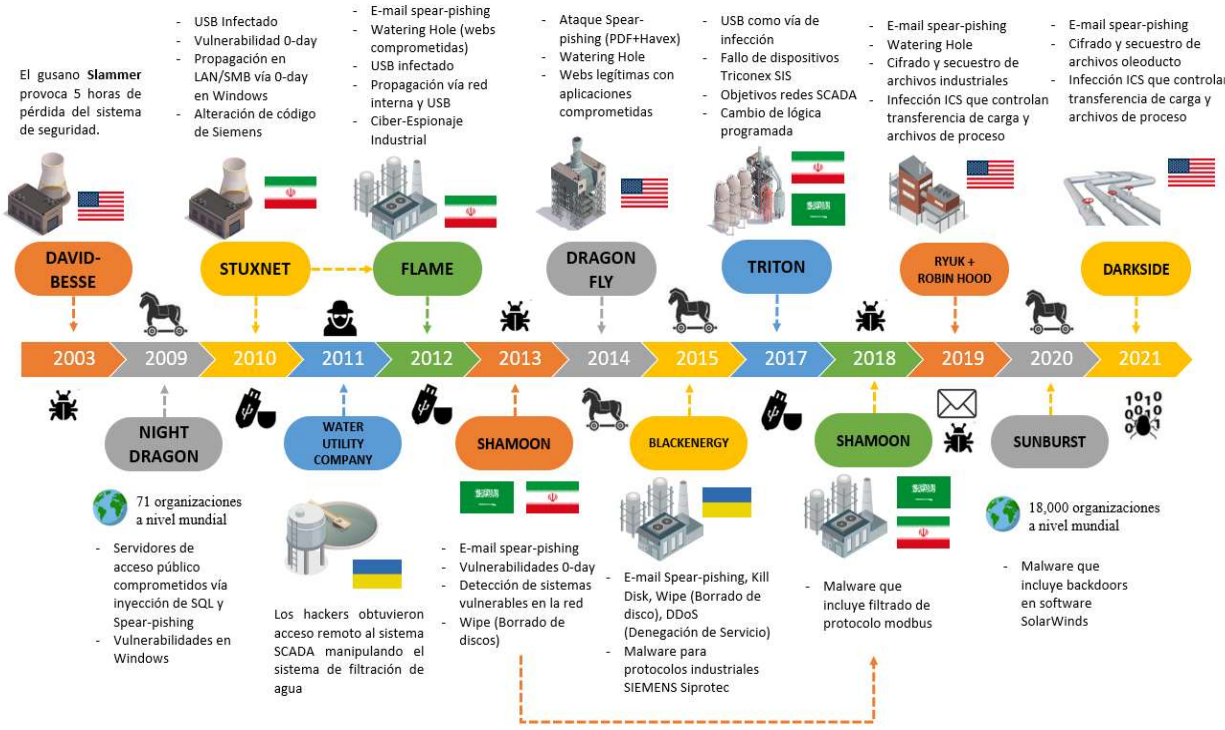


Figura 1. Cronograma de los principales ataques a ICS. Autor

Una vez que ocurren problemas de seguridad de la información sobre un ICS, no solo se disminuye el rendimiento del sistema y la pérdida de funciones, sino que también pueden provocar víctimas, contaminación ambiental y otros accidentes importantes, incluso poniendo en peligro la seguridad nacional. A continuación, los principales ataques a ICS a nivel mundial que afectaron la disponibilidad, integridad y confiabilidad de la infraestructura críticas [12]:

AÑO	LUGAR	TIPO DE ATAQUE	DETALLE
2003	Ohio, EEUU Central Nuclear	Slammer Worm [1]	Generó una obstrucción de los servidores de Microsoft. El gusano primero se incrustó en la computadora de un contratista de David-Besse y encontró su camino hacia los sistemas de control de procesamiento del reactor nuclear porque el sistema de control de procesamiento estaba conectado a la red corporativa pública.
2009	71 organizaciones a nivel mundial	Night Dragon Troyano[2]	Comprometieron servidores web públicos con inyección SQL e instalaron malware y RAT (Remote Access Trojan). Utilizaron los servidores web infectados para atacar objetivos internos, robo de información y ataques a ICS.
2010	Irán, Central Nuclear	STUXNET [15]	Los inspectores de la Agencia Internacional de Energía Atómica que visitaban una planta nuclear en Natanz, Irán, notaron que las centrifugadoras usadas para enriquecer uranio estaban fallando, se afectaron en total más de 1000 centrifugadoras.
2012	Irán	Flame[3] (STUXNET, DUQU)	Está diseñado principalmente para espiar a los usuarios de computadoras infectadas y robarles datos, incluidos documentos, conversaciones grabadas y pulsaciones de teclas.
2013	Arabia Saudita y los Emiratos Árabes Unidos	Shamoon[4]	A diferencia de los ataques anteriores de Flame + STUXNET, estos últimos ataques involucran una segunda pieza nueva de malware de limpieza (Trojan.Filerase). Este malware eliminará y sobrescribirá archivos en la computadora infectada.
2014	Estados Unidos y Europa.	Dragon Fly	El malware Dragonfly infectó cientos de computadoras comerciales en un intento a menudo exitoso de recopilar información sobre los sistemas de control industrial.
2015	Ucrania	BlackEnergy3[5] Troyano	Troyano que se utiliza para realizar ataques DDoS, ciberespionaje y ataques de destrucción de información. Comenzó a implementar complementos relacionados con SCADA para las víctimas en los ICS y los mercados energéticos de todo el mundo.
2017	Medio Oriente	Triton[6] TRISIS o Hatman	Diseñado para atacar los sistemas de seguridad industrial de empresas petroquímicas. En particular, Triton aprovecha las vulnerabilidades del sistema instrumentado de seguridad Triconex de Schneider.

[1] http://large.stanford.edu/courses/2015/ph241/holloway2/

[2] https://www.industrialcybersecuritypulse.com/threats-vulnerabilities/throwback-attack-night-dragon-one-of-the-first-attacks-to-target-the-energy-industry/

[3] https://www.wired.com/2012/05/flame/

[4] https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/shamoon-destructive-threat-re-emerges-new-sting-its-tail

[5] https://www.kaspersky.com/resource-center/threats/blackenergy

[6] https://blogs.cisco.com/security/how-does-triton-attack-triconex-industrial-safety-systems

2018	Arabia Saudita, y los Emiratos Árabes Unidos	Shamoon*[7]	Saipem, una empresa italiana de exploración de petróleo y gas sufrió un ataque cibernético que destruyó el 10 % de sus datos de mainframe.
2019	Maryland + Florida, EEUU	Robin Hood + RYUK[8] Ramsomware	Los servidores de todos los servicios de la ciudad se desactivaron y fueron cifrados por el ramsomware, con la excepción de los departamentos de policía y bomberos, que estaban en un servidor separado. El sistema telefónico de la administración de la ciudad también estaba caído.
2021	Colonial PipeLine, EEUU	DarkSide[9]	El ataque comenzó cuando un grupo de piratas informáticos identificado como DarkSide accedió a la red Colonial Pipeline. Los atacantes robaron 100 gigabytes de datos en un lapso de dos horas.

Tabla 1. Principales ataques a los ICS a nivel mundial. Autor

Los fabricantes de ICS utilizan la tecnología usada en el dominio IT para aumentar la conectividad y las capacidades de acceso remoto, minimizar los gastos generales operativos y también lograr la utilización óptima de los recursos y la globalización del mercado. Por lo tanto, los ICS se han vuelto más similares a los sistemas de Tecnología de la Información (TI) [22].
Según el informe de Kaspersky en la segunda mitad del 2022 [28], el porcentaje de computadoras ICS asociadas al negocio del Oil&Gas en las que se bloquearon objetos maliciosos fue mayor respecto al promedio mundial (39.6% vs 31.8%), esto quiere decir que el nivel de exposición y vulnerabilidad va en aumento, probablemente en el contexto de integraciones entre los dominios de las tecnologías IT y OT. Figura 2.

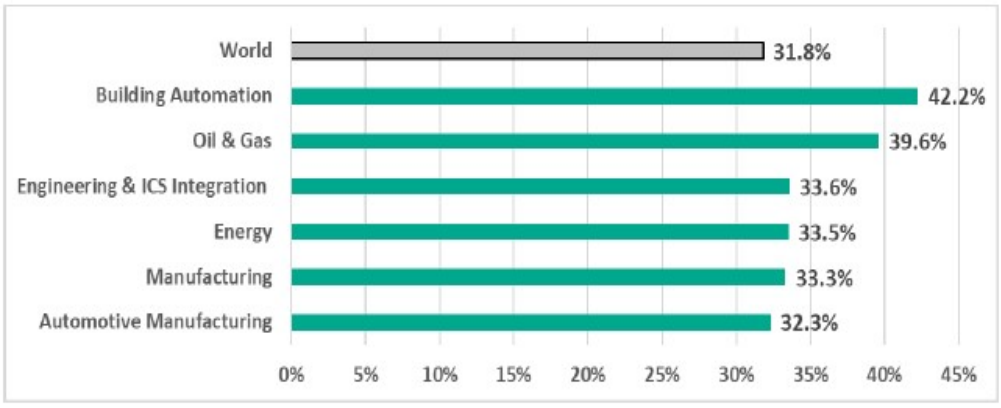


Figura 2. Porcentaje de equipos ICS en los cuales se bloquearon objetos maliciosos en el 2022 por tipo de Industria. Fuente [28]

[7] https://www.cybersecurity-insiders.com/shamoon-malware-behind-saipem-cyber-attack/
[8] https://ics-cert.kaspersky.com/publications/reports/2020/04/24/threat-landscape-for-industrial-automation-systems-ransomware-and-other-malware-key-events-of-h2-2019/
[9] https://www.techtarget.com/whatis/feature/Colonial-Pipeline-hack-explained-Everything-you-need-to-know

Adicionalmente, se evidencia que aproximadamente el 41,6% de las computadoras ICS en el sector del Oil&Gas fueron infectadas por malwares en el proceso de conectividad a Internet, probablemente, estos resultados se encuentren orientados a la carencia en los esquemas de defensa en profundidad o estrategias de ciberseguridad corporativa. Figura 3.

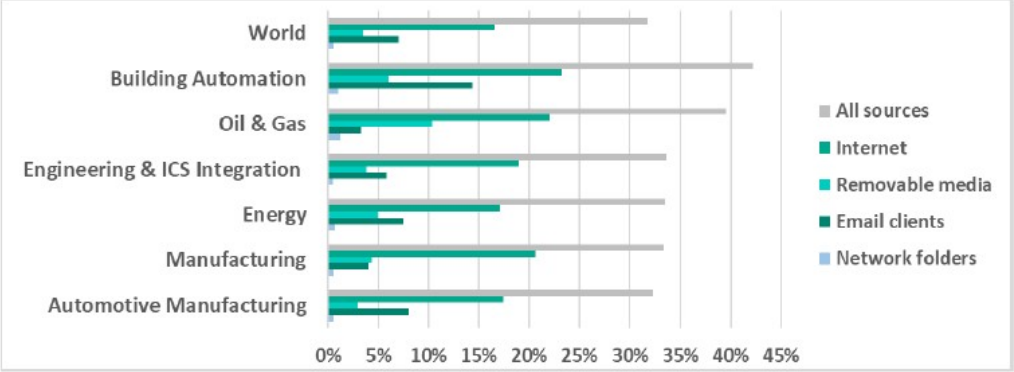


Figura 3. Porcentaje de equipos ICS que bloquearon malware según fuente de amenaza en el 2022 por tipo de Industria. Fuente [28]

Según como se puede observar en los resultados del estudio, se cree que un porcentaje tan alto (10.4%) de computadoras en las que se bloquea el malware cuando se conectan medios extraíbles en el sector del Oil&Gas, puede deberse principalmente a la forma en que opera esta industria. Una gran cantidad de empresas distribuidas geográficamente, a menudo con canales de comunicación deficientes que conectan sistemas y sitios remotos, significa que los medios extraíbles deben usarse más ampliamente cuando se realizan tareas de mantenimiento, en comparación con otras industrias. Figura 4.

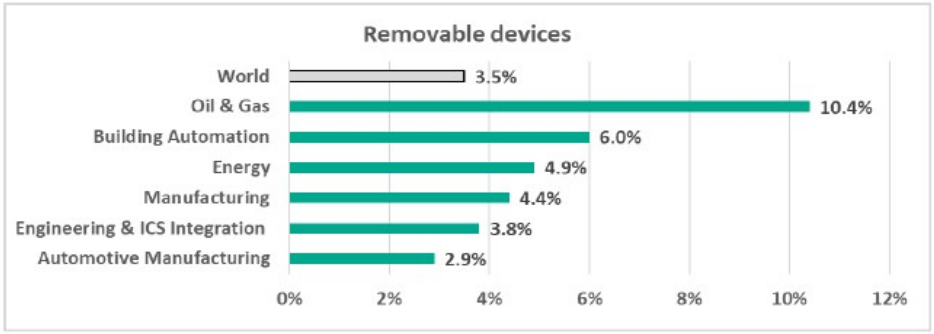


Figura 4. Porcentaje de equipos ICS que bloquearon malware provenientes de dispositivos removibles en el 2022 por tipo de Industria. Fuente [28]

Como era de esperar, dado que el sector Oil&Gas lidera las categorías de medios extraíbles y carpetas de red, lidera de igual manera en el segmento de los archivos o carpetas infectados y bloqueados que se encuentran red, ya que los virus y gusanos se propagan principalmente a través de redes, a través de medios extraíbles y carpetas de red. Figura 5 y 6.

Por lo tanto, existe un requerimiento importante para asegurar estos sistemas. Se necesitan estrategias, incluidas las relacionadas con las personas, los procesos y las tecnologías, para mejorar la seguridad de ICS de manera integral. Con la integración cada vez más profunda de la informática y la industrialización, los ICS se enfrenta a amenazas de seguridad cada vez más graves al mismo tiempo que se desarrolla rápidamente.

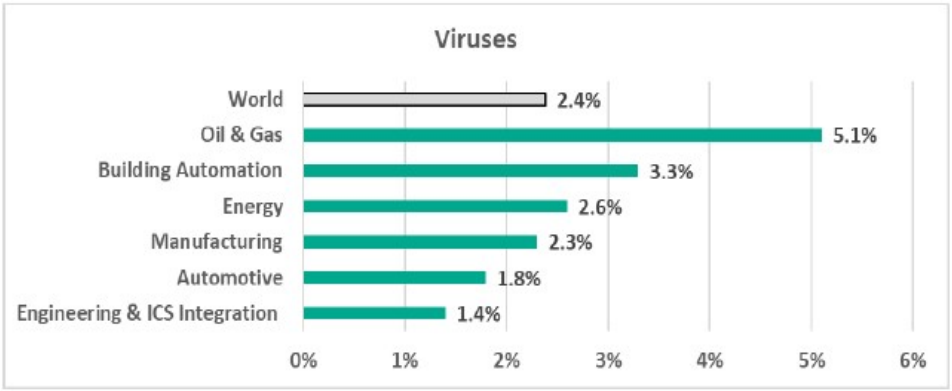


Figura 5. Porcentaje de equipos ICS que bloquearon Virus por tipo de Industria. Fuente [28]

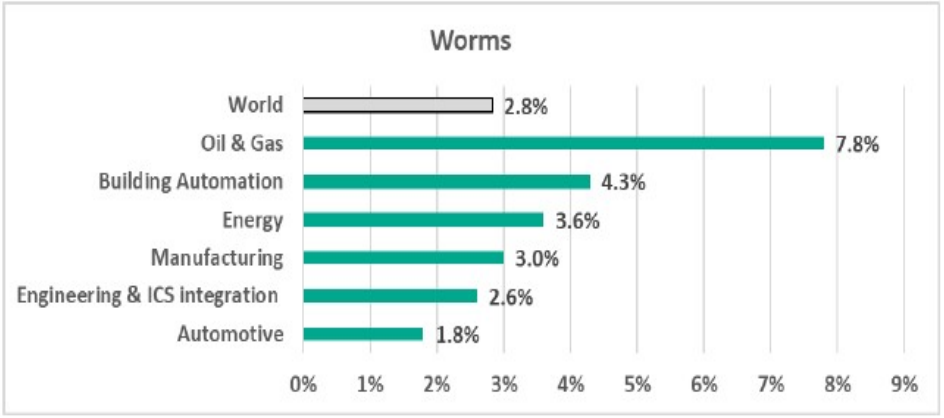


Figura 6. Porcentaje de equipos ICS que bloquearon worms por tipo de Industria. Fuente [28]

La convergencia entre los dominios IT y OT conlleva a que se requieran de forma proactiva, esquemas operativos de colaboración entre los centros de operaciones de seguridad (SOC) y profesionales en ciberseguridad industrial. Es la integración de estos dos grupos, los que reúnen a personas, procesos y tecnología no solo para responder a las amenazas de seguridad, sino también para mitigarlas. Sin embargo, aún más importante que la tecnología adecuada para gestionar la ciberseguridad en entornos industriales; es contar con el conocimiento y personal idóneo en las áreas de proceso, pues son ellos quienes conocen el verdadero riesgo de sus operaciones, es decir; contar con personas familiarizadas con los entornos de OT, sus redes, sus tecnologías y sus prioridades empresariales, aunado a los conocimientos necesarios en torno a las normas de seguridad industrial y a las mejores prácticas en instalaciones industriales.

De lo observado en el inicio de este capítulo, para el presente trabajo, se plantea uno de los mecanismos de monitoreo pasivo (es decir sin afectar a la red o el tráfico existente) sobre los mecanismos que puedan afectan la disponibilidad de los ICS. Con el monitoreo de seguridad pasivo, no se introduce tráfico adicional a una red de control ni se utilizan recursos adicionales de puntos finales conectados a la red para recopilar los datos que nos interesan en el análisis, esto se centra en capturar paquetes de red y examinar los paquetes capturados en busca de contenido o comportamiento malicioso extraño o conocido. El monitoreo de seguridad pasivo es el método preferido para el monitoreo de seguridad de la red ICS, ya que no ejerce un cambio sobre los controladores industriales y los equipos de automatización. Además, el monitoreo pasivo de la red no requiere la instalación de software adicional en los dispositivos finales, lo que a menudo no es posible ni factible con el equipo ICS. La clasificación del tráfico de red industrial juega un papel importante en la *gestión de la seguridad,* el análisis a puertos de comunicación, carga útil y estadísticas de tráfico hacen parte de la estrategia de análisis del presente trabajo. El método basado en puertos clasifica el tráfico asumiendo el uso constante de números de puerto TCP o UDP; sin embargo, la existencia de puertos privados y dinámicos hace que estos métodos se vuelvan poco confiables e inexactos por ello que es importante el conocimiento de las redes de comunicación y protocolos industriales [4].

Se propone en el presente trabajo la aplicación de algoritmos de ML de clasificación binaria para la construcción de un modelo de detección de anomalías en sistemas ciberfísicos industriales en el cual se reflejen condiciones ***normales*** y ***no normales*** de las comunicaciones entre los dispositivos interconectados a dicha red. Mediante la implementación de estadísticas, se determinarán los parámetros, aplicabilidad y eficiencia de los modelos propuestos para revelar las anomalías en los sistemas ICS y se confirmará la capacidad del modelo para revelar y distinguir las características propias de los ataques sobre los activos industriales de la infraestructura crítica de una compañía [3].

El presente trabajo está compuesto por cuatro secciones. La primera sección del documento proporciona una introducción y antecedentes sobre las tendencias de ciberseguridad en los ICS. La segunda sección proporciona diferentes métodos para caracterizar el tráfico de red de control industrial. La tercera sección proporciona algoritmos de ML para construir un modelo de detección de anomalías y la evaluación del modelo en un escenario controlado similar al tomado en el dataset del proyecto. La última sección es el resultado y las discusiones. Finalmente, nuestro trabajo concluye y también proporciona direcciones futuras para el uso del aprendizaje automático en el entorno de redes de control industrial.

Capítulo 4
MARCO TEÓRICO Y ESTADO DEL ARTE

4.1 MARCO TEÓRICO

4.1.1 Sistemas de Control Industrial ICS

Los Sistemas de Control Industrial (ICS) juegan un papel clave en la infraestructura crítica de la nación, entre ellos se pueden encontrar sistemas de fabricación, energía, redes eléctricas, plantas de tratamiento de agua, refinerías de gas y petróleo, y atención de la salud [13].

Actualmente pertenecen al dominio de las tecnologías de la operación TO, y se encuentran compuestos por dispositivos electrónicos programables, procesos, equipos de campo, switches, entre otros. En la medida en que esos elementos puedan compartir la información de forma automática, la eficiencia será mayor, ya que se disminuirán los tiempos de respuesta y se eliminarán o reducirán errores. [8]

La Norma ISA-95 es un estándar internacional que facilita la integración de todos los sistemas de información que puedan estar involucrados en un entorno industrial, desde las funciones empresariales hasta los sistemas de control en planta. La ISA-95 define 5 niveles de operaciones en la automatización industrial (Figura 7):

- **Nivel 0**: el propio proceso productivo.
- **Nivel 1**: los propios dispositivos que procesan y manipulan el producto en sí (robots, actuadores, instrumentación). Los DCS se ubican en este nivel, ya que combinan tecnologías de control (los propios controladores) con el software de supervisión ligado a dichos controladores de proceso.
- **Nivel 2**: los dispositivos que monitorizan y controlan el proceso productivo (HMI, SCADAs).
- **Nivel 3**: los dispositivos que controlan el work flow y las recetas del proceso productivo y que almacenan toda la información sobre el mismo (MES, Batch, Historian, LIMS)
- **Nivel 4**: el nivel que contiene la infraestructura de logística, inventario, ERP 8 o planificación.

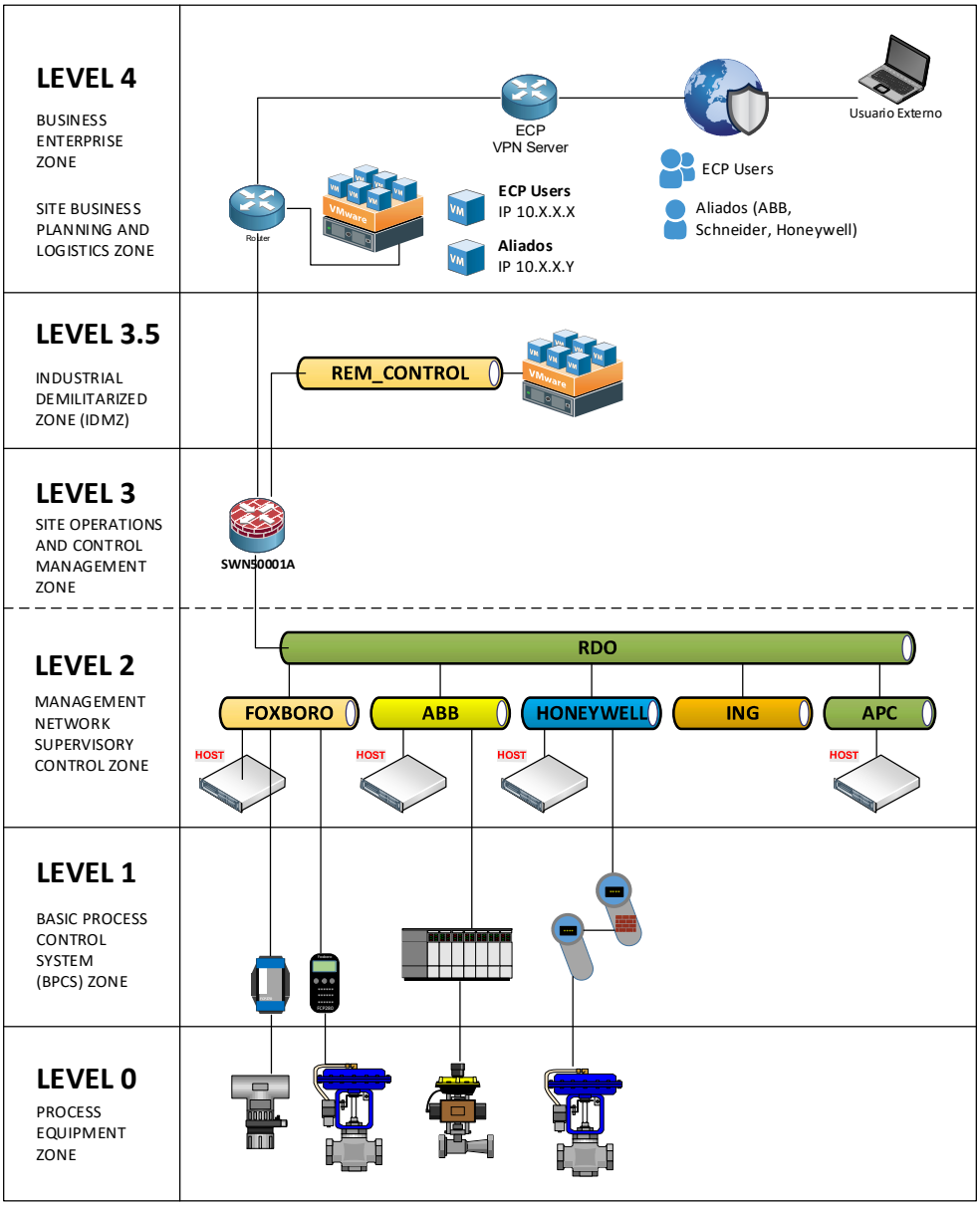


Figura 7. Arquitectura típica de un sistema de control. Autor

Si bien estos sistemas ciberfísicos permiten nuevos caminos hacia una mayor eficiencia, las oportunidades de uso indebido y abuso aumentan por las interconexiones entre los dispositivos mecánicos que afectan el mundo real y las redes de comunicación convencionales.

Históricamente, las redes ICS y sus componentes estaban protegidos contra ataques cibernéticos, ya que se trabajaba con hardware y software propietarios y estaban conectados a

redes aisladas sin conexión a Internet. Sin embargo, a medida que el mundo se interconectó, ha habido una necesidad de conectar los componentes del ICS a otras redes, para permitir el acceso remoto y funcionalidades de monitoreo. Como resultado, los ICS exponen sus protocolos especiales de conectividad entre dispositivos a los riesgos asociados en la afectación de conexiones de equipos no autorizados a la red ICS o inserción de tráfico no autorizado en la red del ICS.

4.1.2 Seguridad en los ICS

Considerando un ejemplo práctico de sistemas ICS en red en una empresa en la que los ICS se encuentran interconectados a la red corporativa como se muestra en la Figura 7, varios ejemplos críticos de amenazas a la seguridad de una amplia variedad de fuentes pueden clasificarse en dos categorías principales [8]:

1. **Amenazas externas**: **(a)** Conexiones remotas no autorizadas a la puerta de enlace o servidores desde Internet, **(b)** ataques de Denegación de Servicio (DoS) entrantes de piratas informáticos o usuarios remotos a aplicaciones, servidores de control y RTU/PLC de vital importancia. Conducen a la interrupción del servicio, **(c)** Ataques de sondeo entrantes de intrusos cerca de las estaciones inalámbricas para recopilar información, **(d)** Ataques directos a las vulnerabilidades del sistema operativo ICS para comprometer el servidor.
2. **Amenazas internas**: **(a)** Desbordamiento de búfer o ataques de fallas en el programa de personas internas descontentas a los servidores de control. Estos ataques ayudan a obtener privilegios de root y se clasifican como ataques User-to-Root (U2R), **(b)** Escaneo de puertos de información privilegiada a redes internas. Se puede utilizar para el descubrimiento de vulnerabilidades, **(c)** Varios ataques de penetración desde la red troncal de igual a igual o redes de proveedores a redes internas.

En respuesta a estas amenazas de seguridad sobre los ICS, la detección y prevención de intrusiones hace parte de estrategias técnicas adoptadas por sectores industriales y se están convirtiendo en tácticas para proteger la infraestructura crítica cuando, según MITRE[10]:

[10] https://attack.mitre.org/tactics/ics/

a) TA0108 El adversario está tratando de ingresar a su entorno ICS.
b) TA0104 El adversario intenta ejecutar código o manipular funciones, parámetros y datos del sistema de forma no autorizada.
c) TA0100 El adversario está tratando de recopilar datos de interés y conocimiento del dominio en su entorno ICS para informar su objetivo.

Para efectos del desarrollo del presente trabajo se usarán elementos de detección de intrusos para bloquear el tráfico en los límites de la red teniendo en cuenta el monitoreo y detección de tráfico de red sospechoso de manera que no se interrumpan protocolos o comunicaciones responsables de las funciones en tiempo real relacionadas con el control o la seguridad[11].

4.1.3 Protocolos de comunicación en los ICS

Comprender cómo funcionan las redes industriales requiere una comprensión básica de los protocolos de comunicaciones subyacentes que se utilizan, dónde se utilizan y por qué. En la figura 8 se presentan algunos de los protocolos industriales de mayor frecuencia de uso, según las capas del modelo ISA99, En ellas se pueden identificar los protocolos de instrumentación industrial que se conectan a las variables de proceso (Level 0), los protocolos que llevan la información a los buses de control (Level 1) en el cual se recogen los estados de la instrumentación de campo y se procesan para el control de los procesos, y finalmente los protocolos de interconexión con las redes de comunicaciones estándares (Level 2) en donde se procesan e interactúan con las tecnologías de la información.

Hay muchos protocolos altamente especializados que se utilizan para la automatización y el control industrial, la mayoría de los cuales están diseñados para la eficiencia y la fiabilidad para apoyar la economía y los requisitos operativos de las grandes arquitecturas de sistemas de control industrial (ICS).

[11] https://attack.mitre.org/mitigations/M0931/

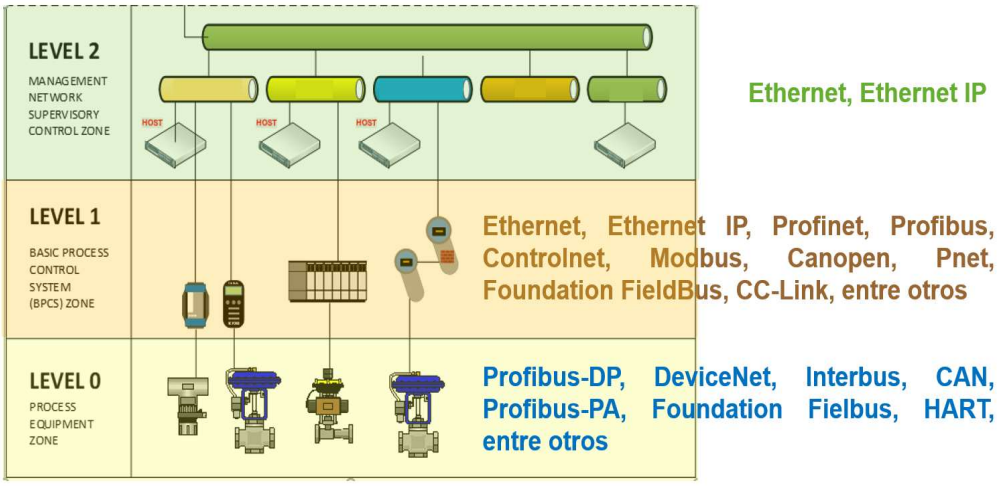


Figura 8. Protocolos Industriales según el Modelo ISA 99. Autor

Los protocolos industriales están diseñados para operar en tiempo real para respaldar operaciones de precisión implicando la comunicación determinista de los datos de seguimiento y control. Esto significa que la mayoría de los protocolos industriales renuncian a cualquier característica o función que no sea absolutamente necesario en aras de la eficiencia. Más desafortunado es que esto a menudo incluye la ausencia incluso de funciones de seguridad básicas, como la autenticación o el cifrado, ambos requieren gastos generales adicionales.

Para complicar aún más las cosas, muchos de estos protocolos se han modificado para ejecutarse a través de Ethernet y Protocolo de Internet (IP) redes a medida que los proveedores se alejaron de las redes propietarias y el hardware de red y tecnologías comerciales listas para usar (COTS) aprovechadas. Esto, sin embargo, ahora ha dejado estos protocolos "frágiles" potencialmente vulnerables a los ataques cibernéticos [9]. La migración a TCP/IP también implicó la estandarización e implementación de nuevos protocolos sobre los ICS capaces de entender conexiones TCP/IP. Actualmente, existen varios protocolos ICS basados en IP, como Modbus/TCP, DNP3, IEC-104 e ICCP/TASE2 [21].

A continuación, se analizan algunos de los más importantes protocolos de comunicación industrial[12] usados en la conectividad para el control de los procesos industriales [22]:

Protocolo	DETALLE
PROFIBUS DP	**Velocidad**: Dependiente de la velocidad en baudios del enlace serial; **Topología física**: uno-a-uno o autobús; **Topología lógica**: centralizado maestro / esclavo]; **Número máximo de dispositivos**: 247; Longitud de la red: longitud máxima se basa en las normas EIA (RS)-485 1200 m (depende de la velocidad). Puede alcanzar una longitud más larga con repetidores; **Método de transmisión**: típicamente EIA (RS)-485; **Sesgos de industria**: ampliamente utilizado en todas las industrias.
DeviceNet, Allen-Bradley	**Velocidad**: Típicamente 500kbit/s; **Topología física**: Bus; **Topología lógica**: centralizado, maestro/esclavo número máximo de dispositivos: 64; **Longitud de la red**: método de transmisión dependientes (sin repetidores) de la tarifa de datos de 5000m **Método transmisión**: CAN bus, cable de par trenzado con alimentación de red; **Sesgos de industria**: ampliamente utilizado en una amplia gama de industrias, tales como maquinaria de producción general, automoción, etc.
Interbus	**Velocidad**: 500kbit/s y 2 Mbit/s; **Topología física**: repite autobús; **Topología lógica**: centralizada; **Número máximo de dispositivos**: 512, 4096 puntos de la entrada-salida; **Longitud de la red**: 400m entre dispositivos, hasta 13 km de longitud red, **Método de transmisión**: de par trenzado o fibra óptica; **Sesgos de industria**: ampliamente adoptado por los fabricantes de automóviles.
Modbus RTU	**Velocidad**: Dependiente de la velocidad en baudios del enlace serial; **Topología física**: uno-a-uno o autobús; **Topología lógica**: centralizado maestro / esclavo]; **Número máximo de dispositivos**: 247; **Longitud de la red**: longitud máxima se basa en las normas EIA (RS)-485 1200 m (depende de la velocidad). Puede alcanzar una longitud más larga con repetidores; **Método de transmisión**: típicamente EIA (RS)-485; **Sesgos de industria**: ampliamente utilizado en todas las industrias.
EtherCAT	**Velocidad**: 100 Mbit/s; **Topología física**: repite autobús; **Topología lógica**: basado en Ethernet estándar, típicamente repetidos autobús; **Número máximo de dispositivos**: 65536, sin embargo, la tasa de actualización de red se verá afectado; **Longitud de la red**: teóricamente ilimitado, sin embargo, la tasa de actualización en última instancia restringirá; **Método de transmisión**: tecnología Ethernet estándar, con gestión del tiempo; **Sesgos de industria**: control de movimiento general y alto rendimiento. Los mensajes pueden transportarse directamente en una trama Ethernet o encapsularse como carga útil UDP mediante el puerto 34980
S7 / S7+	Es más famoso por ser el equipo y el protocolo que se aprovechó en el ataque Stuxnet, que involucró al programa nuclear iraní. S7+ se introdujo para proporcionar funciones más seguras y ricas para abordar los riesgos de seguridad de los ataques de repetición. Los puertos típicos que se utilizan son 102 y 1099.
PROFINET	**Velocidad**: 100 Mbits/s y superior; **Topología física**: generalmente estrellas; puede ser autobús, árbol o malla; **Topología lógica**: centralizada; **Número máximo de dispositivos**: 200 puntos de la entrada-salida son posibles en una red PROFINET. En la red eléctrica la distancia máxima entre dos dispositivos cualesquiera es 100m. **Método de transmisión**: Ethernet basado con VLAN; **Sesgos de la industria**: automotriz
DNP3	Este es un protocolo utilizado por los sistemas SCADA para equipos de proceso de interconexión utilizados en las industrias de energía y agua. Es un estándar abierto que ha ganado tracción internacional; sin embargo, lo encontrará más comúnmente utilizado en el mercado norteamericano. El puerto típico utilizado es 20000.
Modbus TCP/IP	**Velocidad**: 10Mbit/s, 100 Mbits/s, 1Gbit/s; **Topología física**: generalmente estrellas basado; puede ser de bus, anillo, árbol o malla topología lógica: centralizado, maestro/esclavo; **Longitud de la red**: teóricamente ilimitado número de nodos y distancia, aunque realista limitada por la velocidad de actualización; **Método de transmisión**: Modbus IP utiliza estándar Ethernet industrial, **Sesgos de industria**: General. Los puertos típicos utilizados son 502, 5020 y 7701
Ethernet/IP	**Velocidad**: 10Mbit/s, 100 Mbits/s, 1Gbit/s; **Topología física**: generalmente estrellas; puede ser autobús, árbol o malla; **Topología lógica**: centralizado, maestro/esclavo; **Longitud de la red**: teóricamente ilimitado, sin embargo, tiempos de ciclo y rendimiento de la red será un factor limitante; **Método de transmisión**: utiliza Ethernet estándar; **Sesgos de industria**: maquinaria en General, las máquinas de la industria y la producción auto. Los puertos típicos que se utilizan son 44818 y 2222.

Tabla 2. Principales protocolos industriales. Adaptación de [22]

[12] http://www.ieec.uned.es/investigacion/Dipseil/PAC/archivos/Formacion_Especifica_Tarea_ISE4_3_1.pdf

4.1.4 Metodologías para el Análisis de Datos de un ICS

El tráfico de red industrial de un ICS contiene información que se transfieren entre dispositivos de campo, controladores, estaciones y servidores, Allí pueden encontrarse dependiendo del protocolo utilizado en una aplicación industrial determinada (por ejemplo, comunicaciones TCP/Modbus entre PLC y servidor), patrones que permitan detectar posibles anomalías en el tráfico que eviten inactividad de los equipos industriales y que con un buen planteamiento para el análisis de estos, permitan desarrollar una clasificación de los posibles vectores de ataques que pueden o no afectar a una red de control industrial [5].

A continuación, se presenta un planteamiento de clasificación del tráfico de red ICS teniendo en cuenta una amenaza interna y monitoreo e identificación de tráfico sospechoso, según se explica en el apartado anterior. Las metodologías que permitirán caracterizar el tráfico son: la basada en el puerto, la basada en la carga útil, la basada en estadísticas de flujo y la basada en el host [7]:

- **Metodología basada en puertos**

La clasificación del tráfico utilizando el número de puerto es un procedimiento para distinguir el tráfico de la red industrial, teniendo en cuenta que el paquete tiene tamaño, protocolo, IP origen, IP destino, puerto origen, puerto destino [7].

Es necesario identificar la aplicación que utiliza este tráfico por que de ello depende el tipo de aplicación que se encuentra en tránsito entre los dispositivos de campo. Los clasificadores basados en puertos utilizan los encabezados de los segmentos TCP/UDP para recopilar los datos sobre el número de puerto. Posterior a la determinación del número de puerto, se realiza la correlación con el tipo de protocolo TCP/UDP asignado para así determinar si es una comunicación entre controladores de proceso (UDP) o son comunicaciones con terceros (TCP) que participen en el control, administración o configuración de los equipos de la red de gestión.

El método basado en puertos clasifica el tráfico asumiendo el uso constante de números de puerto TCP o UDP; sin embargo, la existencia de puertos privados y dinámicos hace que estos métodos se vuelvan poco confiables e inexactos.

- **Metodología basada en la carga útil**

Un enfoque alternativo para la clasificación es una técnica basada en la carga útil (también llamada inspección profunda de paquetes DPI) [7]. En esta técnica, la carga útil de los paquetes se inspecciona en busca de firmas características de aplicaciones conocidas. En este método, se realiza una inspección profunda de paquetes (DPI). Esta técnica comprueba primero el número de puerto de la aplicación con la comprobación de firma y después la comprobación con el tipo de protocolo usado. La precisión de la técnica basada en la carga útil es cercana al 79% [7].

Las tasas de detección de este método son buenas, sin embargo, esta técnica también tiene algunas desventajas. Primero, existe una regulación de privacidad o una preocupación legal con los datos de usuario de la carga útil del paquete. La segunda se debe a las técnicas criptográficas utilizadas para cifrar la carga útil del paquete también limitan el uso de técnicas basadas en la carga útil. En tercer lugar, esta técnica debe estar actualizada con el conocimiento de la semántica del protocolo de aplicación que para este caso sería sobre ICS.

- **Metodología basada en estadísticas de tráfico**

Esta técnica depende de los datos que se pueden obtener del encabezado del del segmento TP/UDP (por ejemplo, los bytes enviados, los tiempos entre llegadas, el tamaño de la ventana TCP...) [5]. Dependen de los datos de alto nivel del encabezado no accesibles por DPI, lo que los convierte en una alternativa superior para administrar no accesibles por DPI o situaciones en las que se manejan puertos dinámicos. Esta técnica utiliza características estadísticas de datos de tráfico de Internet en la capa de red. Estas características estadísticas son longitud del paquete, duración del flujo, tiempo entre llegadas del paquete, desviación estándar. Estas características son diferentes para diferentes aplicaciones industriales, por lo que son capaces de distinguir varias aplicaciones.

Definir el objetivo de clasificación es lo primero para construir un clasificador de tráfico de red. Para lograr este objetivo, se categorizarán las clases de tráfico en función de los ICS y sus elementos de control interconectados, como se expone en los siguientes elementos:

- Protocolos (por ejemplo, Modbus, UDP, TCP, DNP3, S7, Melsec, Profinet, Etercap, etc).
- Aplicación (por ejemplo, integraciones con interfaces gráficas de usuario HMI, comunicaciones con bases de datos externas OPC, cierres contables de sistemas de telemetrías, sistemas de turbo-maquinaria, sistemas de parada de emergencia, sistemas de detección y extinción, etc)
- Tipos de tráfico: (por ejemplo, lectura o escritura a una variable, diagnóstico de un estado de control, etc).
- Clasificación binaria (por ejemplo, anomalía o normal).

- **Técnicas basadas en el host**

Se intenta clasificar el tráfico monitorizando el tráfico enviado o recibido desde el mismo host, pero pierde efectividad cuando el host usa más de una aplicación [1][5][6].

4.1.5 Ciencia de datos y sus etapas

En el ciclo de vida de la ciencia de datos se describen las etapas sugeridas que siguen los proyectos para poder brindar un desarrollo organizado. A continuación (Figura 9) se menciona la metodología aplicada al presente trabajo:

Figura 9. Ciclo de datos y sus etapas. Adaptación Autor a Referencia [1]

Este ciclo de vida está diseñado para los proyectos de ciencia de datos que se enviarán como parte de aplicaciones. Estas aplicaciones implementan modelos de aprendizaje o inteligencia artificial de máquina para realizar un análisis predictivo que puede ayudar en la identificación del objetivo. El proceso del ciclo de vida de los datos utilizado en este trabajo [1] se muestra en la Figura 8. Los principales componentes y etapas consisten en:

- **El conocimiento del negocio**: En esta fase, es importante comprender e identificar los problemas a identificar o mejorar en el negocio. Deberá formularse preguntas que definan los objetivos empresariales y a las que puedan aplicarse las técnicas de ciencia de datos.
- **Adquisición:** Una vez identificado el objetivo del ejercicio, deberá extraerse los datos de manera que se puedan trabajar en el preprocesamiento. En esta etapa inicial de recopilación de datos, se identifican y reúnen los recursos de datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del problema.
- **Análisis exploratorio de los datos**: Antes de entrenar los modelos, se debe desarrollar una comprensión sólida de los datos. En ciertas ocasiones dependiendo de la fuente, los conjuntos de datos reales contienen valores nulos, errados y hasta faltantes. En esta fase deberá utilizarse funciones de ajustes y acondicionamiento de datos para dar la información que se necesita para procesarlos y dejarlos preparados para el modelado. Es esta fase se pueden encontrar procesos de selección y extracción de características objetivo, conversión de formato, limpieza, balanceo, codificación de etiquetas y validación de datos.
- **Modelado (Entrenamiento y Evaluación)[13]:** La etapa de modelado utiliza la primera versión del conjunto de datos preparado y se enfoca en desarrollar modelos predictivos o descriptivos según el enfoque analítico previamente definido. El proceso de modelado normalmente es muy iterativo, ya que las organizaciones están adquiriendo *insights* intermedios, lo que deriva en ajustes en la preparación de datos y en la especificación del modelo. En esta fase se caracteriza por la clasificación de aprendizaje automático: capacitación (80 % del conjunto de datos) y pruebas (20 % del conjunto de datos). En el proceso de Evaluación, se tiene del modelo implica el cálculo de varias medidas de

[13] https://learn.microsoft.com/es-es/azure/architecture/data-science-process/lifecycle-modeling

diagnóstico y de otros resultados, como tablas y gráficos, lo que permite interpretar la calidad y la eficacia del modelo en la resolución del problema.

- **Implementación**: Cuando el modelo satisfactorio ha sido desarrollado y aprobado por los promotores del negocio, se implementa en el entorno de producción o en un entorno de pruebas comparable.

4.1.6 Algoritmos de detección de anomalías

ML es una aplicación de inteligencia artificial que utiliza algoritmos para permitir que una computadora aprenda de los datos de entrada. Los datos proporcionados a un algoritmo de aprendizaje automático se utilizan para generar una comprensión de los datos, en función de lo que algoritmo puede "*aprender*" de la entrada que se le da. El objetivo de cualquier técnica de aprendizaje automático es identificar patrones (a menudo muy complejos) en grandes cantidades de datos, patrones que incluso un ser humano muy hábil no sería capaz de reconocer debido a la enorme cantidad de información que a menudo se presenta [11].

Estas técnicas aprovechan la capacidad de una computadora para procesar cantidades masivas de datos rápida y eficientemente. En este trabajo, se utiliza el aprendizaje automático para proporcionar clasificación. En clasificación, los datos se entregan a un algoritmo de aprendizaje automático como instancias múltiples, cada uno consiste en un conjunto de valores. Estos valores pueden tomar múltiples formas, incluyendo cadenas (una secuencia de caracteres, como una palabra), un número entero, un número real o una selección de uno de múltiples valores discretos (como verdadero o falso). Cada uno de estos valores se llama un atributo, o una característica y es la unidad con la que inicia a trabajar la clasificación. La formación del entendimiento de este atributo se utiliza para formar un modelo [12].

A continuación, se presenta una descripción de los diferentes algoritmos en el ML:

- **Aprendizaje automático supervisado**

El algoritmo de aprendizaje automático supervisado contiene un objetivo (o variable dependiente) que es ser pronosticado a partir de un conjunto dado de predictores (variables

independientes). Usando estos conjuntos de variables, generando una función que asigna las entradas a las salidas deseadas. El proceso de formación continua hasta que el modelo alcanza el nivel deseado de precisión en los datos de entrenamiento.

Ejemplos de Aprendizaje Supervisado: **Regression, Decision Tree, Random Forest, K-Nearest Neighbor, Logistic Regression, etc**. [14].

- **Aprendizaje automático no supervisado**

ML no supervisado no tiene ningún objetivo o variable de resultado para predecir/estimar. Está utilizado para agrupar la población en diferentes grupos, que es ampliamente utilizado para segmentar clientes en diferentes grupos para una intervención específica.

Ejemplos de aprendizaje no supervisado: **K-means.**

- **Aprendizaje automático de refuerzo**

En el Reinforcement Machine Learning la máquina está entrenada para tomar decisiones específicas. Funciona de esta manera: la máquina está abierta a un entorno en el que se entrena continuamente usando prueba y error Esta máquina aprende de experiencias pasadas y trata de capturar lo mejor posible conocimiento para tomar decisiones acertadas.

Ejemplo de aprendizaje por refuerzo: proceso de decisión de **Markov**

4.1.6.1 Algoritmos de ML para la clasificación de tráfico en ICS

Como se presenta en el objetivo general de este trabajo, se requiere trabajar en un sistema de análisis de tráfico de red industrial que analice el estado "**normal**" de uno "**anormal**", por lo que la metodología a trabajar nos permite identificar esquemas o patrones de forma que, al suministrar los paquetes restantes sin enseñar la etiqueta, puede intentar predecir a qué etiqueta pertenece a partir de los patrones que ha encontrado con los datos de entrenamiento.

Teniendo en cuenta el requerimiento anterior sobre los datos, el mejor modelo a trabajar es ML Supervisado. En un modelo de aprendizaje supervisado, el algoritmo necesita un conjunto de datos etiquetados para el aprendizaje, se necesita supervisión para entrenar el modelo. Para evaluar su precisión, se comprobará la eficacia de esos algoritmos a la hora de clasificar los paquetes de un tráfico previamente capturado en una red de control industrial real en el cual se conoce el estado "*normal*" y la manipulación de una muestra para el entrenamiento en modo "*anormal*" teniendo en cuenta características determinísticas en ataques reales llevados a cabo sobre redes de área local que afecten las características de análisis planteadas en el trabajo.

Así las cosas, los modelos a trabajar teniendo en cuenta el ML Supervisado con el cual se determinará el estudio, serán los siguientes [14]:

- **Regresión Logística**

Un uso muy popular de los modelos lineales generalizados (GLM) es a través de la regresión logística. Este es el tipo de regresión que se usa cuando la variable objetivo es un valor binario, es decir, solo puede tomar dos valores. Estos dos valores binarios suelen tomar la forma:

Tráfico Anormal (valor binario = 1)

Tráfico Normal (valor binario = 0)

- **K-Nearest Neighbor (KNN)**

El algoritmo K-Nearest Neighbor (KNN) es un método para clasificar instancias según las instancias de entrenamiento más cercanas en el espacio de características. KNN es un tipo de aprendizaje basado en instancias donde la función solo se aproxima localmente y todos los cálculos se posponen hasta la clasificación. El KNN es el método de clasificación importante y más simple cuando es poco o ningún conocimiento previo aproxima la distribución de los datos. Esta regla simplemente conserva el conjunto de entrenamiento integral durante el aprendizaje y asigna a cada solicitud una clase representado por la etiqueta popular de sus k-nearest más cercanos en el conjunto de entrenamiento.

- **Máquina de soporte vectorial (SVM)**

Support Vector Machine (SVM) es un algoritmo de aprendizaje automático supervisado. Cuando se usa como algoritmo de clasificación, separa un conjunto de datos de entrenamiento etiquetado dado con un hiperplano es la distancia máxima a ellos (hiperplano de margen máximo).

- **Naive Bayes**

Este algoritmo tiene el objetivo de clasificar los datos, de forma que se ordenen de forma filtrada y así puedan ser utilizados para optimizar el proceso de toma de decisiones. Los mecanismos Naive Bayes son objetivos, ya que el alcance de este mecanismo permite obtener dos respuestas: alta o baja probabilidad. Por tanto, se posibilita una decisión asertiva a partir de los datos analizados. Así, da como resultado una tabla de probabilidades que luego será analizada por ML , que a su vez se encarga de separar y clasificar la información,

- **Decision Tree**

Los árboles de decisión se consideran un buen modelo predictivo para empezar y tienen muchas ventajas. Entran en juego la interpretabilidad, la selección de variables, la interacción de variables y la flexibilidad para elegir el nivel de complejidad de un árbol de decisiones. Los Decision Tree clasifican las ocurrencias organizándolas en función de los valores de características denominados árboles de decisión. Cada nodo en un árbol de decisión representa una característica en una ocurrencia que se va a clasificar y cada rama representa un valor que el nodo puede asumir. Las instancias se clasifican a partir del nodo raíz y ordenados según sus valores característicos para:

• Comprobar que todos los casos pertenecen a la misma clase.
• Calcular información y ganancia de información.
• Encontrar la mejor división del atributo.

- **Random Forest**

Un bosque aleatorio es un estimador que ajusta una serie de clasificadores de árboles de decisión en varias submuestras del conjunto de datos y utiliza el promedio para mejorar la precisión predictiva y controlar el sobreajuste.

4.2 ESTADO DEL ARTE

Antes de poder realizar el estudio objeto de este trabajo, se identificaron trabajos similares y relacionados con el análisis de tráfico de red anómalo evidenciando trabajos desarrollados sobre bases de datos de tipo industriales. En la presente sección se mencionan algunos de los trabajos de los cuales se tomó importantes elementos de análisis del tráfico ICS propio de un sistema industrial. La elección los features, características, parámetros o campos a usar con los algoritmos de ML en el presente trabajo, así como el establecimiento de las clases en las que se clasifica el tráfico de red industrial, se ha realizado teniendo en cuenta la tabla 3, la cual exponen la clasificación de tráfico en trabajos documentados y se indica cuál es la relación que se tiene con el presente trabajo expresado en el elemento de la tabla "DESARROLLO DIRIGIDO A" :

REF	PARÁMETROS DEL DATASET	MODELOS UTILIZADOS	DESARROLLO DIRIGIDO A	MÉTODOS UTILIZADOS EN EL DESARROLLO	PROPÓSITO DEL MODELO
[1]	El conjunto de datos del banco de pruebas de ICS utilizado en esta investigación es un subconjunto de un conjunto de datos público disponible en Internet. Se utilizan un total de 120025 paquetes de red que consisten en benignos (operación normal del banco de pruebas, sin ataques simulados) y anómalos (ataques simulados).	Ada Boost Classifier, Random Forrest Classifier, Decision Trees, Voting Ensemble Classifier, K-Nearest, Neighbors	Internet of Things (IoT) Industrial Internet of things (IIoT)	Uso de 5 ataques simulados diferentes en un conjunto de datos de un banco de pruebas del Sistema de control industrial (ICS)	Modelo biclasificatorio

[2]	Según el análisis profundo del protocolo S7COMM, los tipos de PDU incluyeron: jO B , AC _k, AC _k—DATOS _ _Ay DATOS DE USUARIO	Support Vector Machine (SVM) Decision tree K-Nearest Neighbors (KNN) Random Forrest Classifier AdaBoost	PLC seleccionado permite que el software SCADA se comunique con el PLC mediante el protocolo de propiedad industrial S7COMM basado en la conexión TCP	El sistema SCADA se comunica con el PLC de la serie S7-300 de Siemens para simular el tráfico de red normal de la fábrica y recopilar información de estado. El host de ataque se conectó a la red SCADA a través de la conexión directa con el conmutador, para implementar ataques MITM por envenenamiento ARP.	Modelo biclasificatorio
[3]	Los conjuntos de funciones de esta aplicación se centran en los valores específicos asociados con los datos RTU y los resultados de pruebas simples que proporcionan controles de la integridad de los datos y el protocolo de las transacciones.	Naïve Bayes Random Forests OneR, J48 NNge, SVM	SCADA	Los datos utilizados para este experimento son una colección de flujos de telemetría RTU etiquetados de un sistema de gasoducto en el Centro de Protección de Infraestructura Crítica de la Universidad Estatal de Mississippi	Modelo biclasificatorio
[4]	Clasificar el tráfico de red con precisión (QoS) de la red, particularmente para aplicaciones sensibles al tiempo, y optimizar el ancho de banda disponible.	Random Forest (RF) K-Nearest Neighbors (KNN) Deep Learning (DL)	Internet	La mayoría de los trabajos seleccionaron conjuntos de datos públicos como ISCX VPN-non VPN e ISCX2012.	Clasificación binaria
[6]	En este trabajo se simularon ataques de denegación de servicio, ataques de intento de reconocimiento y ataques de acceso no autorizado contra protocolos Modbus.	Random Forrest Classifier Decision Tree Support Vector Machine (SVM)	Industrial Control Systems (ICS)	Se utilizó como servidor un VPS ubicado en los Estados Unidos y se implementó el honeypot de Conpot en el servidor para simular el PLC SIMATIC S7-200.	Clasificación binaria
[8]	Resultados experimentales de los datos de referencia de KDD-Cup99 IDS.	(ACCM) Ant Colony Clustering Model	Industrial Control Systems (ICS)	El conjunto de datos KDD-Cup99 del repositorio de la base de datos de UCI (http:/wwwJcs.uciedu/~mlearn/) se usa comúnmente como datos de referencia para la evaluación de IDS	Clasificación cluster
[9]	Sistemas de control de supervisión que interactúan con varios dispositivos electrónicos inteligentes complementados con dispositivos de monitoreo de red como los sistemas SNORT y Syslog	OneR NNge Random Forests Naïve Bayes Support Vector Machine (SVM)	SCADA, Power system, Industrial control systems, such as those used in the Smart Electric Grid	La red está compuesta por 4 interruptores controlados por relés electrónicos inteligentes.	Multiclass Binary

[11]	La Universidad Estatal de Mississippi y el Laboratorio Nacional de Oak Ridge implementaron una versión reducida de un marco de sistema de energía.	Bayesian Network, Naive Bayes J48 Decision Tree, Support Vector Machine	Intrusion Detection Systems (IDS) Industrial Control Systems (ICS)	La red está compuesta por 4 interruptores controlados por relés electrónicos inteligentes.	Multiclass Binary
[12]	Comparación de trabajos relacionados con el tema de aplicaciones, desafíos y recomendaciones en ML.	Deep learning	Critical infrastructures IoT, Industrial Control Systems	Banco de pruebas de ICS de gas relativamente simple de la Universidad Estatal de Mississippi (MSU).	Investigación

Tabla 3. Comparación entre los trabajos relacionados. Autor

Aunque no hay una evidencia exacta sobre los parámetros utilizados en cada trabajo, se tiene que la mayoría trabajan con datasets originado por entornos industriales simulados. En el presente trabajo, se toman para el proceso de investigación parámetros como son:

- Tamaño de paquetes
- Protocolo,
- IP origen, IP destino,
- Puerto origen, puerto destino
- Tiempo entre paquetes, entre otros.

La cantidad de tráfico para lograr un buen entrenamiento del modelo es también una de las variables que se estudia en este proyecto, teniendo en cuenta el procesamiento de los datos, y así se comprobará si un algoritmo necesita un entrenamiento con más paquetes para poder clasificar el tráfico con mayor precisión y a partir de qué número de paquetes no sigue mejorando la clasificación.

Los paquetes del tráfico industrial son seleccionados de manera aleatoria para evitar que el ejercicio del cual se basa el presente trabajo se vea afectado por el orden de los paquetes o un conjunto específico de ellos. Para cada uno de los trabajos mencionados, se tiene que es importante contar con la suficiente cantidad de tráfico este estudio para así poder proporcionar una buena clasificación de tráfico. De manera muy similar, se tienen Indicadores de Compromiso (IoC) [1] y de calidad del servicio [5] para cada uno de los diferentes ataques de red simulados. Aun cuando no específicamente se tratan sobre tráfico de red ICS, los trabajos [3] [9] [20] tratan de análisis

sobre sistemas SCADA, relacionados entre sí con los ICS, aplica en conceptos y modos de análisis, pues en los primeros la característica especial hace referencia a diferentes protocolos industriales a nivel de capa 2 según el modelo OSI. El rendimiento de los algoritmos depende de las diferencias entre los algoritmos elegidos y su configuración específica, así como de los parámetros o características elegidos para la clasificación [3].

Capítulo 5
METODOLOGÍA

Este trabajo se centra en una evaluación de métodos de aprendizaje automático como discriminadores de tráfico en las comunicaciones de sistemas de control industriales del tipo maliciosas. En el contexto de la seguridad cibernética, el presente trabajo se centra en discriminar el tráfico de red malicioso del tráfico no malicioso [3].

Para desarrollar el sistema de análisis de tráfico de red se usará Scikit-Learn que es una librería de código abierto enfocada en la implementación de ML en Python. Su objetivo es el minado y análisis de datos, por lo tanto, ofrecen herramientas que facilitan este objetivo ocultando la matemática que hay detrás. [14].

A continuación, se presenta la metodología con la cual se ha desarrollado el presente trabajo, basada en el capítulo 4.1.5:

5.1 Conocimiento del Problema

Con el propósito de construir un modelo de clasificación que distingue entre tráfico "**normal**" y "**anormal**" en una red de control industrial, se toma una recopilación de tráfico real de un ICS compuesta por controladores, servidores, historiadores, switches, firewall, estaciones de operación e instrumentación de campo de una infraestructura crítica. Tabla 4.

nombre	cantidad	descripción
firewall	1	Firewall interconectado entre red de control y corporativa
HISTORIAN	1	Equipo servidor donde se guardan las tendencias y configuraciones del sistema de control
ESTACIONES OPERACIÓN	6	Estaciones cliente del servidor principal desde donde el operador controla y manipula los elementos de campo.
SERVIDOR	2	Servidores Principal / Secundario donde reside la base de datos principal del ICS.
SWITCHES	5	Equipos donde se encuentran interconectadas estaciones de operación, servidores, historiador, procesadores de control y servidores OPC
SERVIDOR OPC	1	Servidor de comunicaciones con terceros, los cuales se encuentran interconectados por protocolo OPC.

Tabla 4. Arquitectura ICS para la recolección del dataset. Autor

5.2 Adquisición de Datos

Para el conjunto de datos (o dataset) de pruebas, se instaló un equipo para recolectar tráfico real de un ICS en lugar de seleccionarlos de una base de datos abierta donde se pueden obtener trazas de tráfico. El hecho de tomar las características específicas del sistema en consideración sobre el tipo de tráfico de control industrial, facilita la labor de identificación, que luego servirá para el entrenamiento y la validación.

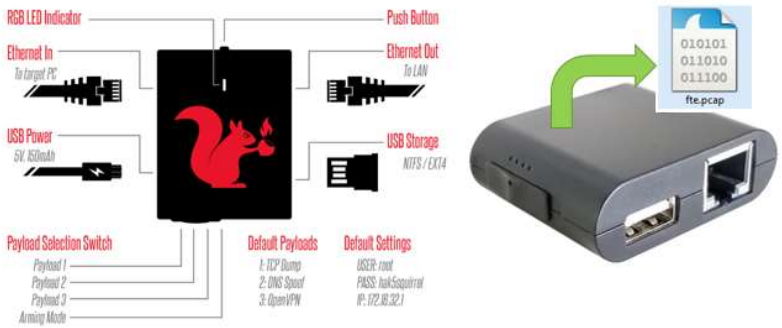


Figura 10. Dispositivo de recolección del dataset. Autor

El dataset denominado ***fte.csv*** es un conjunto datos que representan el tráfico real del ICS según la arquitectura de la figura 9, fue adquirido de mediante un dispositivo electrónico recolector de tráfico denominado packet squirrel[14] el cual es una herramienta de conexión Ethernet diseñada para capturas de paquetes, instalado en el puerto espejo de uno de los switches de red de administración del sistema de control.

El packet squirrel interconectado a un puerto en modo espejo pasivo de unos de los switches troncales recolectó de manera offline, un total de 3 meses de datos asociados al tráfico de la arquitectura propuesta en la figura 11 y lo guardó en un archivo con extensión. ***pcap*** que representaba todo el tráfico del segmento de red determinado para el posterior análisis. Para poder realizar el análisis inicial fue revisado en el software Wireshark y después se exportó en un archivo .csv para poderlo analizar por los algoritmos ML.

[14] https://docs.hak5.org/packet-squirrel/getting-started/packet-squirrel-basics

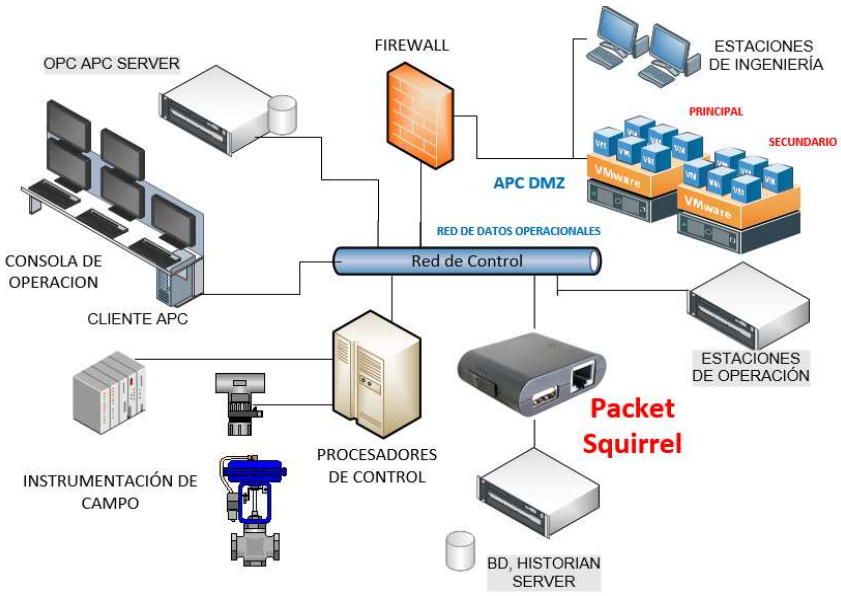


Figura 11. Arquitectura ICS para la recolección del dataset. Autor

5.3 Análisis exploratorio de los datos (EDA)

En esta fase del proceso se utilizan el análisis exploratorio de datos (EDA) para analizar e investigar conjuntos de datos y resumir sus principales características, a menudo empleando métodos de visualización de datos orientados a los resultados del dataset de datos de tráfico de red. Asimismo, es indispensable en el proceso de acondicionar los datos para obtener las respuestas necesitadas en la investigación orientadas al análisis de tráfico OT, lo que facilita en el presente trabajo a descubrir patrones, detectar anomalías, probar una hipótesis o comprobar supuestos.

En consecuencia, se etiquetan estos "*features*" o características como "**anormal**", mientras que los rastros de tráfico limpio, obtenidos directamente en la red se etiquetan como "**normal**". El propósito principal del análisis exploratorio es permitir realizar procesos de identificación y validación d ellos datos antes de presentar un juicio. Es indispensable identificar errores, así como comprender mejor los patrones dentro de los datos, detectar valores atípicos o eventos anómalos.

Selección y Extracción de las características (features)

El objetivo es identificar a qué clase (también llamada etiqueta) pertenece cada instancia en el conjunto de prueba. Los administradores de sistemas y redes tienen como elementos bases de análisis tuplas de 5 para identificar los requisitos clave para crear una conexión de red segura, operativa y bidireccional entre dos o más máquinas remotas y locales, es decir, **IP de origen, IP de destino, número de puerto de origen, número de puerto de destino, protocolo**. Figura 11.

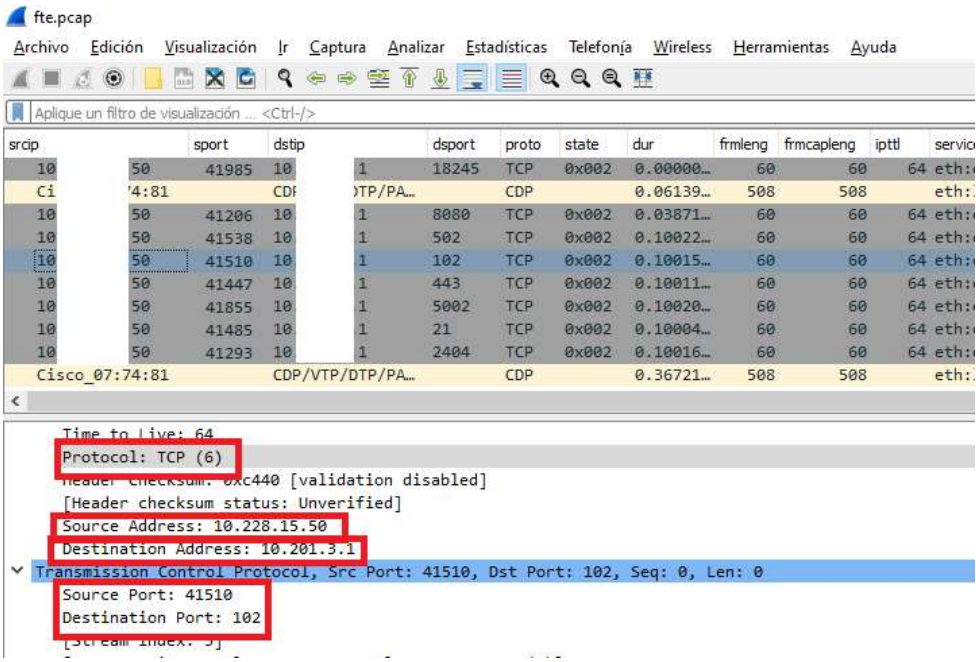


Figura 12. Tupla 5-TCP en archivo fte de tráfico ICS. Autor

Sin embargo, para el desarrollo del presente trabajo no solo se tomaron las características asociadas a las tuplas, sino también las referenciadas a tamaños del tráfico, duración del tiempo de la comunicación, la pérdida de datos generado entre los datos en tránsito y los datos capturados, entre otros. La tabla 5 representa las características de extracción y selección para el tráfico de red ICS adquirido en el proceso de extracción de datos una vez han sido analizados y seleccionados por el software Wireshark en el momento de la exportación del archivo ***fte.cap*** a ***fte.csv***.

item	features	definición	TIPO	description
1	srcip	srcip	object	Dirección IP fuente
2	sport	sport	float64	Puerto fuente
3	dstip	dstip	object	Dirección IP destino
4	dsport	dsport	float64	Puerto destino
5	proto	proto	object	Protocolo usado en la comunicación
6	state	tcp.flags	object	Conversaciones TCP están completas cuando tienen protocolos de enlace tanto de apertura como de cierre, independientemente de cualquier transferencia de datos. (0x002:SYN, 0x010:SYN-ACK, 0x014:ACK, 0x018:DATA, 0x012:FIN, 0x011:RST)
7	dur	frame.time_delta	float64	Duración del paquete
8	frmleng	frame.length	int64	Bytes enviados
9	frmcapleng	frame.cap_len	int64	Bytes recibidos, longitud de fotograma almacenada en el archivo de captura.
10	lbytes	Calculed in dataset	int64	Bytes perdidos en la transacción
11	ipttl	ip.ttl	int64	Tiempo de vida del paquete (Time To Live)
12	service	frame.protocols	object	Tipo de servicio en el protocolo de comunicaciones
13	tcpwinsize	tcp.window_size_value	float64	Tamaño de la ventana de la cabecera TCP
14	iplen	ip.len	float64	Longitud total del protocolo de comunicaciones
15	tcplen	tcp.len	float64	Longitud del segmento TCP
16	tcpwinsz	tcp.window_size	float64	Tamaño de la ventana calculada
17	tcphdrlen	tcp.hdr_len	float64	Longitud del encabezado TCP
18	tcprtt	tcp.analysis.initial_rtt	float64	Tiempo desde el SYN hasta el ACK en el handshake
19	Label		int64	0 para Normal y 1 para Anormal

Tabla 5. Descripción de los parámetros o características del dataset. Autor

Conversión de formatos para los datos

La conversión de datos es el proceso de traducir los datos de un formato a otro, manteniendo su viabilidad y calidad según sea la necesidad (ejemplo: int62 a float64).

Codificación de etiquetas

En el aprendizaje automático, generalmente tratamos con conjuntos de datos que contienen múltiples etiquetas en una o más de una columna. Estas etiquetas pueden tener la forma de palabras o números. Para que los datos sean comprensibles o legibles por humanos, los datos de entrenamiento a menudo se etiquetan con palabras o números.

La codificación de etiquetas se refiere a convertir las etiquetas en un formato numérico para convertirlas en un formato legible por máquina. Se analizará el proceso de decodificación en el presente trabajo teniendo en cuenta la dificultad de manejar una dirección IP o una palabra.

5.4 Modelado

Entrenamiento

Los archivos de características resultantes de la fase anterior se proporcionan como entrada a los algoritmos de clasificación. La intención es construir modelos que tengan la capacidad de distinguir entre tráfico "**normal**" y tráfico "**anormal**".

¿Qué algoritmo utilizar?

La mayoría de los trabajos mencionados en la tabla 3 se basan en aprendizaje supervisado, es decir; que dependen de información que ya se encuentra etiquetada y en la cual se va a desarrollar el análisis basado en un conjunto de datos. En este caso; se utilizan los algoritmos para analizar datos de entrenamiento y un aprendizaje de la función que se asigna sobre la salida requerida. Para nuestro caso en particular, no tenemos salidas como funciones de aprendizaje, pues la determinación de un tráfico anómalo se basa en el conocimiento de la hipótesis de lo que, si se considera tráfico normal. Así las cosas; el modelo de aprendizaje supervisado no aplicaría para este trabajo.

El planteamiento de la solución de detección requiere de una predicción numérica rápida que aunque no presente un etiquetado como elemento directo de identificación, permita con una cantidad limitada de datos etiquetados dados por la experiencia de lo conocido, mejorar la precisión del aprendizaje. El modelo para desarrollar presenta un esquema de clasificación de forma ***semisupervisada***, pues no se tienen las características del estado "*anormal*" en las redes ICS en las cuales se desarrolla el análisis, por el contrario, solo se conoce la condición "*normal*" de las mismas teniendo en cuenta los modos normales de funcionamiento del sistema industrial. En las clasificaciones de forma semisupervisadas se conoce la existencia de las anomalías del tráfico ICS a analizar, pero estas no se contemplan en el conjunto de datos original.

Con el conjunto de datos de tráfico ICS se llega a definir la normalidad del esquema operacional, un supuesto en el que es necesario establecer un comportamiento normal se da a cabo cuando hay tráfico, por ello que es importante afrontar este problema modelando de manera adecuada el comportamiento de las características (features) con el fin de detectar las anomalías existentes.

Así las cosas, la anomalía que se va a tratar en el tráfico ICS en estudio, es el estado diferente del resto del dataset y representa una condición de funcionamiento operativo en la red industrial de manera normal para el cual se permite generar etiquetas en el modo ***test*** para el estado "anormal", construyendo de manera manual paquetes que no hacen parte de la normalidad de la red ICS e inyectándolos al modelo para ver si el modelo es capaz de identificar que son anormales.

No obstante, como la condición de anomalía es una variación de la normalidad, se necesita un criterio o medida para definir este hecho. Se tiene entonces que, para poder desarrollar estos criterios, es importante tener en cuenta las siguientes recomendaciones: [15]

- **Definir el criterio normal**: el cual limita o establece la posibilidad de determinar el conjunto de los datos anómalos.
- **Identificación de acciones maliciosas**: las cuales establecen el patrón de medición y generación de modelos para su identificación y clasificación.
- **Características especiales de conocimiento**: mediante las cuales se analiza la anomalía desde los diferentes puntos de vista de especialistas en el campo de aplicación, para este caso del conocimiento de sistemas de control industrial.
- **Distinguir ruido de anomalía**: el ruido puede llegar a ser similar al conjunto de datos anómalos, lo que hace que la diferenciación entre estos dos sea una tarea desafiante.

Usando la librería Scikit-learn se analiza el archivo .csv que proviene de la configuración realizada con Wireshark y se procesa para ajustar solo con la información que nos interesa que use el algoritmo de Machine Learning, algunos de estos parámetros se obtienen directamente de la cabecera de cada paquete mientras que otros son calculados (por ejemplo, el tiempo transcurrido entre el paquete actual y el anterior recibidos que pertenecen al mismo socket o la cantidad de paquetes perdidos teniendo en cuenta paquetes en tránsito vs paquetes capturados) [12].

Los parámetros que usaremos con los algoritmos de Machine Learning para analizar los datos se presentan en la tabla 5. Una vez hemos procesado el dataset, esta información se guarda como

[15] https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/63253/TFG-Arnott%20Iglesias%2C%20Ignacio.pdf?sequence=1

archivos CSV ya que las librerías que usamos ya implementan métodos que nos permiten leer fácilmente este tipo de archivos. Se usarán los siguientes algoritmos de aprendizaje supervisado: Regression, Decision Tree, Random Forest, K-Nearest Neighbor, Logistic Regression para construir modelos de detección de anomalías, y sobre estos últimos indicar cuales se van a utilizar en este proyecto.

Para el presente trabajo se tomará el 80% de los paquetes para entrenamiento y el 20% para las pruebas. [10].

Evaluación de los Clasificadores

Los modelos de clasificación semisupervisado se ven perjudicados por cometer demasiados falsos positivos cuando un nuevo dato entra en el modelo y es clasificado incorrectamente, de manera que se considera una anomalía y no tiene por qué ser cierto. La manera de solucionar este tipo de problemas es generando clasificadores basados en reglas particulares del tráfico ICS, teniendo en cuenta que solo existe una hipótesis real de lo que es normal en el tráfico tomado sobre el objetivo, y que cualquier elemento de juicio adaptado sobre este por parte del analista estará basado en conceptos y experticia en temas como protocolos industriales, riesgo sobre la seguridad del proceso industrial y sobre todo el posible impacto de la materialización de una vulnerabilidad en la operación. En el trabajo presentado, se considera que un evento anómalo contiene características de verificación diferentes de un tráfico normal, entre ellos; diferentes valores de carga en los puertos de origen y destino fuera d ellos rangos ya conocidos, evidenciando con eso, posibles pérdidas de tráfico en el tiempo de vida del paquete.

Para medir el rendimiento de los algoritmos de aprendizaje automático, utilizamos tres evaluaciones que son puntos de referencia: Matriz de confusión, Precisión (AC), Tasa de falsos positivos (TPR) y Tasa de falsos negativos (FNR) [2].

- **Matriz de confusion**

Una matriz de confusión es un resumen de los resultados de predicción de un problema de clasificación. Los números de predicciones correctas e incorrectas se resumen con valores de conteo y se desglosan por cada clase. La matriz de confusión muestra las formas en que el modelo

de clasificación se confunde cuando hace predicciones. Nos da una idea no solo de los errores que se están cometiendo por un clasificador, pero lo que es más importante, los tipos de errores que se están cometiendo [5]. Para evaluar el desempeño del modelo de clasificación se utiliza la matriz de confusión. La matriz compara los valores objetivos reales y el valor previsto proporcionado por el modelo. Esto brinda información sobre si nuestro modelo comete errores o sobre qué tan bien se está desempeñando nuestro modelo de clasificación. La matriz de confusión tiene cuatro términos:

1. **True Positives (TP)**: aquellos valores que se predicen como positivos y la salida real también es positiva.
2. **True Negatives (TN)**: aquellos valores que se predicen como negativos y la salida real también es negativa.
3. **False Positives (FP)**: aquellos valores que se predicen como positivos y la salida real es negativa.
4. **False Negatives (FN)**: aquellos valores que se predicen como negativos y la salida real es positiva.

		PREDICCIÓN	
		POSITIVO	NEGATIVO
CLASE REAL	POSITIVO	True Positives (TP)	False Negatives (FN)
	NEGATIVO	False Positives (FP)	True Negatives (TN)

Figura 13. Matriz de Confusión. Autor

- **Precisión**

La evaluación de nuestros datos para su desempeño se realiza a través de sus predicciones positivas. La precisión se define por la Ecuación (1) de la siguiente manera:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

Ecuación (1)

- **Tasa de falsos positivos (FPR) y Tasa de falsos negativos (FNR)**

Por otro lado, el FPR y FNR reflejan el número total de flujos mal clasificados. Es más, el FPR está dedicado a los flujos fuera de clase que se clasifican erróneamente como en clase, y el FNR es reservado para el flujo en clase que está mal clasificado como fuera de clase. La ecuación (2 y 3) define formalmente FPR y FNR respectivamente. El objetivo es lograr una alta precisión con un bajo número de falsos (positivos y negativos).

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \qquad TNR = \frac{TN}{N} = \frac{TN}{TN + FP}$$

Ecuación (2)

$$FPR = \frac{FP}{N} = \frac{FP}{TN + FP} \qquad FNR = \frac{FN}{P} = \frac{FN}{TP + FN}$$

Ecuación (3)

Capítulo 6
RESULTADOS Y DISCUSIÓN

6.1 Etapa de recolección

El desarrollo del dataset denominado ***dataset_ICS_traffic.csv*** se encuentra basado en el tráfico de red de un sistema de control distribuido DCS marca Honeywell (Figura 11), el cual tiene como red de control switches que interconectan los controladores, los servidores, estaciones de ingeniería y operaciones, descritos en la tabla 4 del presente documento. Esta red interconecta comunicaciones entre los elementos de procesamiento de señales de campo y controladores de control, además de los HMI (Human Machine Interfaces) y sistemas terceros, entre ellos los de parada de emergencia, sistemas de detección de fuego y administración de turbomaquinaria de una planta industrial que tiene una corrida de operación de 5 años, es decir trabaja las 24 horas del día, 7 días a la semana durante 5 años seguidos.

Cualquier indisponibilidad al sistema de control, protección o turbomaquinaria causada por un tráfico anómalo puede generar afectación al medio ambiente, a personas, imagen de la empresa o pérdida variables y dependiendes de la funcionalidad dentro de la cadena de producción, en las cuales el lucro cesante podría oscilar entre $MUSD 0,5 – 1,5, de manera que evitar una materialización de este vector de ataque (tráfico de red de control) es posible lograrlo si se encuentra siendo monitoreado constantemente para detectar alguna anomalía en las comunicaciones.

Los features configurados en la tabla 5 del dataset, hacen referencia a los elementos característicos de la configuración básica de Wireshak en adición con características específicas del frame dentro del protocolo TCP/IP, con el cual se evidencia la interconectividad y conexión entre los equipos de la red de control. El dataset, es una combinación de tráfico del mismo segmento de red de control industrial, tomados durante 3 días seguidos, cada 15 días usando el mismo procedimiento para garantizar el proceso de repetibilidad de las muestras en el mismo punto de toma de datos.

Los datos adquiridos en formato .pcap por el dispositivo packet squirrel fueron suministrados en cinco (5) archivos etiquetados desde *fte_01.pcap* hasta fte_05.pcap en donde se tomaron datos de aproximadamente dos (2) meses de tráfico industrial. Ver Tabla 6.

nombre	Fecha	tamaño	paquetes
dataset_ics_traffic.csv	22/08/22	114.199 KB	756.255
fte_01.pcap	05/09/22	16.676 KB	156.021
fte_02.pcap	19/09/22	14.257 KB	111.349
fte_03.pcap	03/10/22	17.172 KB	171.987
fte_04.pcap	17/10/22	14.921 KB	143.939
fte_05.pcap	31/10/22	17.512 KB	172.959

Tabla 6. Características del dataset_ICS_traffic.csv. Autor

Una vez recolectados los datos del tráfico, se fusionaron en Wireshark para suministrar un archivo exportado denominado ***dataset_ics_traffic.csv.*** Para este caso, Wireshark nos permite adecuar el tráfico capturado en la red de control y que una vez desarrollada esta tarea; se extraerá la información útil de cada paquete para luego etiquetarlos. Esta información se guardará en un archivo CSV (Comma-Separated Values) para que un programa en lenguaje Python pueda leer y analizar los datos de manera adecuada. Ver figura 14.

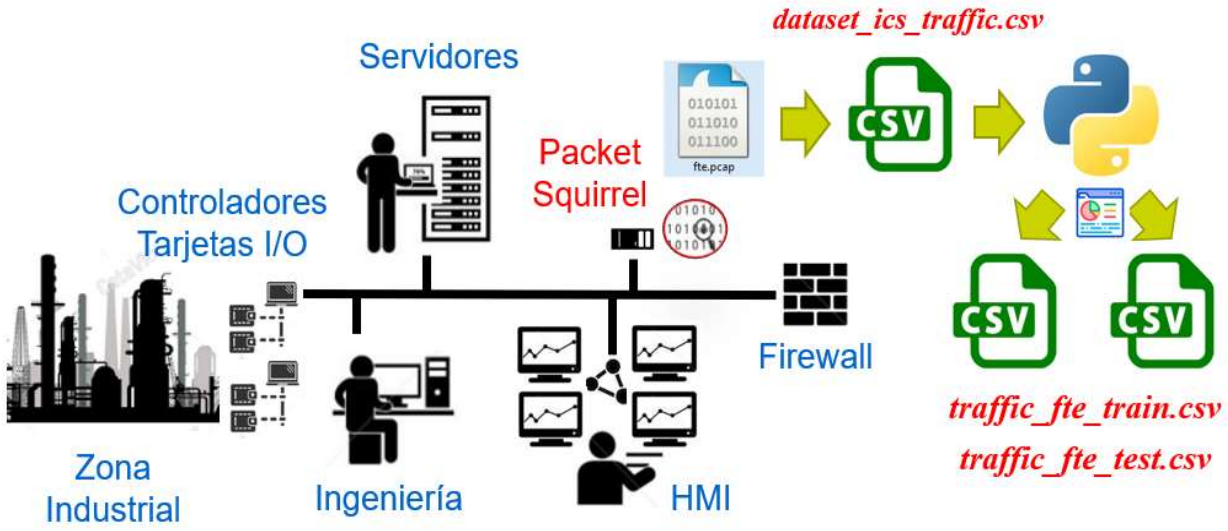


Figura 14. Metodología para extraer dataset entrenamiento y prueba. Autor

6.2 Etapa de Análisis Estadístico

Para iniciar esta etapa, una vez capturado el tráfico de red normal, se transforma una parte como tráfico anormal a partir de la ejecución de tráfico transformado con comportamientos de malware en el entorno ICS. En esta fase preliminar del análisis estadístico, se procede a configurar una transformación de los datos para el dataset de entrenamiento ***traffic_fte_train.csv*** (Figura 14) y así generar un nuevo esquema para la creación de datos no convencionales o de situaciones anómalas dentro del tráfico de red de ICS que nos sirva de dataset de prueba ***traffic_fte_test.csv***.

El dataset es cargado en el lenguaje de programación Python[16], puesto que éste es un lenguaje de programación versátil que cuenta con una licencia de código abierto que permite la automatización de procesos en el dominio de la inteligencia artificial (AI), al igual que se destaca en aplicaciones web y big data, entre otros. Para que funcionen las características del trabajo, deberán cargarse librerías que contienen herramientas eficientes para aprendizaje automático y modelado estadístico incluyendo clasificación, regresión agrupación y reducción de dimensionalidad (sklearn), generación de gráficos a partir de listas o arrays (matploit, seaborn), manejo del cálculo numérico y análisis de datos para altos volúmenes de datos (numpy, pandas) y hasta el manejo de estadística cuando se tienen series temporales (statsmodels), entre otras, según el requerimiento que se presente.

Una vez se cargan las librerías y se carga el dataset en Python se inicia un proceso de generación del modelo, sin embargo, teniendo en cuenta que nuestro objetivo dentro del esquema de trabajo es poder generar un dataset de prueba basado en mecanismos o patrones anómalos de un dataset que solo contiene datos normales; dado que es un modelo semisupervisado, se procede a realizar lectura y comprensión de los datos para proceder con el proceso de configurar la transformación.

En el análisis exploratorio de los datos del dataset ***dataset_ics_traffic.csv*** el cual contiene datos reales del tráfico de un ICS, se entenderá desde ahora que la etiqueta "**Label**" como característica, identificará el estado de normalidad (0) o anormalidad (1) en el tráfico recolectado, sin embargo, como en la adquisición de datos no existía la característica "Label" inicialmente, ésta

[16] https://www.python.org/

ha sido incluida y configurada como normal en este dataset. Se realiza un análisis a los datos no numéricos, y después a los datos numéricos teniendo en cuenta que el dataset contiene elementos con valores NaN otros sin valores, para ello se hace una limpieza de datos, y adecuación de tipos de datos teniendo en cuenta el valor decimal (,).

En la figura 15 se puede observar estado y cantidades de las etiquetas No Numéricas que el dataset contiene; entre ellas valores en cero (0) para la etiqueta “Label” lo que garantiza que el hace parte del entrenamiento en el modelo semisupervisado, la cantidad de protocolos y servicios representativos en el tráfico y la etiqueta “lbytes” que identifica los bytes perdidos en la comunicación de las dos partes (IP origen e IP destino).

Nota: Es de aclarar que como el dataset hace parte de un sistema de control real, se reserva el derecho a la exposición de información relevante a la ciberseguridad de los activos industriales y elementos como direccionamiento y puertos quedarán en un modo incognito.

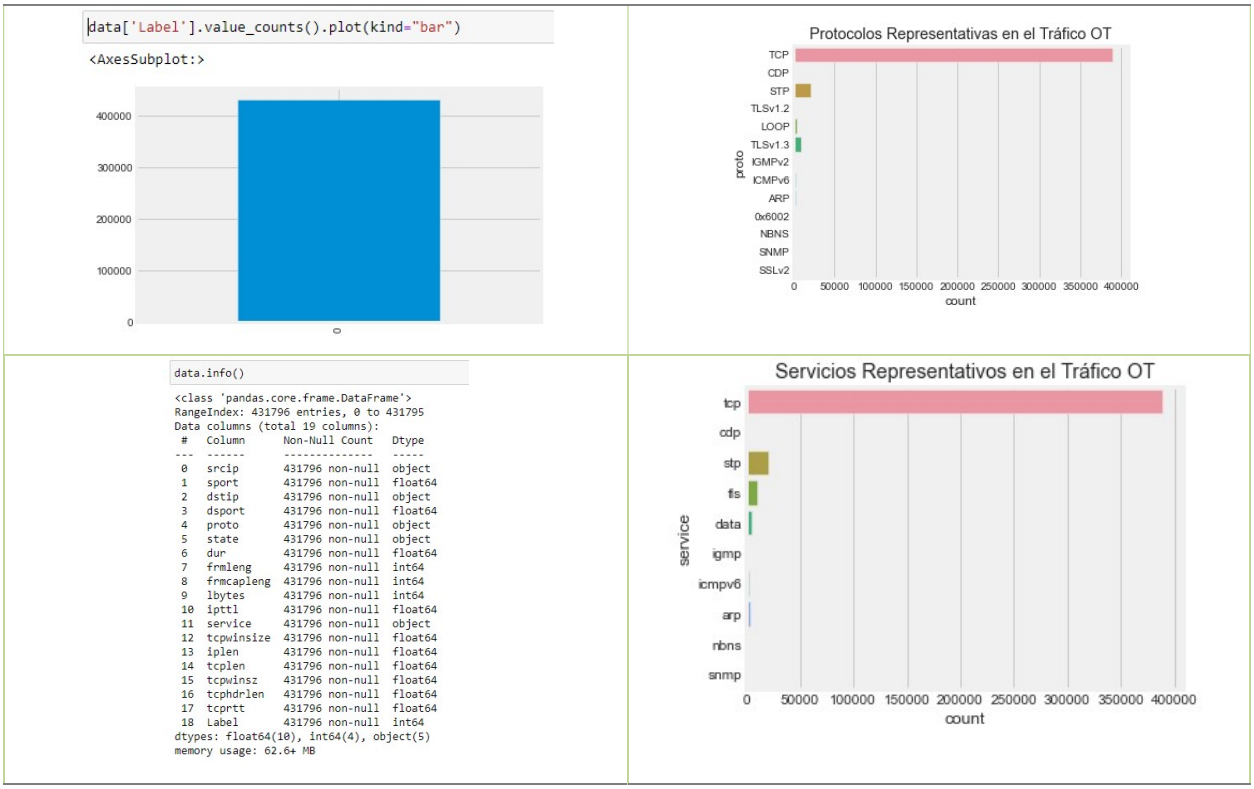


Figura 15. Análisis exploratorio de los datos No Numéricos del dataset inicial. Autor

En la figura anterior se evidencia igualmente, los tipos de datos que estarán en el trascurso del trabajo, en los cuales se desarrollarán los métodos de modelamiento. Se tomarán solo una porción de datos originales teniendo en cuenta el manejo computacional que se dará en el tiempo de procesamiento del dataset original.

El análisis de los datos de tipo Numéricos se observa en la figura 16 y representa las características específicas del tráfico en el segmento de red del sistema de control de una unidad de producción en la Gerencia Refinería de Barrancabermeja.

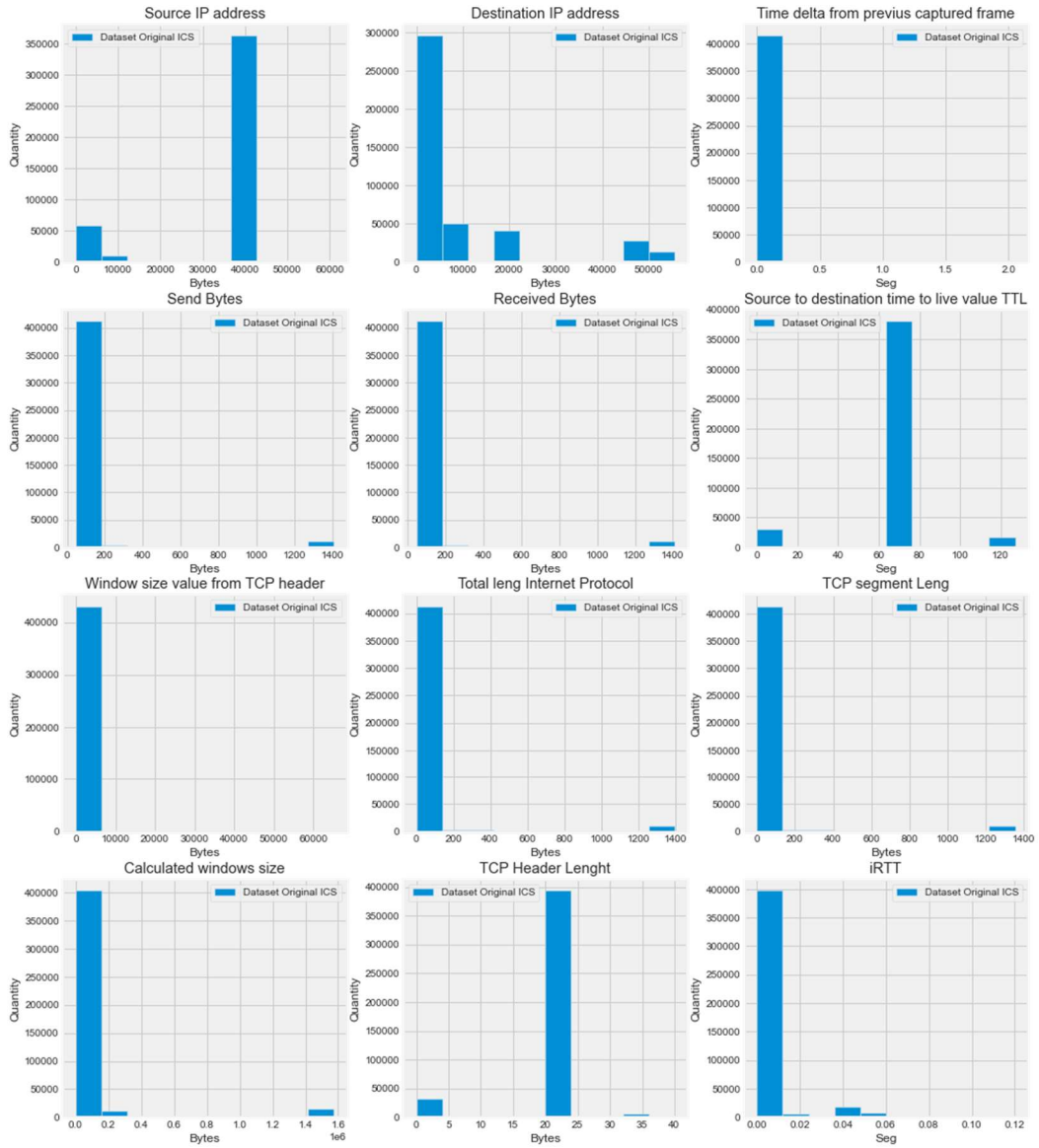


Figura 16. Análisis exploratorio de los datos numéricos del dataset inicial. Autor

La figura 16 representa el estado *normal* del tráfico de red ICS referenciado en la etiqueta "Label", desde este punto se presenta el hecho que cualquier variación del patrón o modelo de datos existentes será denominada posteriormente como un estado ***anormal*** en el dataset.

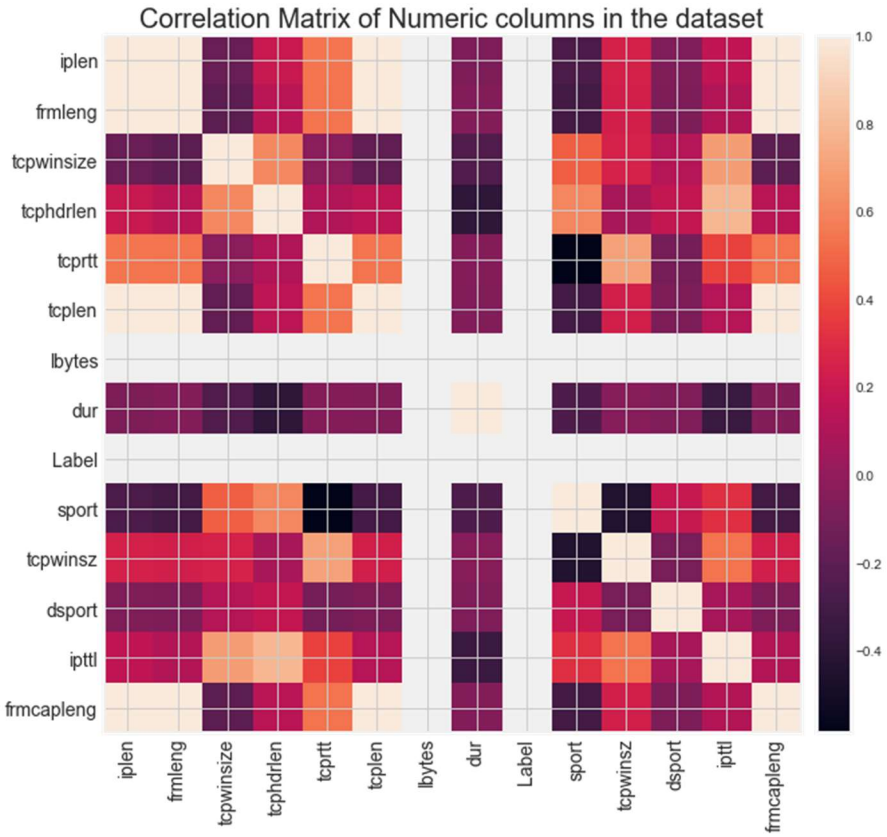


Figura 17. Matriz de Correlación entre los datos Numéricos del dataset. Autor

La figura 17 representa la matriz de correlación entre los datos numéricos del datset y se evidencia que las variables "Label" y "lbytes" contienen la mayor correlación con el resto de las variables dado que inicialmente estas etiquetas tienen el valor cero (0) por ser normales (Label) y por no presentar pérdidas de paquetes (lbytes) en las comunicaciones. En este punto se crea ***traffic_fte_train.csv***, que representa el dataset de entrenamiento.

6.2.1 Transformación Preanalítica de Datos[17]

Se define a la Transformación Preanalítica de Datos, como la ***Transformación sobre el dataset del tráfico normal de un ICS en el cual se crean patrones que representan ejemplos de situaciones anómalas en el mismo, sin afectar el entorno ICS.*** La transformación preanalítica de datos permitirá caracterizar el tráfico anómalo mediante:

- Transformación de puertos (origen /destino)
- Transformación de protocolos
- Transformación de estadísticas de flujo
- Transformación de host (origen / destino)

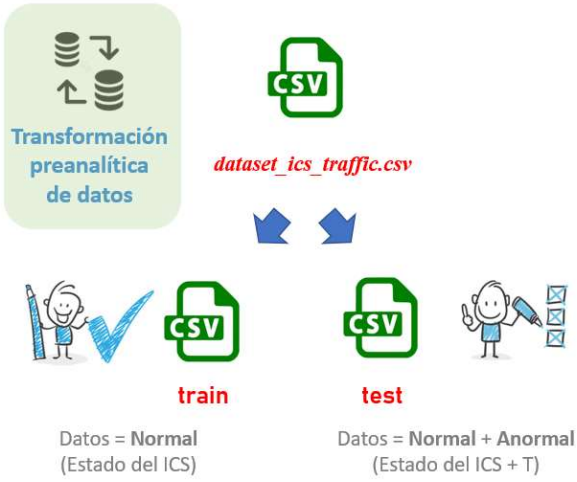


Figura 18. Esquema de la transformación preanalítica de datos

A continuación, se detalla el proceso de transformación del dataset original para generar un patrón de anomalías representativas en el tráfico ICS inicial y con el cual el modelo a desarrollar se probará. Estas pautas se encuentran orientadas al planteamiento de clasificación del tráfico de red ICS teniendo en cuenta una amenaza interna y monitoreo e identificación de tráfico sospechoso, según se explica en el apartado anterior.

[17] Definición presentada por el autor para especificar la característica de cambio entre los datasets train / test

Recomendaciones en la Transformación preanalítica de datos

- Conocer detalladamente el tráfico del **ICSuC[18] (Industrial Control System under Consideration)** sistema de control industrial en consideración, teniendo en cuenta que las variaciones del mismo originarán las condiciones anormales del experimento.
- Debe desarrollarse **un análisis preliminar de estado NORMAL** de la red ICSuC.
- Identificar segmentos de red y reglas de firewalls (si aplica) en el tráfico seleccionado para evaluar los posibles patrones de asociación básicos al estado de ANORMALIDAD a la que se va a transformar el dataset original.
- El ambiente simulado deberá ser lo más real posible, es decir, deberá contarse con **habilidades básicas de networking** para asegurar la fiabilidad de los patrones anómalos.

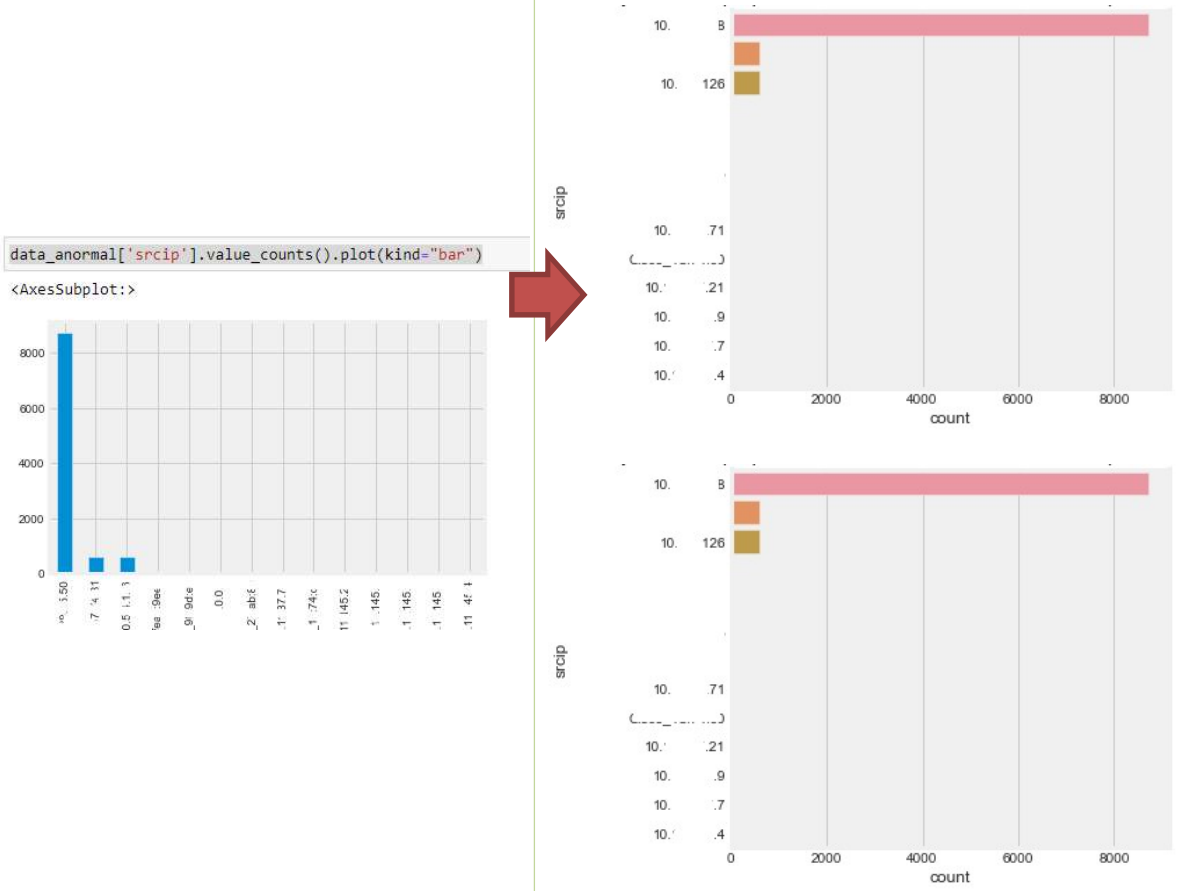
Las metodologías que permitirán caracterizar el tráfico son: la basada en el puerto, la basada en la carga útil, la basada en estadísticas de flujo y la basada en el host, relacionadas en el apartado 4.1.4 del presente documento.

- **Generación de estados Anormales en dataset original dado por "srcrip"**

Este tipo de transformación toma el direccionamiento IP fuente para enmascararlo en otro del mismo segmento de red con el fin de suplantar el primero (Figura 18). El tipo de ataque simulado permitirá detectar tráfico anormal entre un dispositivo (servidor, estación de operación u otro dispositivo electrónico con direccionamiento IP) y la red, generando un tráfico asociado posiblemente a consultas o exploración no autorizada.

En estado normal, se tiene que el equipo con direccionamiento de mayores conteos es el servidor de dominio y éste tiene su configuración especial de tráfico, protocolo y destinatarios.

[18] Adaptación al témino SuC (System under Consideration) de la ISA/IEC62443. Autor



Tráfico Normal desde servidor principal	Transformación de IP Fuente en el tráfico

Figura 19. Generación de estados Anormales en dataset original dado por "srcrip"

- **Generación de estados Anormales en dataset original dado por protocol (proto)**

La presente transformación se encuentra asociada a la suplantación del mayor tipo de protocolo existente en la muestra de estado normal, el TCP por ARP[19] (Figura 19). El protocolo ARP es muy importante para la transmisión de datos en redes Ethernet porque las tramas de datos solo se pueden enviar si se tiene un direccionamiento físico (MAC) de los equipos destino y los guarda en una tabla interna del del protocolo, en este caso el protocolo TCP no puede realizar esta tarea por sí mismo. Como el TCP no tiene la posibilidad de guardar el direccionamiento completo de la red,

[19] https://www.rfc-editor.org/rfc/rfc826

el ARP se vuelve bastante útil. Mediante esta modificación se presenta un posible vector de ataque referenciado a descubrimientos en la red, lo que ocasionaría la exposición de direccionamiento IP o la MAC del equipo y con un tratamiento profundo de esta información, la identificación de una posible vulnerabilidad ya sea de puertos o de sistema operativo, puede llevar a su explotación y posterior afectación de disponibilidad operativa.

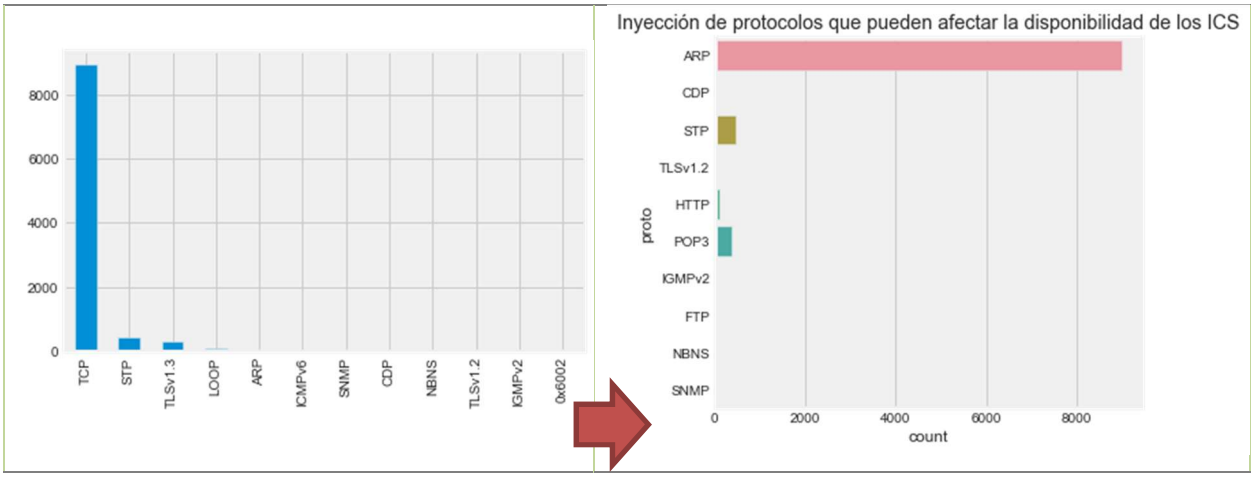


Figura 20. Generación de estados Anormales en dataset original dado por Protocolo (proto)

- **Generación de estados Anormales en dataset original dado por duration de paquetes (dur)**

Esta transformación del dataset se hace una modificación a la variable dur (Figura 20), que en definición se tiene como: *Time delta from previous captured frame (frame.time_delta)* o lo que se podría identificar como tiempo pasado desde la última trama capturada, de manera que una modificación a esto, generaría sospechas en el modelo, para el caso en particular, se darían condiciones anormales del tiempo desde la última trama mayores a 0.5 ms.

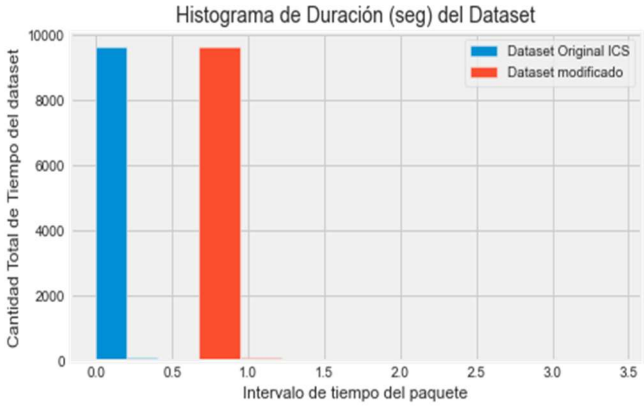


Figura 21. Generación de estados Anormales en dataset original dado por "dur"

- **Generación de estados Anormales en dataset original dado por source port (sport)**

La figura 21 representa la transformación del puerto origen teniendo en cuenta que el dataset original opera en un intervalo de puertos (1 - 1200 y 38000 – 42000), por lo que cualquier modificación a este valor, claro está no permitiendo valores mayores a 65000, puedan simular un ataque a protocolos o inclusive a la instalación y uso de backdoors[20]. Para la transformación ejecutada se adicionó un valor real de puertos a los existentes, lo que generó un corrimiento de estos en intervalos diferentes a los normales.

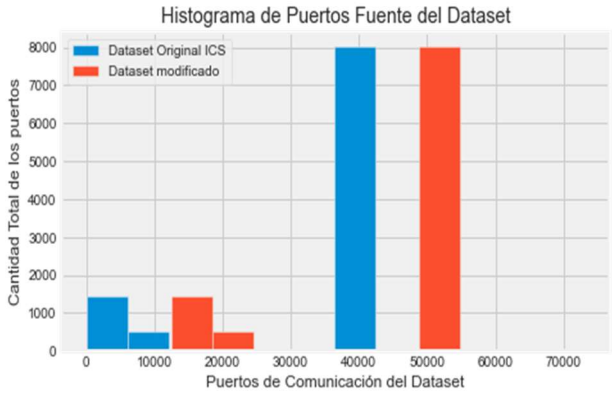


Figura 22. Generación de estados Anormales en dataset original dado por source port (sport)

[20] En la informática, una puerta trasera (en inglés, backdoor) es una secuencia especial o un término trasero dentro del código de programación, mediante la cual se pueden evitar los sistemas de seguridad del algoritmo (autenticación) para acceder al sistema. https://es.wikipedia.org/wiki/Puerta_trasera

- **Generación de estados Anormales en dataset original dado por " lbytes"**

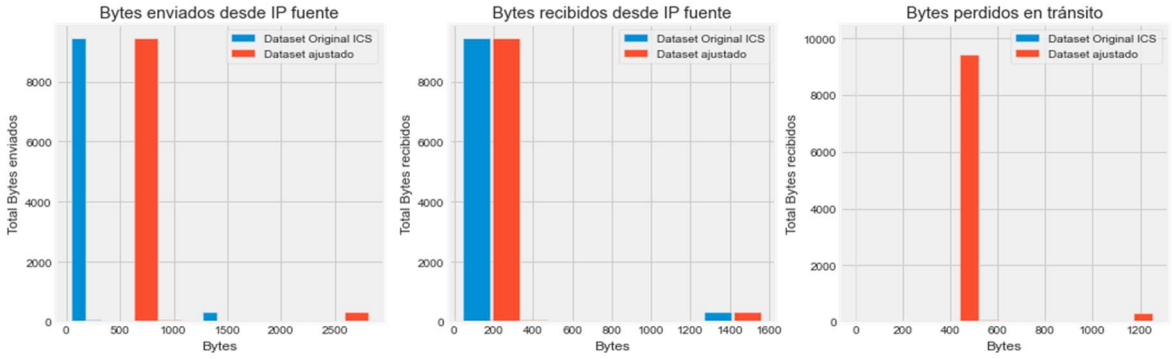


Figura 23. Generación de estados Anormales en dataset original dado por "lbytes"

La transformación de los paquetes enviados vs los recibidos de la figura 22, representa la etiqueta "lbytes" que en un estado original del dataset tiene valor cero (0) pues los bytes enviados son iguales a los recibidos. Esta transformación basada específicamente en el incremento proporcional con una desviación a los valores de cada característica (bytes en tránsito – bytes recibidos) permite representar un estado de anormalidad al tener datos perdidos mayores a cero, lo que podría representar una falla en la comunicación dado por latencias en la red, dispositivos fuente con bloqueos, pérdidas por canal de comunicación o inclusive un vector de ataque o de intento de suplantación (MITM – man in the middle).

6.2.2 Selección de ventanas de entrenamiento - prueba

Así las cosas, se crea este nuevo dataset en los cuales existen estados alterados para generar datos no convencionales o situaciones anómalas dentro del tráfico de red de ICS que nos sirva de modelo para entrenar. Para dar una referencia al modelo anormal, la columna etiquetada como "Label" en este dataset será seteada con valor uno (1), de manera que podamos identificar y referenciar los patrones anómalos en un nuevo dataset combinado.

Para dar lugar al dataset ***traffic_fte_test.csv*** de prueba, se combinan los porciones iguales de estados normales **(Label = 0)** y anormales **(Label = 1)** originados en la sección anterior. Una vez está preparado el dataset, podremos usarlo con los algoritmos de ML. Estos algoritmos usarán parte del tráfico ya etiquetado para entrenamiento ***traffic_fte_train.csv*** (el 80%) y el resto para

comprobar la tasa de acierto ***traffic_fte_test.csv*** (el 20%), la división será siempre 80-20 ya que es la forma más común de dividir los datasets. Permitirán así la construcción de un conocimiento de referencia para un tráfico "*normal*" teniendo en cuenta el normal funcionamiento del sistema de control y de tráfico "*anormal*", mediante un acondicionamiento del tráfico adquirido representando un tráfico de red con afectación orientada a una de las metodologías para el análisis de datos de un ICS.

Hasta el momento se han creado los dataset de entrenamiento y prueba y se requiere realizar un análisis al dataset combinado de los dos para evidenciar los cambios respecto al original. Para realizar esta tarea se genera un nuevo dataset con los dos esquemas, y así iniciar el análisis de datos, en la figura 23 se evidencia la combinación de los estados normales y anormales para iniciar el modelamiento de los datos totales, dados en una pequeña porción de los originales debido a la limitación en el procesamiento de los datos en el equipo donde corren los modelos. Sin embargo, para mejorar los resultados, se generaron procesos de análisis de datos en deepnote[21], obteniéndose las combinaciones de los dataset normales y anormales (Figura 24) diferentes a los referenciados en la figura 15 del dataset original.

[21] Deepnote es una aplicación web compatible con Jupyter Notebooks que provee de un entorno de trabajo interactivo para ciencia de datos a nivel de proyectos. https://deepnote.com/home

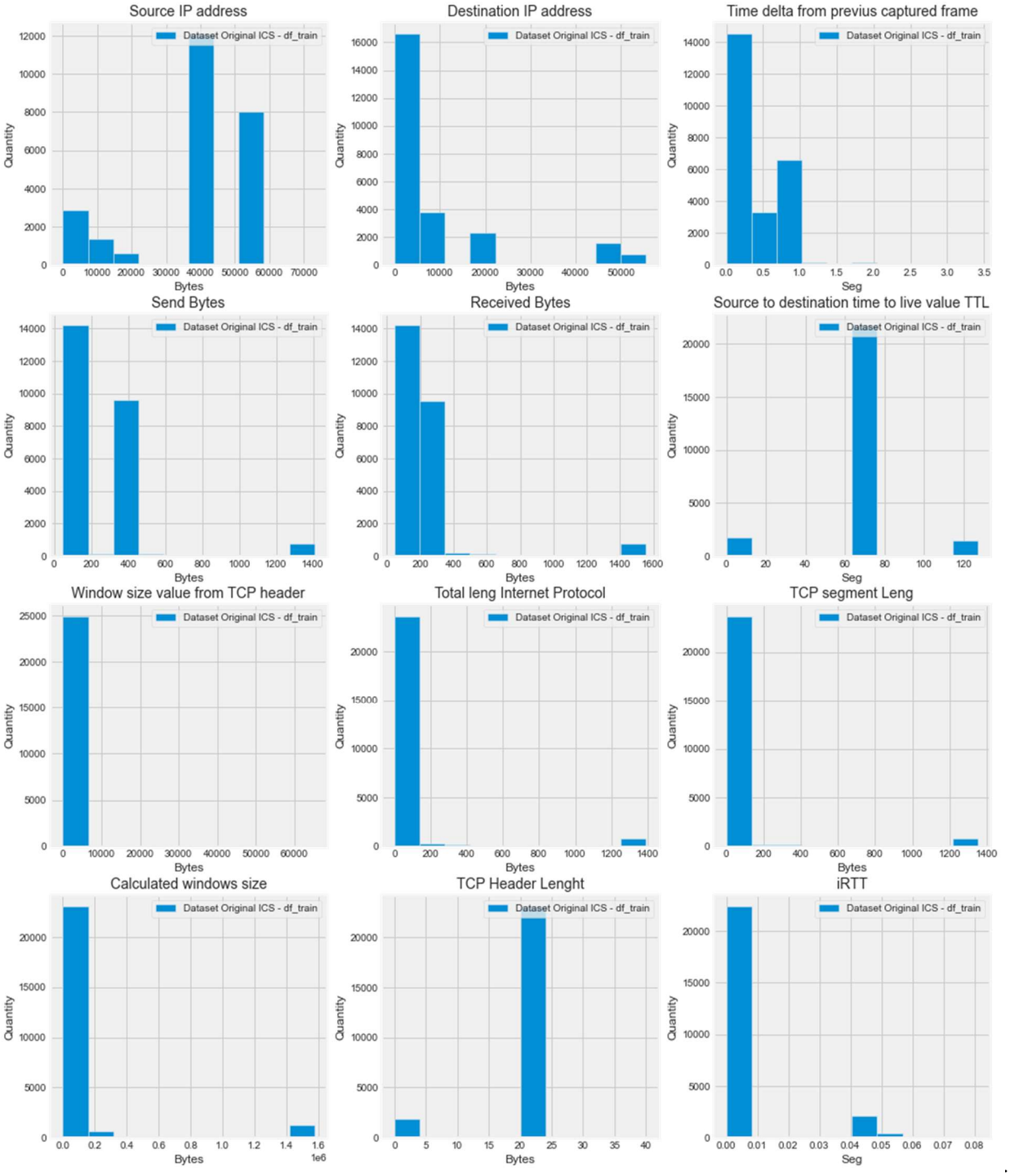


Figura 24 Combinación de dataset características normales y anormales.

En el análisis de datos Numéricos, se tienen las combinaciones en un solo dataset de las transformaciones desarrolladas en el apartado anterior al igual que una nueva matriz de correlación la cual plantea un dimensionamiento diferente a la existente y generada inicialmente.

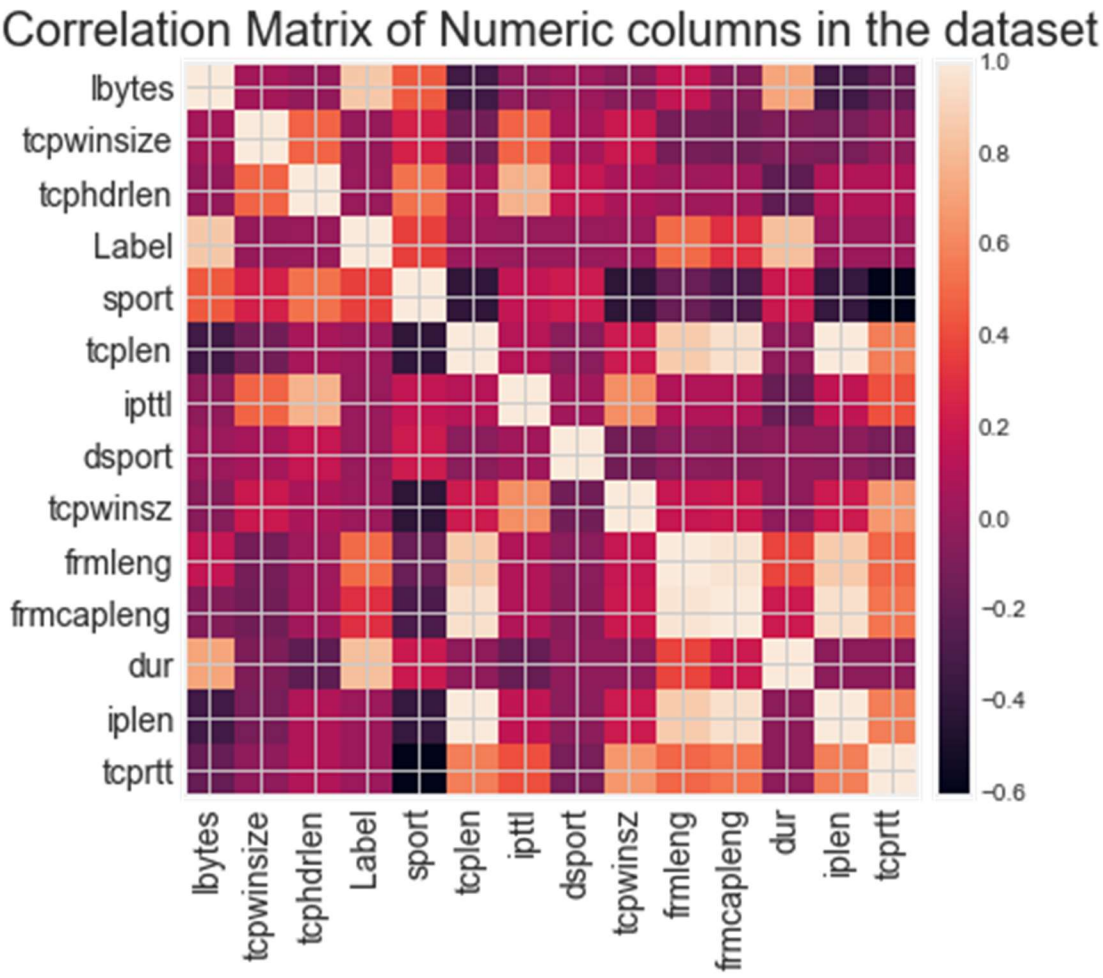


Figura 24. Matriz de correlación con dataset transformado

En esta matriz se evidencia una correlación ya no marcada en “Label” y “lbytes” ya que los datos han sido transformados para obtener anomalías dentro del mismo dataset.

Para poder trabajar con los datos totales deberá ejecutarse un paso de transformación adicional de codificación de las variables No Numéricas usando datos categóricos. Como opción

se codificarán los datos categóricos para que nuestro modelo pueda aprender mejor las características de cada una de ellas utilizando Label Encode.

Las variables categoricas en las cuales nos debemos enfocar son: **srcip** (IP Fuente) , **dstip** (IP Destino), **proto** (Protocolo), **state** (Estado de Comunicación) y **service** (Servicio). A continuación, se ajustarán los dataset de entrenamiento y prueba para poder trabajar los modelos, teniendo en cuenta lo siguiente:

- **X_train** : Datos de tráfico ICS (0: Normal)
- **y_train** : Datos de tráfico ICS (0: Normal), con las "etiquetas" de los resultados esperados de X_train
- **X_test** : Datos de tráfico ICS manipulado (0: Normal + 1: Anormal)
- **y_test** : Datos de tráfico ICS manipulado (0: Normal + 1: Anormal), con las "etiquetas" de los resultados de X_test

6.3 Etapa de modelado

El entorno de hardware del experimento de simulación es una CPU Intel i5-7200U de 2,70 GHz, 16 GB de memoria; el entorno de software es el sistema operativo Windows 10 Ultimate de 64 bits, Python3.7, Wireshark1.10.8.

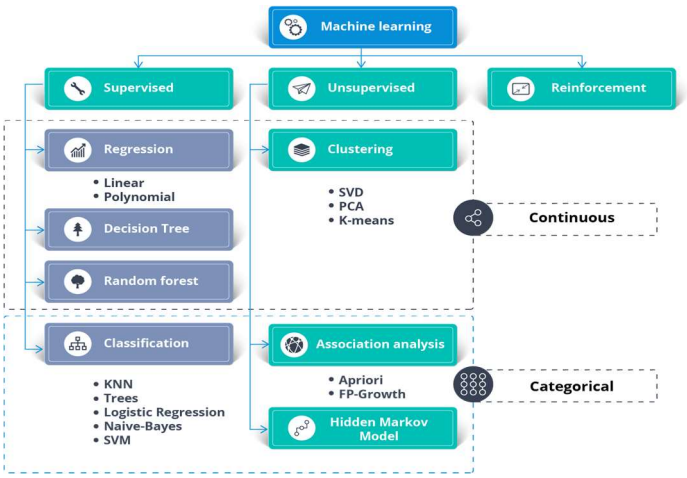


Figura 25. Tipos de algoritmos en los modelos de ML. Image credit:edureka.co

Sin embargo, el procesamiento de los datos demoró muchas de las tareas por más de 4 horas en ciertos intentos, por lo que se optó por adquirir una MV en DeepNote con características de 16 núcleos y 64 GB de RAM.

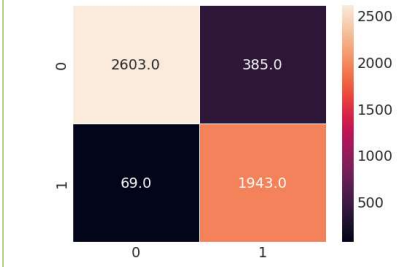
Para el presente desarrollo se usaron los modelos más asociados al aprendizaje supervisado (Figura 25), teniendo en cuenta que en la detección de anomalías no se contiene toda la información de clasificación necesaria, por lo que se desarrolla un modelo intermedio. Los siguientes fueron modelos de ML a trabajar:

- Regresión Logística
- KNN o k-Nearest Neighbors
- SVM o Support Vector Machine
- Naive Bayes
- Decision Tree
- Random Forest

6.3.1 Regresión Logística

```
Promedio Validación cruzada Regresión logística:  0.9149499999999999
Varianza de Regresión logística:  0.00963444341931593
Execution time (seg): 2.3057053089141846
```

	precision	recall	f1-score	support
0	0.97	0.87	0.92	2988
1	0.83	0.97	0.90	2012
accuracy			0.91	5000
macro avg	0.90	0.92	0.91	5000
weighted avg	0.92	0.91	0.91	5000



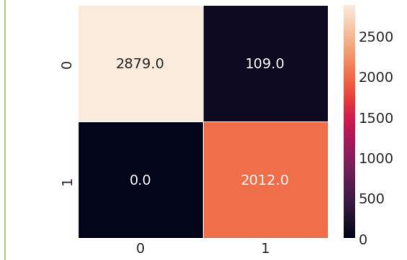
Accuracy: 90.92%
Recall: 90.92%
Precision: 91.80%
F1-Score: 91.00%
time to train: 1.80 s
time to predict: 0.00 s
total: 1.81 s

Figura 26. Matriz de Confusión - Regresión Logística

6.3.2 KNN o k-Nearest Neighbors

```
Promedio Validación cruzada KNN:  0.9988999999999999
Varianza de KNN:  0.0007348469228349661
Execution time (seg): 11.891858577728271
```

	precision	recall	f1-score	support
0	1.00	0.96	0.98	2988
1	0.95	1.00	0.97	2012
accuracy			0.98	5000
macro avg	0.97	0.98	0.98	5000
weighted avg	0.98	0.98	0.98	5000



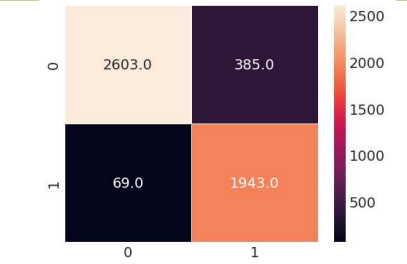
Accuracy: 99.90%
Recall: 99.90%
Precision: 99.90%
F1-Score: 99.90%
time to train: 0.00 s
time to predict: 1.97 s
total: 1.98 s

Figura 27. Matriz de Confusion - k-Nearest Neighbors

6.3.3 Support Vector Machine

```
Promedio Validación cruzada SVM:  0.9342
Varianza de SVM:  0.004802082881417157
Execution time (seg): 132.4317901134491
```

	precision	recall	f1-score	support
0	0.97	0.87	0.92	2988
1	0.83	0.97	0.90	2012
accuracy			0.91	5000
macro avg	0.90	0.92	0.91	5000
weighted avg	0.92	0.91	0.91	5000



Accuracy: 97.2%
Recall: 97.2%
Precision: 97.2%
F1-Score: 97.2%
time to train: 0.1 s
time to predict: 1.1 s
total: 1.2 s

Figura 28. Matriz de Confusión - Support Vector Machine

6.3.4 Naive Bayes

```
Promedio Validación cruzada NB:  0.9798499999999999
Varianza de NB:  0.004006557125513128
Execution time (seg): 0.09542989730834961
```

	precision	recall	f1-score	support
0	1.00	0.96	0.98	2988
1	0.95	1.00	0.97	2012
accuracy			0.98	5000
macro avg	0.97	0.98	0.98	5000
weighted avg	0.98	0.98	0.98	5000

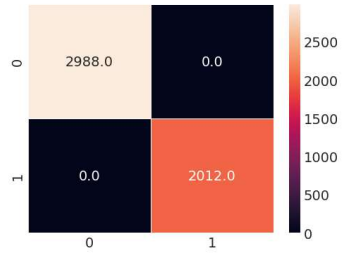
Accuracy: 98.3%
Recall: 98.3%
Precision: 98.3%
F1-Score: 98.3%
time to train: 0.08 s
time to predict: 1.6 s
total: 1.68 s

Figura 29. Matriz de Confusión - Naive Bayes

6.3.5 Decision Tree

```
Promedio Validación cruzada Decision Tree:  1.0
Varianza de Decision Tree:  0.0
Execution time (seg): 0.1506965160369873
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2988
1	1.00	1.00	1.00	2012
accuracy			1.00	5000
macro avg	1.00	1.00	1.00	5000
weighted avg	1.00	1.00	1.00	5000



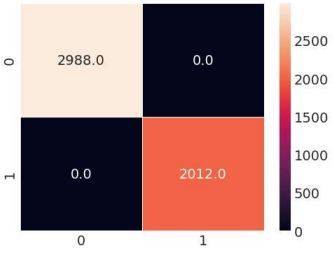
Accuracy: 100.00%
Recall: 100.00%
Precision: 100.00%
F1-Score: 100.00%
time to train: 0.05 s
time to predict: 1.4 s
total: 0.45 s

Figura 30. Matriz de Confusión - Decision Tree

6.3.6 Random Forest

```
Promedio Validación cruzada Random Forest:  1.0
Varianza de Random Forest:  0.0
Execution time (seg): 4.316094160079956
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2988
1	1.00	1.00	1.00	2012
accuracy			1.00	5000
macro avg	1.00	1.00	1.00	5000
weighted avg	1.00	1.00	1.00	5000



Accuracy: 100.00%
Recall: 100.00%
Precision: 100.00%
F1-Score: 100.00%
time to train: 0.61 s
timetopredict: 0.05 s
total: 0.65 s

Figura 31. Matriz de Confusión - Random Forest

6.3.7 Comparación entre algoritmos

El análisis de tráfico para identificación de anomalías sobre los ICS actualmente se encuentra direccionado a mecanismos o Sistemas de Identificación de Intrusos (IDS) en el que información de tráfico adquirida por sensores (dispositivos electrónicos conectados a una arquitectura industrial de manera que puedan adquirir tráfico directo) permite analizar características anómalas en estas arquitecturas. Esta tarea se encuentra basada en la verificación

de las vulnerabilidades CVE[22] (Common Vulnerabilities and Exposures) existentes sobre las firmas de seguridad, protocolos de comunicación, productos y versiones de programas, sistemas operativos y firmware existentes en los equipos interconectados a la red de control y con ello. Sin embargo; es requerido que este servicio IDS sea respaldado por proveedores y fabricantes y que mantengan las actualizaciones contantes para soportar los modelos de decisión.

EL uso de ML para el análisis de tráfico de red industrial permite para el presente trabajo de manera adicional sobre un sistema IDS, no solo el análisis de las vulnerabilidades sino también caracterizar este tráfico según sea conveniente al tipo de industria y negocio.

A continuación, se exponen los seis (LR, KNN, SVM, NB, DT, y RF) modelos trabajados se han de comparar en términos de exactitud y precisión para los datos adquiridos en el tráfico de red industrial. De la tabla 7 se tienen los resultados de la validación de todos los modelos de aprendizaje automático entrenados para el presente trabajo. La puntuación media de validación varió de 0,9092 por Regresión Logística a 1.0 por el modelo de Decision Tree al igual que Random Forest.

puesto	nombre	accuracy	recall	precisión	F1 Score	time to train	time to predic	total
1	**Decision Tree**	**1.0**	**1.0**	**1.0**	**1.0**	**0.40 s**	**0.05 s**	**0.45 s**
2	Random Forest	1.0	1.0	1.0	1.0	0.61 s	0.05 s	0.65 s
3	KNN	0.999	0.999	0.999	0.999	0.0 s	1.97	1.98
4	Naive Bayes	0.983	0.983	0.983	0.983	0.08 s	1.6	1.68
5	SVM	0.972	0.972	0.972	0.972	0.1 s	1.1	1.2
6	Regresión Logística	0.9092	0.9092	0.9180	0 – 0.92 1 – 0.90	0.0 s	1.97 s	1.98 s

Tabla 7. Comparación entre algoritmos

[22] https://www.cve.org/ Organización encaragada de identificar, definir y catalogar las vulnerabilidades de seguridad cibernética divulgadas públicamente

puesto	nombre	accuracy	recall	precisión	F1 Score	time to train	time to predic	total
1	**Decision Tree**	**1.0**	**1.0**	**1.0**	**1.0**	**0.22 s**	**0.11 s**	**0.31 s**
2	Random Forest	1.0	1.0	1.0	1.0	0.74 s	0.22 s	1,4 s
3	KNN	1.0	1.0	1.0	1.0	0.86 s	0.74 s	2,7 s
4	Naive Bayes	0.993	0.993	0.993	0.993	0.14 s	1.9	2.7
5	SVM	0.98	0.98	0.98	0.98	0.17 s	1.8	1.7
6	Regresión Logística	0.96	0.96	0.96	0.96	0.4s	2,4 s	1.7 s

Tabla 8. Comparación entre algoritmos

Según estos resultados, se puede decir que el algoritmo de clasificación Decisión Tree tiene el mejor rendimiento en el procesamiento de conjuntos de datos para la detección de anomalías del tráfico ICS analizado. La medida F1 para los modelos descritos determinan el rendimiento de la clasificación en términos de precisión y recuperación, se obtienen modelos con una caracterización de 100%, entre ellos **Decision Tree** al igual que **Random Forest**, como se esperaba resultado del párrafo anterior.

6.4 Etapa de pruebas

Por último, se evalúa la capacidad predictiva del modelo final empleando el conjunto de pruebas. A continuación, se describen los resultados del procedimiento de validación cruzada como método estándar para estimar el rendimiento de un algoritmo o configuración de aprendizaje automático en un conjunto de datos. Para nuestro caso en particular los modelos se corrieron con un número de K = 5, 10 y 15, los resultados se pueden evidenciar en las tablas 8, 9 y 10:

puesto	nombre	validación[23] cruzada	varianza	execu. Time (s)
1	**Decision Tree**	**1.0**	**0.0**	**0.21**
2	Random Forest	1.0	0.0	10.48
3	KNN	0.99879	0.00048	19.69
4	Naive Bayes	0.9794	0.00075	0.46
5	SVM	0.9341	0.00347	72.72
6	Regresión Logística	0.94525	0.01772	6.12

Tabla 9. Comparación validación cruzada k=5

puesto	nombre	validación cruzada	varianza	execu. Time (s)
1	**Decision Tree**	**1.0**	**0.0**	**0.15**
2	Random Forest	1.0	0.0	4.31
3	KNN	0.99889	0.00073	11.89
4	Naive Bayes	0.97984	0.00400	0.095
5	SVM	0.9342	0.00480	132.043
6	Regresión Logística	0.91494	0.00963	2.30

Tabla 10. Comparación validación cruzada k=10

puesto	nombre	validacion cruzada	varianza	execu. Time (s)
1	**Decision Tree**	**1.0**	**0.0**	**0.69**
2	Random Forest	1.0	0.0	18,28
3	KNN	0.9990	0.0005	22.04
4	Naive Bayes	0.97969	0.00491	0.24
5	SVM	0.9342	0.0056	272.50
6	Regresión Logística	0.92955	0.0281	24.82

Tabla 11. Comparación validación cruzada k=15

En las tablas 8, 9 y 10 se tiene que los mejores modelos de la validación cruzada son el **Decisión Tree** y el **Random Forest** con valor de 100% y tiempo de ejecución entre 0,21 y 0,61 para el primero y entre 4,31 y 18,28 para el segundo. Los modelos no cambian de lugar con la variación de la validación en K, sin embargo, se observa que computacionalmente, el modelo SVM tiene mayor tiempo y con ello consumo de máquina.

[23] Representa el ratio entre el número de instancias que el modelo ha predicho correctamente, frente al número total de instancias del dataset, multiplicado por 100 para dar un resultado porcentual.

Capítulo 7
EVALUACIÓN DEL MODELO EN ESCENARIO CONTROLADO ICS

Objetivo (Objetivo Específico #3)

- Evaluar el modelo en un escenario controlado similar al tomado en el dataset del proyecto mediante el cual se despliega en un ICS un ataque propio de las redes de control.

Arquitectura escenario

El escenario de prueba del modelo implementado en el presente trabajo es el extrusor **EX2201** de la Unidad de Polietileno I de la Gerencia Refinería de Barrancabermeja. El EX-2201 es un extrusor de alimentación caliente con un tornillo de 12" de diámetro y 163.2" de longitud de la rosca, accionado por un motor de velocidad variable de 400 H.P. y 480 voltios, corriente continua. El movimiento se transmite al tornillo por medio de un reductor de engranajes. La cabeza está compuesta del sistema de cuchillas accionadas por un motor eléctrico que se puede separar del cuerpo principal o barril abriendo hidráulicamente las abrazaderas de ajuste que mantiene unidos los dos cuerpos.[24]

Figura 33. Extrusor EX2201 Planta de Polietileno I

[24] MANUAL DE OPERACIONES PLANTA DE POLIETILENO I, 1160-02-90-P51-TEC-002

El extrusor tiene configurado instrumentación y sistema de control basado en un PLC Allen Bradley Logix5561 el cual controla el sistema en su totalidad. Se interconecta con la red del sistema de control en campo con un switch PhoenixContact, por donde se realizó la conexión de un equipo que simularía una condición anormal con el cual se tomaría tráfico para probar el modelo. Figura 32.

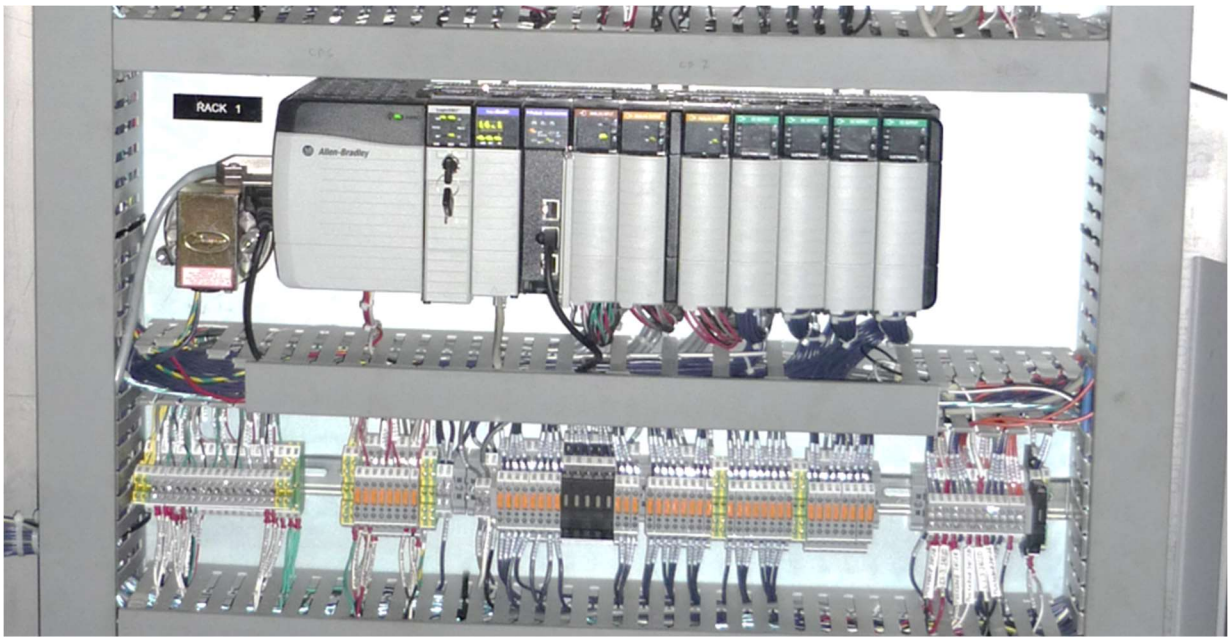


Figura 32. Activo Industrial (PLC) asignado al Extrusor EX2201

Toma de datos en estado Normal

Para dar prueba al modelo seleccionado, se tomó tráfico de red en el switche en el cual la arquitectura de campo se encontraba conectada. Para el estado normal (Fig. 33, izquierda) se evidencia en el software WireShark que existe únicamente el protocolo CIP I/O en el cual se controlan cada una de las variables de campo.

Para el caso de la inyección de tráfico anómalo, se generó una conexión sobre la arquitectura existente y un equipo que simulara la condición anormal (Fig. 33, derecha), de manera que, desarrollando las siguientes tareas en las comunicaciones desde equipo, modificaran el estado inicial “normal”:

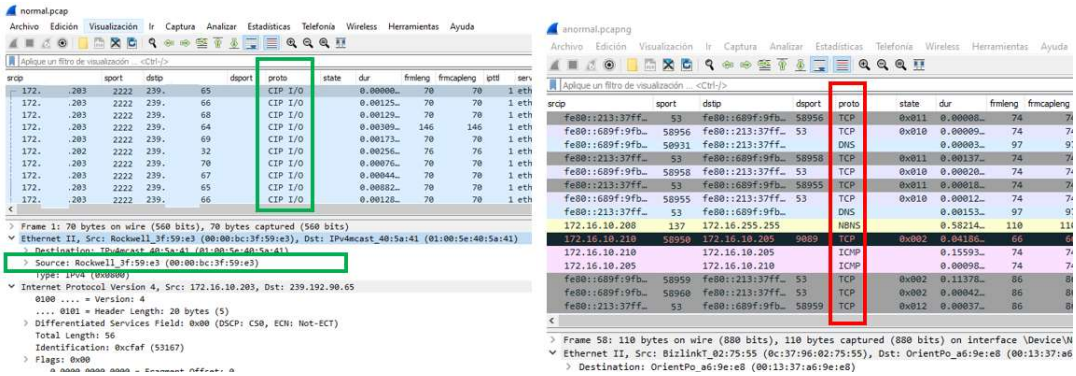


Figura 33. Tráfico del PLC objetivo analizado con software WireShark

- Ping 172.X.X.201 desde el equipo atacante.
- Protocolo RDP por intento de acceso remoto desde equipo atacante.
- Intento de clonación estación HMI mediante VMWare vCenter desde equipo atacante.
- Solicitudes al DHCP de la red de control del PLC (No existente*).*

puesto	nombre	accuracy	recall	precisión	F1 Score	time to train	time to predic	total
1	**Decision Tree**	**1.0**	**1.0**	**1.0**	**1.0**	**0.22 s**	**0.11 s**	**0.31 s**
2	Random Forest	1.0	1.0	1.0	1.0	0.74 s	0.22 s	1,4 s
3	KNN	1.0	1.0	1.0	1.0	0.86 s	0.74 s	2,7 s
4	Naive Bayes	0.993	0.993	0.993	0.993	0.14 s	1.9	2.7
5	SVM	0.98	0.98	0.98	0.98	0.17 s	1.8	1.7
6	Regresión Logística	0.96	0.96	0.96	0.96	0.4s	2,4 s	1.7 s

Tabla 12. Resultados para los modelos desarrollados en escenario controlado ICS.

Como resultado de la intervención atacante o acción por parte del equipo atacante, se tienen que el algoritmo de ***Decision Tree***, presenta los mejores resultados al igual que ***Random Forest***, con la diferencia del tiempo de entrenamiento, situación semejante a lo evidenciado en las pruebas a tráfico adquirido en el estudio preliminar del presente trabajo. Tabla 12.

Capítulo 8
CONCLUSIONES Y RECOMENDACIONES

En el presente trabajo se han tomado puntos de referencia para poder enfocar el análisis de anomalías de tráfico direccionado a un ambiente industrial no simulado que aborda todos los problemas propios de la naturaleza del activo. Esta implementación considera que la confiabilidad y disponibilidad de las comunicaciones en el sistema de control industrial son de vital importancia, así que teniendo en cuenta las posibles formas de afectar equipos interconectados en una red de control se implementó un modelo de ML de tipo semisupervisado para realizar identificación anomalías desarrolladas a partir de vectores de ataque asociados al tráfico de red real de una unidad operativa. En el desarrollo del modelo se logró clasificar de manera confiable las anomalías del sistema con bajas tasas de falsos positivos, por ello que se puede concluir que los modelos de ML hacen parte de un enfoque viable para proporcionar un apoyo de decisión confiable en la identificación de posibles afectaciones internas o externas sobre ciberseguridad de la infraestructura crítica de una organización.

La modalidad de la clasificación semisupervisada planteada en el presente trabajo, desarrolló la caracterización del tráfico de red industrial para generar un esquema de transformaciones a ciertas características del dataset que representaba tráfico de control industrial real, a partir de la ejecución de tráfico transformado con comportamientos de malware en el entorno ICS. Esto hace necesario realizar un análisis preliminar de los datos considerados como *patrones normales* para después modificarlos como *anómalos* e inyectarlos en el modelo para su entrenamiento y posterior prueba, por ello que se puede decir que la precisión y confiabilidad de cada modelo de detección de anomalías se basa en la calidad de los datos de entrenamiento. Los resultados experimentales muestran que el método propuesto tiene una mayor precisión de detección y una menor tasa de falsas alarmas. Por ello, es de gran importancia que quien desarrolle este tipo de modelos sea personal idóneo con conocimiento en infraestructura de redes y que conozca con un nivel de detalle el objetivo que está siendo analizado.

Con el desarrollo del presente trabajo, se construyó un modelo de detección de anomalías para tráfico ICS el cual contempló seis (LR, KNN, SVM, NB, DT, y RF) de los principales algoritmos del ML, iniciando con la preparación de datos, el preprocesamiento, el diseño de entrada del modelo y la arquitectura del modelo para así evaluar el rendimiento y realizando un estudio comparativo entre ellos, para lo cual se observó que los modelos de aprendizaje automático Decisión Tree y Random Forest pueden abordar el análisis de los flujos de tráfico de la red en cantidad y calidad de los datos con mayor eficiencia y precisión. Esto proporcionó como ganancia adicional la posibilidad reconocer no solo tráfico anómalo sobre un ICS sino la posibilidad de integrar en cierta forma el modelo a un esquema SOC (Security Operation Center) corporativo, con el fin de dimensionar una excelente infraestructura de ciberseguridad operativa que permitiera con un significado práctico y valor de aplicación, la seguridad de sus activos industriales.

REFERENCIAS

[1] V. Atluri and J. Horne, "A Machine Learning based Threat Intelligence Framework for Industrial Control System Network Traffic Indicators of Compromise," SoutheastCon 2021, 2021, pp. 1-5, doi: 10.1109/SoutheastCon45413.2021.9401809.

[2] J. M. Beaver, R. C. Borges-Hink and M. A. Buckner, "An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications," 2013 12th International Conference on Machine Learning and Applications, 2013, pp. 54-59, doi: 10.1109/ICMLA.2013.105.

[3] H. Lan, X. Zhu, J. Sun and S. Li, "Traffic Data Classification to Detect Man-in-the-Middle Attacks in Industrial Control System," 2019 6th International Conference on Dependable Systems and Their Applications (DSA), 2020, pp. 430-434, doi: 10.1109/DSA.2019.00067.

[4] S. M. Rachmawati, D. -S. Kim and J. -M. Lee, "Machine Learning Algorithm in Network Traffic Classification," 2021 International Conference on Information and Communication Technology Convergence (ICTC), 2021, pp. 1010-1013, doi: 10.1109/ICTC52510.2021.9620746.

[5] S. P. Khedkar and R. AroulCanessane, "Machine Learning Model for classification of IoT Network Traffic," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 166-170, doi: 10.1109/I-SMAC49090.2020.9243468.

[6] H. Li and S. Qin, "Optimization and implementation of industrial control system network intrusion detection by telemetry analysis," 2017 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 1251-1254, doi: 10.1109/CompComm.2017.8322743.

[7] H. Singh, "Performance Analysis of Unsupervised Machine Learning Techniques for Network Traffic Classification," 2015 Fifth International Conference on Advanced Computing & Communication Technologies, 2015, pp. 401-404, doi: 10.1109/ACCT.2015.54.

[8] Chi-Ho Tsang and S. Kwong, "Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction," 2005 IEEE International Conference on Industrial Technology, 2005, pp. 51-56, doi: 10.1109/ICIT.2005.1600609.

[9] E. D. Knapp, J.T. Langill, “Industrial Network Security, Securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control Systems”, ISBN: 978-0-12-420114-9.

[10] W. Chen, T. Liu, Y. Tang, D. Xu, “Multi-level adaptive coupled method for industrial control networks safety based on machine learning”, Safety Science, Volume 120, 2019, Pages 268-275, ISSN 0925-7535, https://doi.org/10.1016/j.ssci.2019.07.012.

[11] E. Anthi, L. Williams, M. Rhode, P. Burnap, A.Wedgbury, “Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems”, Journal of Information Security and Applications, Volume 58, 2021, 102717, ISSN 2214-2126, https://doi.org/10.1016/j.jisa.2020.102717.

[12] J. Pei, K. Zhong, M. Ahmad Jan, J. Li, “Personalized federated learning framework for network traffic anomaly detection”, Computer Networks, Volume 209, 2022, 108906, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2022.108906.

[13] A. Shahraki, M. Abbasi, A. Taherkordi, A.Delia Jurcut, “A comparative study on online machine learning techniques for network traffic streams analysis”, Computer Networks, Volume 207, 2022, 108836, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2022.108836.

[14] J. Vávra, M. Hromada, L. Lukáš, J.Dworzecki, “Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment, International Journal of Critical Infrastructure Protection”, Volume 34, 2021, 100446, ISSN 1874-5482, https://doi.org/10.1016/j.ijcip.2021.100446.

[15] M. A. Umer, K. N. Junejo, M. T. Jilani, A. P. Mathur, “Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations”, International Journal of Critical Infrastructure Protection, Volume 38, 2022, 100516, ISSN 1874-5482,

https://doi.org/10.1016/j.ijcip.2022.100516.

[16] I. Chakraborty, B. M. Kelley, B. Gallagher, "Industrial control system device classification using network traffic features and neural network embeddings", Array, Volume 12, 2021, 100081, ISSN 2590-0056, https://doi.org/10.1016/j.array.2021.100081.

[17] Yask & B. Suresh Kumar (2019), "A review of model on malware detection and protection for the distributed control systems (Industrial control systems) in oil & gas sectors", Journal of Discrete Mathematical Sciences and Cryptography, 22:4, 531-540, DOI: 10.1080/09720529.2019.1642623

[18] J. F. Brenner (2013), "Eyes wide shut: The growing threat of cyber-attacks on industrial control systems", Bulletin of the Atomic Scientists, 69:5, 15-20, DOI: 10.1177/0096340213501372

[19] C. Bronk & E. Tikk-Ringas (2013), "The Cyber Attack on Saudi Aramco", Survival, 55:2, 81-96, DOI: 10.1080/00396338.2013.784468.

[20] S. Bagui, X. Fang, E. Kalaimannan, S.C. Bagui & Joseph Sheehan (2017), "Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features", Journal of Cyber Security Technology, 1:2, 108-126, DOI: 10.1080/23742917.2017.1321891

[21] M. Safari, E. Parvinnia, A. K. Haddad, "Industrial intrusion detection based on the behavior of rotating machine", International Journal of Critical Infrastructure Protection, Volume 34, 2021, 100424, ISSN 1874-5482, https://doi.org/10.1016/j.ijcip.2021.100424.

[22] P. Ackerman, "Industrial Cybersecurity, Efficiently secure critical infrastructure systems", Published by Packt Publishing Ltd, ISBN 978-1-78839-515-1.

[23] J. McCarthy, E Division, D Faatz, "Securing the Industrial Internet of Things: Cybersecurity for Distributed Energy Resources", NIST SPECIAL PUBLICATION 1800-32, National Institute of Standards and Technology, https://www.nccoe.nist.gov/iiot