Implementación de un modelo de aprendizaje automático para el laboratorio de análisis de vulnerabilidades en el CERT Académico de la ESPE

Ponce Almachi, John Francisco y Villarreal Campaña, Luigi Damián

Departamento de Ciencias de la Computación

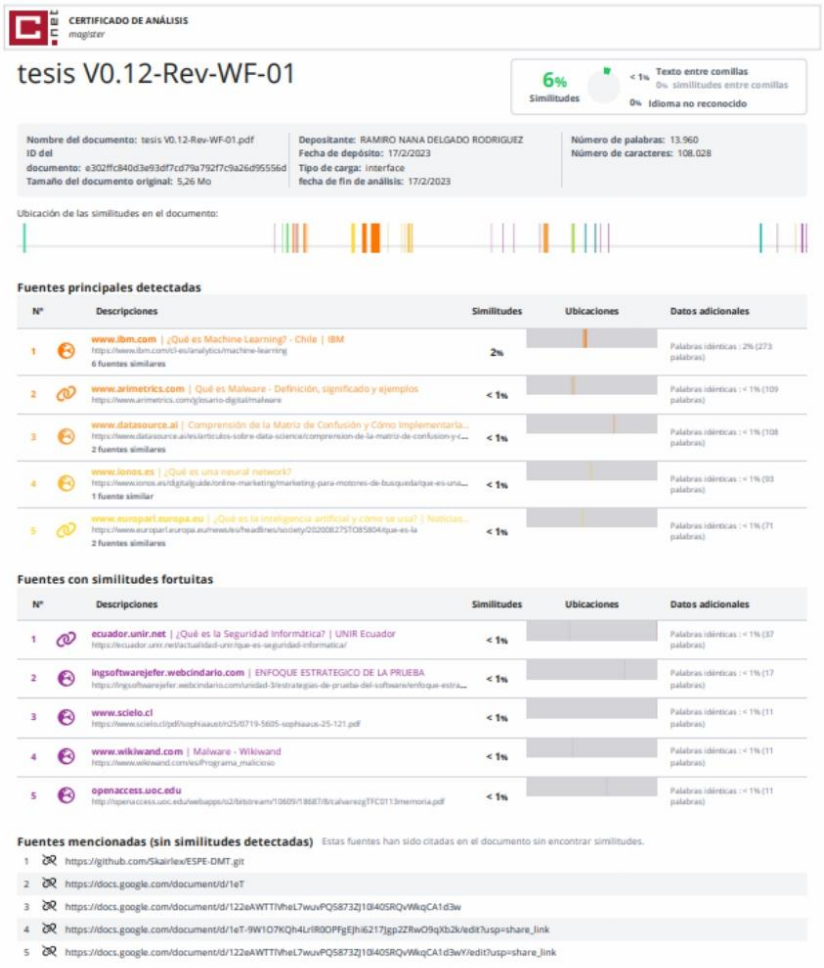
Carrera de Ingeniería de Sistemas e Informática

Trabajo de titulación, previo a la obtención del título de Ingeniero en Sistemas e Informática

Ing. Fuertes Díaz, Walter Marcelo PhD

15 de febrero del 2023

Reporte de Similitud de Contenidos



CERTIFICADO DE ANÁLISIS
magister

tesis V0.12-Rev-WF-01

6% Similitudes
< 1% Texto entre comillas
0% similitudes entre comillas
0% Idioma no reconocido

Nombre del documento: tesis V0.12-Rev-WF-01.pdf
ID del documento: e302ffc840d3e93df7cd79a792f7c9a26d95556d
Tamaño del documento original: 5,26 Mo
Depositante: RAMIRO NANA DELGADO RODRIGUEZ
Fecha de depósito: 17/2/2023
Tipo de carga: interface
fecha de fin de análisis: 17/2/2023
Número de palabras: 13.960
Número de caracteres: 108.028

Ubicación de las similitudes en el documento:

Fuentes principales detectadas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	www.ibm.com | ¿Qué es Machine Learning? - Chile | IBM https://www.ibm.com/cl-es/analytics/machine-learning 6 fuentes similares	2%		Palabras idénticas : 2% (273 palabras)
2	www.arimetrics.com | Qué es Malware - Definición, significado y ejemplos https://www.arimetrics.com/glosario-digital/malware	< 1%		Palabras idénticas : < 1% (109 palabras)
3	www.datasource.ai | Comprensión de la Matriz de Confusión y Cómo Implementarla... https://www.datasource.ai/es/articulos-sobre-data-science/comprension-de-la-matriz-de-confusion-y-c... 2 fuentes similares	< 1%		Palabras idénticas : < 1% (108 palabras)
4	www.ionos.es | ¿Qué es una neural network? https://www.ionos.es/digitalguide/online-marketing/marketing-para-motores-de-busqueda/que-es-una-... 1 fuente similar	< 1%		Palabras idénticas : < 1% (93 palabras)
5	www.europarl.europa.eu | ¿Qué es la inteligencia artificial y cómo se usa? | Noticias... https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la 2 fuentes similares	< 1%		Palabras idénticas : < 1% (71 palabras)

Fuentes con similitudes fortuitas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	ecuador.unir.net | ¿Qué es la Seguridad Informática? | UNIR Ecuador https://ecuador.unir.net/actualidad-unir/que-es-seguridad-informatica/	< 1%		Palabras idénticas : < 1% (37 palabras)
2	ingsoftwarejefer.webcindario.com | ENFOQUE ESTRATEGICO DE LA PRUEBA https://ingsoftwarejefer.webcindario.com/unidad-3/estrategias-de-prueba-del-software/enfoque-estra...	< 1%		Palabras idénticas : < 1% (17 palabras)
3	www.scielo.cl https://www.scielo.cl/pdf/sophiaaust/n25/0719-5605-sophiaaus-25-121.pdf	< 1%		Palabras idénticas : < 1% (11 palabras)
4	www.wikiwand.com | Malware - Wikiwand https://www.wikiwand.com/es/Programa_malicioso	< 1%		Palabras idénticas : < 1% (11 palabras)
5	openaccess.uoc.edu http://openaccess.uoc.edu/webapps/o2/bitstream/10609/18687/8/ralvarezgTFC0113memoria.pdf	< 1%		Palabras idénticas : < 1% (11 palabras)

Fuentes mencionadas (sin similitudes detectadas) Estas fuentes han sido citadas en el documento sin encontrar similitudes.

1 https://github.com/Skairlex/ESPE-DMT.git
2 https://docs.google.com/document/d/1eT
3 https://docs.google.com/document/d/122eAWTTlVheL7wuvPQ5873ZJ10l40SRQvWkqCA1d3w
4 https://docs.google.com/document/d/1eT-9W1O7KQh4LrlR0OPFgEJhi6217JgpZZRwO9qXb2k/edit?usp=share_link
5 https://docs.google.com/document/d/122eAWTTlVheL7wuvPQ5873ZJ10l40SRQvWkqCA1d3wY/edit?usp=share_link

Firma:

Firmado electrónicamente por:
WALTER MARCELO
FUERTES DIAZ

.....

Ing. Fuertes Díaz, Walter Marcelo, PhD.

Director

Departamento de Ciencias de la Computación

Carrera de Ingeniería de Sistemas e Informática

Certificación

Certifico que el trabajo de titulación: **"Implementación de un modelo de aprendizaje automático para el laboratorio de análisis de vulnerabilidades en el CERT Académico de la ESPE"** fue realizado por los señores **Ponce Almachi, John Francisco** y **Villarreal Campaña, Luigi Damián**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE. Además, fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 06 de marzo del 2023

Firma:

.....

Ing. Fuertes Díaz, Walter Marcelo, PhD.

C. C. 1707017701

Departamento de Ciencias de la Computación

Carrera de Ingeniería en Sistemas e Informática

Responsabilidad de Autoría

Nosotros, **Ponce Almachi, John Francisco y Villarreal Campaña, Luigi Damián** con cédulas de ciudadanía 1723428304 y 1750142109 respectivamente, declaramos que el contenido, ideas, y criterios del trabajo de titulación: "Análisis e implementación de un modelo de aprendizaje automático con el fin de detectar tráfico malicioso, la el laboratorio de análisis de vulnerabilidades del CERT Académico de la ESPE", es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricas, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 16 de febrero del 2023

Firma

Ponce Almachi, John Francisco

C.C: 1723428304

Firma

Villarreal Campaña, Luigi Damián

C.C: 1750142109

Departamento de Ciencias de la Computación

Carrera de Ingeniería en Sistemas e Informática

Autorización de Publicación

Nosotros, **Ponce Almachi, John Francisco y Villarreal Campaña, Luigi Damián** con cédulas de ciudadanía 1723428304 y 1750142109 respectivamente, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: "Análisis e implementación de un modelo de aprendizaje automático con el fin de detectar tráfico malicioso, para el laboratorio de análisis de vulnerabilidades del CERT Académico de la ESPE", en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 16 de febrero del 2023

Firma

Ponce Almachi, John Francisco

C.C: 1723428304

Firma

Villarreal Campaña, Luigi Damián

C.C: 1750142109

Dedicatorias

A Dios quien me dio la oportunidad de estudiar en nuestra privilegiada Institución.

A mis padres Francisco Ponce y Patricia Almachi, por brindarme su apoyo e inspiración para mantenerme firme en el transcurso de mi carrera profesional. Por ser padres ejemplares y colocar en mi corazón ese sustento fundamental de valores y principios en los cuales me baso cada día para vivir.

A mis compañeros por todos los buenos momentos de esfuerzo y risas que pudimos compartir.

John Francisco Ponce Almachi

A Dios quien me dio la oportunidad de estudiar en esta prestigiosa Institución.

A mis padres Luis Alberto Villarreal Cando y Gloria Cecilia Campaña Velasco, por brindarme su apoyo logrando cumplir uno de mis objetivos más importantes como es un título profesional, esperando que se sientan muy orgullosos de mí, como yo me siento muy afortunado de que estén siempre a mi lado día tras día, son los mejores padres del mundo.

También agradezco a mis compañeros que me han acompañado a lo largo de este trayecto pudiendo estrechar lazos muy fuertes de compañerismo con cada uno de ustedes.

Luigi Damián Villarreal Campaña

Agradecimientos

Agradezco a Dios por su sabiduría y conocimiento que me ha brindado en todo este proceso educativo.

Agradezco a mi familia por su apoyo incondicional, por todo ese tiempo y sabios consejos que pude aprovechar. Gracias por ser mi sustento económico que siempre es necesario y muchas veces difícil de conseguir, espero poder compensar algún día todo su esfuerzo brindándoles el orgullo que se merecen.

Agradezco a todo docente que nos entregó su tiempo y experiencias para formarnos como buenos profesionales, además de todo consejo de buen corazón que me impulsó a seguir en este camino y aún buscar nuevas metas.

John Francisco Ponce Almachi

Agradezco a Dios por brindarme sabiduría y fuerza en mis momentos de vulnerabilidad, para cumplir con este proceso educativo.

Agradezco a mi familia que me han apoyado tanto económica y moralmente en todo momento, para poder finalizar mis estudios sin decaer en el camino

Finalmente, agradezco a todos esos docentes que nos instruyeron a lo largo de la carrera, brindándonos todo ese conocimiento y formando bases sólidas para podernos defender en la vida como buenos profesionales, especialmente a nuestro tutor Ing. Walter Fuertes quien no solo fue nuestro tutor sino un gran amigo que, con su vasta experiencia y paciencia, nos ayudó a poder completar este proyecto de manera exitosa. De igual manera agradecer a nuestro director de carrera Ing. Fernando Galarraga por toda su ayuda con los trámites correspondientes a este proceso de titulación.

Luigi Damián Villarreal Campaña

Tabla de contenidos

Índice de tablas

Índice de Figuras

Resumen

En la actualidad el consumo de la tecnología sigue en crecimiento gracias a la demanda de recursos tecnológicos que ayudan acelerar procesos. De igual manera el uso de software amplía las posibilidades de encontrar vulnerabilidades en los sistemas informáticos. Por tanto, las exigencias en la seguridad de la información es cada vez trascendental aplicarlo en todo ámbito de trabajo, al ser las redes informáticas un punto crítico a analizar, debido a que toda la información fluye a través de las mismas.

Las redes informáticas por lo general tienen diversos mecanismos de seguridad como firewalls, IDS, IPS, entre otros, los cuales implementan diversas técnicas para el análisis de vulnerabilidades. La seguridad de las redes puede mejorar gracias a la aplicación de aprendizaje automático que mejora la detección de amenazas y las referidas vulnerabilidades.

La presente tesis tiene como objetivo desarrollar algoritmos de aprendizaje automático para el análisis y detección de tráfico malicioso, que incluye toda actividad inusual en el flujo de paquetes de datos, tales como los ataques DOS, XSS y ataques de fuerza bruta. Para lograrlo, se aplicó la metodología ágil SCRUM, que permitió generar un artefacto de software que comprende la arquitectura de cuatro capas funcionales con algoritmos de inteligencia artificial para la detección de tráfico malicioso en una red. Adicionalmente, se realizó una evaluación comparativa de algoritmos de machine learning y el desarrollo evolutivo referido. Como resultado, se entrega una solución práctica al CERT académico, que busca generar nuevo conocimiento que permita incrementar los niveles de seguridad cibernética de la universidad y del país.

Palabras clave: tráfico malicioso, aprendizaje automático, vulnerabilidades, seguridad informática, CERT

Abstract

Nowadays, technology consumption continues to grow thanks to the demand for technological resources that help speed up processes. In the same way, the use of software increases the possibility of finding vulnerabilities in computer systems. Therefore, the demands on information security are increasingly transcendental to apply in all areas of work, computer networks being a critical point to analyze because all information flows through them.

Computer networks generally have various security mechanisms, such as firewalls, IDS, and IPS, which implement different vulnerability analysis techniques. Network security can improve thanks to the application of automatic learning that enhances the detection of threats and the vulnerabilities mentioned above.

This thesis aims to develop automatic learning algorithms for the analysis and detection of malicious traffic, including any unusual activity in the flow of data packets, such as DOS attacks, XSS, and brute force attacks. The agile SCRUM methodology was applied, allowing the generation of a software artifact that includes the architecture of four functional layers with artificial intelligence algorithms for detecting malicious traffic in a network. Additionally, a comparative evaluation of machine learning algorithms and the referred evolutionary development was carried out. As a result, a practical solution is delivered to the academic CERT, which seeks to generate new knowledge that increases the levels of cyber security of the university and the country.

Keywords: malicious traffic, machine learning, vulnerabilities, computer security, CERT**.**

Capítulo I

Introducción

Antecedentes

Con el avance tecnológico y la exigencia del teletrabajo y tele-educación, la cantidad de ataques cibernéticos en el mundo se han incrementado. Es así que tan solo en el año 2016 ESET compañía de seguridad informática establecida en Bratislava, informó que alrededor 30% en las empresas grandes reportan problemas de vulnerabilidades y 30% en empresas pequeñas (CHANG, 2020). Esta cifra continúa creciendo debido a la gran digitalización impulsada por la pandemia de Covid-19 lo que conlleva a un mundo cibernético más preparado para el manejo de la información.

Una propuesta de solución a esta necesidad ha sido la implementación de un equipo especializado en seguridad informática, los cuales son responsables de prevenir, identificar y responder los incidentes de seguridad informática que se lo denomina CERT.

Para su desarrollo y funcionamiento han surgido distintos foros y organizaciones que coordinan a los diferentes CERTs mundialmente, comparten información sobre amenazas, vulnerabilidades y ataques a nivel global y divulgar medidas tecnológicas que mitiguen el riesgo de ataques a sistemas y usuarios conectados a Internet entre ellos el observatorio de Seguridad de la Información [OSI] (Guacho Morocho, 2014), el cual es un referente mundial al servicio de los ciudadanos, empresas y administraciones para describir, analizar, asesorar y difundir la cultura de la seguridad y la confianza en la Sociedad de la Información.

Las tendencias gubernamentales de los países más avanzados en materia de gestión de la seguridad y lucha contra la delincuencia y el terrorismo, corroborado por entidades supranacionales como la Agencia Europea de Seguridad de las Redes de la Información (ENISA), la Comisión de la Unión Europea, la Unión Internacional de Telecomunicaciones

(UIT)5 o la OTAN6, por citar algunos, apuntan a la creación de organizaciones altamente especializadas, diseñadas con el fin de garantizar la seguridad de los sistemas y redes de información de una nación de los que depende el correcto funcionamiento de la propia sociedad.

En muchos países, los CSIRT surgieron por primera vez como parte de la academia y como redes de investigación, y no en el gobierno ni en las empresas privadas, por lo que su participación es vital para el desarrollo de nuevas soluciones en el campo de la seguridad de la información y lucha contra la delincuencia, el ciberespionaje, la ciberguerra y el terrorismo cibernético (Vicente & Rafael, 2020). En vista de su importancia y dado que forma parte del dominio académico e investigativo de nuestra Universidad el presente proyecto pretende mejorar el alcance que tiene un CERT, desde el ámbito académico e investigativo, para colocar en funcionamiento un CERT-Académico, impulsando una nueva generación de CERTs. En concreto, este proyecto, tiene como propósito diseñar e implementar un modelo de aprendizaje automático para el CERT Académico, de la Universidad de las Fuerzas Armadas ESPE (ESPE-CERT).

Formulación del problema

En la actualidad las amenazas cibernéticas se encuentran en constante crecimiento y evolución. Esto se debe a una amplia necesidad de la población para adaptarse a un ambiente tecnológico que motiva a desarrollar plataformas de fácil acceso y que requieren un alto nivel de seguridad (Medina, 2022). Pese a los altos estándares que se usa, toda persona u organización puede ser víctima de un ataque cibernético lo que ha motivado la institución de diversos organismos que ayudan a resguardar la integridad de la información que se maneja a nivel mundial tales como: la Agencia Europea de Seguridad de las Redes de la Información (ENISA), la Comisión de la Unión Europea, la Unión Internacional de Telecomunicaciones (UIT)5 y la OTAN.

Con la ayuda de estas organizaciones a nivel nacional, cada país puede contar con un marco de gestión general, lo que no siempre se apega a una realidad nacional. Un ejemplo de esto fue en abril de 2019 cuando diversos medios informáticos, nacionales fueron víctimas de un ataque cibernético que registró alrededor de 40.000.000 incidencias en diferentes instituciones privadas y públicas, debido a la decisión del gobierno ecuatoriano que acordó retirar el asilo diplomático al Fundador de WikiLeaks Julián Assange (Telecomunicaciones, 2019). Debido a las incidencias ocurridas a nivel nacional, se ha visto necesaria la constitución de una estructura organizacional para todo tipo de empresa que brinde un sistema de ciberseguridad, que cuente con el equipamiento y capacidades adecuadas para enfrentar diferentes tipos de amenazas, ante cualquier situación referente a ataques informáticos. (Aguirre Ponce, 2017).

A nivel institucional se puede encontrar que no existe el recurso humano y tecnológico para dar respuesta ante incidentes informáticos adecuado a los procesos llevados por la universidad, lo que genera una brecha de seguridad en el manejo de la información. De ello se derivan otros problemas:

1) Carencia de software especializado para servicios reactivos y preventivos para un CERT académico, en ocasiones por la falta de presupuesto.
2) Inexistencia de software especializado propio para el análisis de vulnerabilidades que de soporte a procesos de investigación.
3) Disminución del nivel de seguridad al interior de la ESPE en relación a los ataques cibernéticos.
4) Inexistencia de herramientas especializadas para optimizar los procesos de enseñanza aprendizaje y el entrenamiento especializado al interior de la ESPE y del país.
5) Excesiva dependencia con el software comercial desarrollados por el extranjero.

Justificación

Las vulnerabilidades son cualquier tipo de fallo ya sea de software o hardware que compromete la integridad, disponibilidad o confidencialidad de la información. Estos fallos favorecen a los ciber-delincuentes que pretenden quebrantar el orden de la sociedad, debido a que muchas de las empresas que están al servicio público, consiguen ser afectadas, como son el servicio agua potable, eléctrico, telecomunicaciones, salud, educativo, etc. Por tanto, existe, la necesidad de crear directivas comunes a escala nacional e internacional, con el fin de mitigar los riesgos y amenazas provenientes de la red. Por este motivo, los centros educativos e investigativos como las universidades, se encuentran en el deber de desarrollar soluciones que contrarresten estos delitos tales como el espionaje cibernético, la guerra digital y el ciber terrorismo. La ITU, NIST, OTAN recomiendan la formación de equipos especializados, en la investigación, innovación y el desarrollo tecnológico, para garantizar la seguridad de las personas y organizaciones (Aguirre Ponce, 2017).

En vista de la relevancia que tiene la seguridad de la información alrededor del mundo, y dado que la universidad forma parte del entorno académico e investigativo, el presente trabajo pretende mejorar el alcance que tiene un CERT, que tiene como objetivo optimizar el software disponible para el uso del mismo, además de aportar con un nuevo conocimiento de innovación y desarrollo tecnológico desde la ESPE, a través de investigación aplicada multidisciplinaria, en la solución de problemas reales, actuales, con garantía de calidad técnica e investigativa. Lo que impulsa una nueva generación en las herramientas de desarrollo e investigación para los CERTs. Este proyecto es de gran importancia por su impacto en la sociedad, puesto que es un problema contemporáneo, delicado y es de responsabilidad de FF.AA. En concreto, este proyecto, tiene como propósito diseñar e implementar un modelo de aprendizaje automático de análisis de vulnerabilidades para el uso del CERT Académico, de la Universidad de las Fuerzas Armadas ESPE (ESPE-CERT).

Objetivos

Objetivo General

Diseñar e implementar un modelo de aprendizaje automático para el análisis de tráfico malicioso, y su implementación en el laboratorio de análisis de vulnerabilidades del ESPE-CERT mediante la creación de scripts ejecutables y sus pruebas funcionales.

Objetivos Específicos

a) Evaluar los métodos, técnicas y herramientas de machine learning para el análisis de tráfico malicioso y la creación de scripts ejecutables para este propósito.
b) Diseñar e implementar un modelo de aprendizaje automático para el análisis de tráfico malicioso, a partir de las fases de la metodología ágil SCRUM en el paradigma de Prototipado o framework iterativo.
c) Realizar las pruebas de concepto funcionales y no funcionales del modelo implementado, y colocar en funcionamiento el prototipo en el CERT Académico de la ESPE.
d) Documentar los manuales de operación, técnico y de usuario y difundir sus resultados.

Alcance

El siguiente proyecto, comprende el análisis, desarrollo e implementación de un modelo de aprendizaje automático para el análisis de tráfico malicioso, que comprende el uso de técnicas machine learning, para buscar comportamientos maliciosos como, spam, trafico fraudulento y ataques DoS, lo que dará como resultado un artefacto instalable, para el laboratorio de análisis de vulnerabilidades del CERT académico, el cual provee a este proyecto con los accesos a los logs, de los firewall Fortinet (Forti analyzer).

Este estudio nace en base al proyecto de investigación del CERT académico de la ESPE, el cual proveerá la infraestructura física y tecnológica para su progreso.

Para delinear de forma adecuada el alcance de la investigación planteada, se proponen varias preguntas de investigación asociadas a los objetivos específicos que se muestran en la Tabla 1.

Tabla 1

Preguntas de investigación

Objetivos específicos	Preguntas de investigación
Evaluar los métodos, técnicas y herramientas de machine learning para el análisis de tráfico malicioso y la creación de scripts ejecutables para este propósito	• RQ1- ¿Cuáles son los métodos, técnicas y herramientas más adecuados para generar un modelo de aprendizaje automático enfocado en la detección de vulnerabilidades?
Diseñar e implementar un modelo de aprendizaje automático para el análisis de tráfico malicioso, a partir de las fases de la metodología ágil SCRUM en el paradigma de Prototipado o framework iterativo.	• RQ2- ¿Qué arquitectura debe tener un prototipo de software para análisis de vulnerabilidades que se acople a las necesidades CERT académico de la ESPE?

Objetivos específicos	Preguntas de investigación
Realizar las pruebas de concepto funcionales y no funcionales del modelo implementado, y poner en funcionamiento su prototipo en el CERT Académico de la ESPE.	• RQ3- ¿Cuáles son las pruebas funcionales y no funcionales que se debe hacer en el desarrollo de software orientado al análisis de vulnerabilidades?
Documentar los manuales de operación, técnico y de usuario y difundir sus resultados	• RQ4 ¿Qué puntos se debe tomar en cuenta para realizar el manual técnico y de usuario?

Hipótesis

Un modelo de aprendizaje automático para análisis de vulnerabilidades de tráfico malicioso incrementa la detección de ataques cibernéticos.

Capitulo II

Marco teórico

Red de Categorías

Con el fin de buscar la pertinencia en la fundamentación teórica de la presente investigación, conviene estructurar una red de las principales categorías que intervienen en la explicación y comprensión científica del tema objeto de estudio; dicha red se muestra en las figuras 1 y 2

Figura 1

Modelo de aprendizaje automático

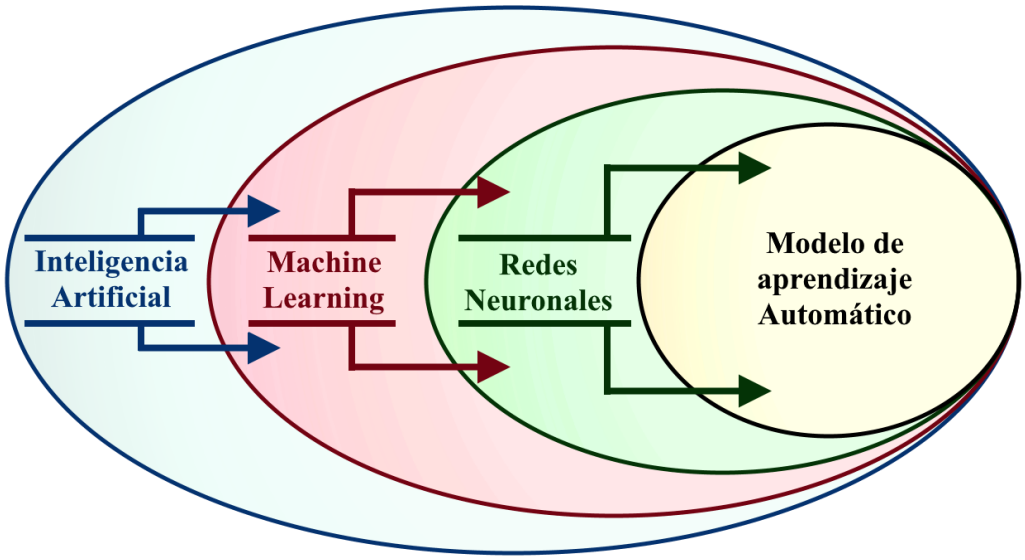
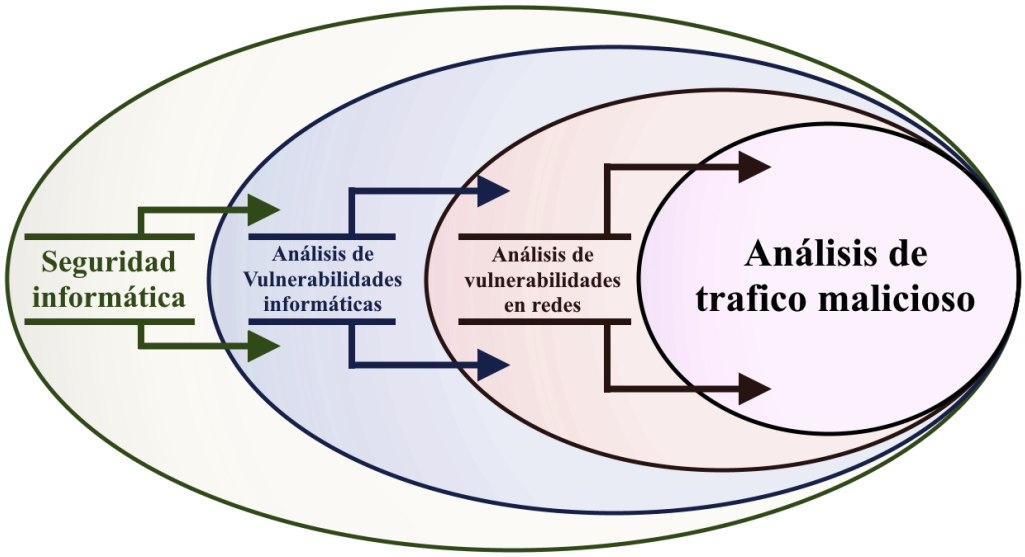


Figura *2*

Análisis de tráfico malicioso



Seguridad Informática

Según (Unir, ¿Qué es la seguridad informática y cuáles son sus tipos?, 2021) se refiere al proceso de prevenir y detectar el uso no autorizado en un sistema informático, al eliminar vulnerabilidades para preservar el software de computadoras, servidores, dispositivos móviles, redes, bases de datos y sistemas electrónicos para evitar la manipulación o procesamientos de malwares informáticos, que busca garantizar la protección anti amenazas de equipos tecnológicos y bases de datos

En concordancia con (Martínez, 2023) el modus operandi del Malware para acceder a los dispositivos y equipos a través de la red ha llevado a que la Seguridad Informática se enfoque en muchas áreas de protección entre ellas el Hardware. Software y Seguridad de red. La mayoría de ataques son debidos a la herramienta llamada internet genera fallos a gran escala en los sistemas de Seguridad informática, representa un riesgo latente que muchas veces implican pérdidas millonarias. Al buscar proteger las redes, equipos, información sensible

de una empresa al identificar y eliminar amenazas que pueden difundirse en la red. Busca minimizar el mantenimiento de la infraestructura, mejorar su seguridad en todos los niveles.

Seguridad informática de hardware

Enfocada al hardware como su nombre lo dice protege el equipo de intrusiones no deseadas a través de firewalls, servidores proxy que buscan controlar el tráfico de red, así como HSM "Hardware Security Module" que trabaja a través de claves criptográficas para la autentificación en los sistemas al identificar errores de seguridad en los equipos. Al iniciar con la configuración o código de ejecución, dispositivos de entrada y salida de datos.

Seguridad informática de software

Según (Medina, 2022), los ataques de software son cada vez más frecuentes al punto que aprovechan cualquier agujero de seguridad informática para entrar y realizar el robo de información, por lo que los fabricantes evitan tener errores mínimos desde el proceso de desarrollo, como defectos de diseño, desbordamiento de buffer, fallos de implementación que pueden abrir la puerta a virus o hackers. La seguridad informática de software, busca proteger las aplicaciones, bases de datos y programas posibles de amenazas exteriores a través de cortafuegos, filtros antispam, antivirus, software de filtro de contenido, entre otros.

Malware

En concordancia con (Belcic, 2019), el malware abreviación de Malicious software, se refiere a cualquier tipo de software malicioso diseñado por ciberdelincuentes o tradicionalmente llamados hackers, que afecta negativamente, perjudican el funcionamiento de un computador, teléfono móvil o cualquier otro dispositivo. Los ejemplos más comunes de malware, son el robo de información, dañar o causar un mal funcionamiento del sistema informático, perjuicios económicos, chantaje a propietarios de datos de sistemas informáticos, acceso de usuarios no autorizados, provocar molestias o una combinación de varias de estas actividades con objetivo de dañar al host ejecutor.

Normalmente el malware tratara de tomar el control de las funciones de un dispositivo con el objetivo de obtener información o sacar dinero al usuario de forma ilícita. Entre los muchos efectos negativos que produce el malware en los equipos se encuentran (arimetrics, 2022):

a) Captación de datos desde el navegador web.
b) Instalación de programas, herramientas, extensiones y/o complementos no deseados.
c) Pérdida de espacio en el disco duro.
d) Ralentización del dispositivo.
e) Oleadas de adware (publicidad no deseada).
f) El sistema se bloquea constantemente
g) Cambios no autorizados en la configuración del dispositivo.

Tipos de Malware

Virus. Se adhiere a un programa como puede ser de uso común el cual, al ejecutarse, produce la multiplicación del mismo al infectar a otros programas con sus bits de código malicioso (Belcic, 2019).

Adware. Consiste en la aparición de anuncios no deseados en el navegador del dispositivo o ventanas emergentes con anuncios (Belcic, 2019).

Troyano. Se hace pasar por la apariencia de un programa o extensión útil para el usuario, se instala en el sistema para robar información importante o instalar otros virus (Belcic, 2019).

Gusanos. Es muy similar a los virus comunes, se expanden por el sistema, lo cual destruye la información y archivos del equipo (Belcic, 2019).

Spyware. Se introduce ilícitamente en el dispositivo y extrae información que comunica al creador del malware para futuros chantajes o mal uso de esta (Belcic, 2019).

Rootkit. Permanece oculto en el sistema después de su ejecución, mientras goza de privilegios de administrador lo que permite el acceso de una atacante al software infectado (Belcic, 2019).

CERT

La facilidad de acceso a internet a través de dispositivos tecnológicos ha causado el rápido avance tecnológico experimentado durante los últimos años ha generado competitividad en la creación de servicios que mejoren la interacción entre cliente empresa. La integración de sistemas acelerados y procesos de desarrollo donde las infraestructuras de TI a nivel mundial han tenido que afrontar grandes dificultades a nivel seguridad.

Los ataques de seguridad cada vez son más abundantes y alarmantes como lo respaldan las estadísticas durante los últimos años, motivo suficiente por el cual las

organizaciones se han enfocado en salvaguardad sus activos informáticos y para solventar la protección de estas se han construido los diferentes CERT.

Definición

CERT es un equipo de profesionales en seguridad TI con capacitad para responder ante un incidente ataque de seguridad con el soporte necesario para este. EL trabajo del CERT o integrantes de este tienen como objetivo menorar el riesgo de las organizaciones que estén bajo su servicio a través de capacidad en seguridad TI, estos buscaran identificar deficiencias y alertar de estas tanto con software y en hardware (Cabezón, 2023).

El CERT cuya abreviación de sus siglas en ingles significa “Computer Emergency Response Team”, es registrado en Estado Unidos por el CERT Coordination Center (CERT/CC), por lo que en la actualidad se encuentran varios nombres que han sido de uso para identificación de estos, algunos ejemplos de estos tenemos: Computer Incident, Computer Security Incident Response Team (CSIRT), Security Emergency Response Team (SERT, Incident Response Team (IRT), Response Team (CIRT).

Tipos de CERT

El CERT para dar un mejor rendimiento debe familiarizarse con el entorno de desarrollo y enfocar el mejor servicio posible conforme a las necesitades identificadas al solventar los objetivos del negocio u organización por lo que la clasificación de los CERT es:

1) CERT Nacional
2) CERT Militar
3) CERT de Soporte
4) CERT Académico
5) CERT Gubernamental
6) CERT Comercial

7) CERT Interno
8) CERT para PYMES “Pequeña y mediana empresa”

Los CERTS/CSIRT certificados alrededor del mundo que fomentan la comunicación y apoyo ante los ataques de Seguridad según FIRST (Forum for Incident Response and Security Teams) presentes en Ecuador son:

1) CERT Radical
2) CSIRT-EPN
3) BLUE HAT CERT
4) EcuCERT
5) CSIRT-CEDIA
6) MAINTLATAM CSIRT
7) CSIRT Telconet

Servicios de un CERT

Los CERT ofrecen servicios según los objetivos y el entorno empresarial que contrato los servicios con el deber de mantener y cumplir las metas del Cliente. Por ende, estos se dividen en 3 categorías descritas a continuación:

Servicios reactivos. Enfocado en dar respuesta de manera rápida a las exigencias de asistencia al tratar de manera activa los incidentes alertados, siento estos alertados por monitoreo o por personal contratado. Para el sustento en todas las áreas de este servicio es necesario el aprendizaje continuo del CERT por lo que se actualizara por medio de los informes y análisis pos mortem, de cada incidente de seguridad mejorando y dando una mejor calidad de servicio por lo que este ofrece algunos subservicios pequeños aparte del general nombrados a continuación (CERT Ecuador/CC, 2023):

1) Manejo de artefactos
2) Manejo de vulnerabilidades
3) Alertar y advertencias
4) Tratamiento de incidentes

Servicios proactivos. Con el objetivo de evitar impactos negativos de las empresas este servicio está diseñado para detectar, informar y actuar con prevención las deficiencias antes del posible ataque de seguridad. Otro de sus objetivos principales es el compartir información a la comunidad de informáticos en servicio para su protección (CCN-CERT, 2023). Los sub servicios ofrecidos por este son:

1) Configuración y mantenimiento de la seguridad
2) Intercambio de inteligencia sobre amenazas
3) Desarrollo de herramientas
4) Detección de intrusiones
5) Auditorias de seguridad

Servicios de Gestión de la Calidad de la Seguridad:

Es el resultado del continuo aprendizaje a través de la información recolectada de los servicios nombrados anteriormente enfocado en la continua mejora de la seguridad de la empresa u organización para la capacitación y evolución de este (incibe-cert, 2023). Los sub servicios que ofrece son:

- Sensibilización
- Análisis de riesgos, continuidad del negocio y recuperación tras un desastre
- Evaluación o certificación de producto
- Consultoría de seguridad

Inteligencia Artificial

La inteligencia artificial o IA, es la imitación de procesos de inteligencia humana a través de la creación y aplicaciones de la computación a base de algoritmos, en otras palabras, la IA es el intento de que los dispositivos o equipos tecnológicos actúen, piensen y tomen decisiones como lo hace la humanidad (NETAPP, 2022). El ser humano desde el siglo 1 A.C. ya se planteó la posibilidad de crear maquinas o equipos con un cerebro parecido al del ser humano, sin embargo, no fue hasta 1955 en que de la mano de John McCarthy acuño el termino IA o inteligencia artificial. En 1956 McCarthy junto a otros se unieron en una conferencia de aprendizaje profundo, predictivo y descriptivo al crear la “Ciencia de los Datos”. No obstante, para poder conseguir dicha IA, son necesarios tres componentes.

1) Sistemas computacionales
2) Datos y gestión de los mismos
3) Algoritmos de IA avanzados (código)

La Inteligencia Artificial, depende de la velocidad de procesamiento para poder tener un mayor parecido al comportamiento humano que se quiera conseguir, por lo que se es necesario que este tenga mayor cantidad de recursos para el proceso (NETAPP, 2022).

La IA permite que los sistemas tecnológicos se relacionen en su entorno, solucionen problemas y tomen decisiones lógicas. Los equipos reciben datos preparados o recopilados a través de sus periféricos. Poseen capacidades de ajuste en sus mecanismos, analizan acciones anteriores y operan de manera autónoma (Europeo, 2020).

Machine Learning

Es la capacidad de aprendizaje de los equipos, está presente muchas veces alrededor de nuestro día a día. Los ejemplos más evidentes del machine learning se los puede identificar en las plataformas de streaming que recomiendan al usuario contenido de acuerdo al gusto de otros usuarios, asimismo puede reconocer en el reconocimiento de voz en asistentes virtuales o la toma de decisión de los automóviles autónomo al escoger el camino por la carretera. En otras palabras, el Machine learning, es una forma de la IA que permite a los sistemas aprender de los datos y no de una programación explícita (Kirsch, 2018).

Las técnicas procedentes del machine learning son muy necesarias para mejorar la precisión predictiva de esta. Esta según el problema empresarial posible se enfocará conforme a la base de datos obtenida o programada y se atenderá según el tipo y volumen de los datos recopilados. Esta se ha divido en las siguientes categorías:

Aprendizaje supervisado. Inicia con un conjunto de registros establecidos y una cierta comprensión de como clasificarlos de la mejor manera posible. Suelen hallar patrones similares en los datos que se pueden aplicar procesos analíticos. ejemplo, Una aplicación de reconocimiento desarrollada con machine learning, basada en imágenes que ayudan a distinguir una serie de animales salvajes.

Aprendizaje no supervisado. Se utiliza cuando se analiza una masiva cantidad datos. Ejemplo. La detección de spam en las redes sociales como Twitter, Instagram y Snapchat con granes bases de datos sin ninguna clasificación ni interacción humana. Al existir una cantidad masiva de variables en cada email o spam los clasificadores de machine learning, basados en clustering y asociación se basan en su base de conocimiento para identificar e-mail no deseado y clasificarlos (Kirsch, 2018).

Aprendizaje de refuerzo. Es un modelo de aprendizaje conductual. El algoritmo recibe retroalimentación del análisis de datos, al llevar al usuario hacia el mejor resultado. Dicho aprendizaje se diferencia de otros tipos de aprendizaje supervisado, porque el sistema no está entrenado con el conjunto de datos iniciales, aprendiendo a través de prueba y error. No obstante, una secuencia de decisiones exitosas fortalece el algoritmo, resolviendo de manera más efectiva (Kirsch, 2018).

Deep learning. Es un método específico de machine learning que incorpora redes neuronales en capas sucesivas, para aprender de los datos por iteraciones. El Deep learning es útil cuando se trata de aprender patrones de datos no estructurados, emulando el funcionamiento del cerebro humano, los equipos pueden ser entrenadas para lidiar con abstracciones y problemas mal definidos. Las redes neuronales, son utilizadas a menudo con el reconocimiento de imágenes, voz y aplicaciones de visión de computadora.

Modelos de aprendizaje

Decisión Tree. Decisión Tree, o en español árbol de decisiones es lo que llamamos un algoritmo de aprendizaje supervisado y no paramétrico. Su aprendizaje se emplea una estrategia de división optima repetitiva de forma recursiva de manera que este termine de clasificar con etiquetas específicas todos o la mayoría de registros Este modelo de aprendizaje es utilizado tanto para clasificación de regresión (IBM, ¿Qué es un arbol de decicion?, 2022).

Tipos de Decision Tree.

Hunt, un algoritmo creado en 1960 para modelar en la psicología el aprendizaje humano es la base de muchos arboles de decisión a continuación veremos los más populares:

a) ID3. Abreviatura de "Iterative Dichotomiser 3" atribuido por el desarrollo de Ross Quinlan, Este algoritmo utiliza y aprovecha la ganancia de información como métricas para evaluar las posibles divisiones de candidatos.

b) C4.5. Considerado una iteración posterior de ID3, desarrollado por Quinlan. Utiliza la ganancia de información en la base de datos o las proporciones de ganancia de esta para la evaluación de los puntos de división dentro de los árboles de decisión.
c) CART. Introducido por Leo Breiman con el significado de su Abreviatura de "árboles de clasificación y regresión" o en inglés “clasificación and regresión tres”. Es un algoritmo que aprovecha la impureza de Gini o la frecuencia de clasificación incorrecta para identificar el atributo ideal para la división y lograr su objetivo.

Random Forest. En español Bosque Aleatorio, es una técnica de aprendizaje automático muy común y popular por su capacidad de generalización para muchos problemas. Los Random Forest, son ensambles de varios árboles de decisión con el objetivo de ver las distintas porciones de datos de cada árbol al exceptuar los datos de entrenamiento de aprendizaje. Cada Árbol se entrenará con las distintas muestras de datos obtenidas (Heras, 2020). Al juntar los resultados obtenidos los errores se compensarán con otros con una predicción mejor generalizada.

K-means. En español “Vecinos cercanos” es un algoritmo de clasificación no supervisado de agrupación de objetos basado en similitud de características en K grupos, al no tener etiquetados los datos o sin categoría ni definición se realiza la suma de distancias entre cada objeto y el centroide de su grupo clúster. Los centroides de un clúster son un conjunto de valores de característicos que definen los grupos obtenidos (aprendeIA, 2022). Los grupos K Means son:

1) Los centroides de los clústeres K, que pueden ser utilizados para etiquetar nuevos datos posibles.

2) Etiquetas para los datos de formación y cada punto de datos será asignado a un único clúster.

Neural networks. En español Red neuronal es un sistema de TI caracterizado por la ayuda que da a los equipos en la solución de problemas por sí mismos y mejora de sus propias capacidades. La red neuronal está inspirada en la estructura del cerebro humano que dota a los equipos o dispositivos de inteligencia artificial. Por sí misma la red neuronal pertenece a la rama de la neuro informática al ser parte de la rama de investigación informática (Digital Guide IONOS, 2020).

Cada uno de los tipos de las diferentes artificial neural networks, ofrece diferentes posibilidades de proceso de información y datos lo que depende del método de IA aplicado. Las redes neuronales artificiales pueden describirse como modelos de dos capas que son la entrada y salida como también se pueden agregar algunas capas intermedias conocidas como hidden layers, su uso dependerá de la complejidad del problema a solucionar (Digital Guide IONOS, 2020). En cada capa se puede encontrar numerosas neuronas artificiales especializadas como las siguientes:

Redes neuronales recurrentes. Incorpora redes neuronales en capas sucesivas, para aprender de los datos por iteraciones. La retroalimentación permite que el sistema desarrolle una memoria. Este tipo de redes se usan para el reconocimiento de voz, la traducción y el reconocimiento de texto manuscrito. (Digital Guide IONOS, 2020).

Redes neuronales convolucionales. Son subcategorías de las redes con muchas capas en general han de tener al menos cinco. En cada una de ellas se realiza un reconocimiento de patrones cuyo resultado se ve transmitido en la siguiente capa. Este tipo de redes neuronales se usa para reconocer imágenes (Digital Guide IONOS, 2020).

Capitulo III

Desarrollo

Metodologías empleadas

Design Science Research (DSR)

Es una guía de investigación que busca direccionar a la resolución de problemas no triviales mediante soluciones útiles y efectivos a partir de la creación de artefactos con procesos y resultados innovadores para los mismos. El desarrollo del artefacto involucra un ciclo de actividades de diseño-construcción-evaluación, que iteran tantas veces como sean necesarias antes que el artefacto sea finalmente validado y comunicado para su utilización (Tebes, y otros, 2020).

Los investigadores deben trabajar en estrecha colaboración con las organizaciones, para probar nuevas ideas en un contexto real. Por lo tanto, se puede utilizar como una forma de producción de conocimiento para lograr dos propósitos diferentes en proyectos de investigación al mismo tiempo, producir conocimiento científico y ayudar a las organizaciones a resolver problemas reales (Dresch, Lacerda, & Antunes, 2015).

La investigación en ciencias del diseño puede verse como una encarnación de tres ciclos de actividades estrechamente relacionados. El ciclo de relevancia inicia la investigación en ciencias del diseño con un contexto de aplicación que no solo proporciona los requisitos para la investigación como insumos, sino que también define los criterios de aceptación para la evaluación final de los resultados de la investigación. El ciclo de rigor aporta conocimientos pasados al proyecto de investigación para asegurar su innovación. Es responsabilidad de los investigadores investigar a fondo y referenciar la base de conocimiento para garantizar que los diseños producidos sean contribuciones de investigación y no diseños rutinarios basados en la aplicación de procesos bien conocidos (Hevner & Chatterjee, 2010). El ciclo de diseño central

itera entre las actividades centrales de construir y evaluar los artefactos de diseño y los procesos de la investigación.

SCRUM.

El desarrollo ágil de software se centra en un grupo de trabajo el cual está preparado para reaccionar rápidamente a los cambios que continuamente requiere el proyecto de manera que el producto final se ajuste a las necesidades y cumpla con el tiempo establecido lo que remueve cambios innecesarios (Pressman, 2005).

Scrum es una metodología ágil que organiza diversos procesos para mejorar el trabajo en equipo y tiene como propósito mejorar continuamente el desarrollo de un artefacto final, este proceso es iterativo y dividido en fases que permiten tener un aplicativo entregable en cortos tiempos de trabajo. Este sistema permite ir resolviendo problemas a la medida que se desarrolla el proyecto lo que requiere mucho esfuerzo en cortos periodos, que permite obtener una gran adaptabilidad ante los problemas (Rodas Neira & Villalva Ayala, 2019).

Scrum se basa en prácticas como reuniones diarias, reuniones de planificación de sprint, revisión de sprint reunión, clasificación de trabajos pendientes y presentación de lanzamientos. Estas reuniones tienen una duración aproximada de 15 minutos, y en ellas se obtiene información relevante, mediante una interacción verbal con preguntas entre los miembros del equipo, esto incumbe lo que todos pretenden hacer y cuáles fueron los obstáculos que se presentaron durante ese día, esto sirve para que los demás miembros puedan apoyar con ideas para resolver diversas problemáticas (Lima, de Castro Freire, & Costa, 2012).

Hipótesis de investigación

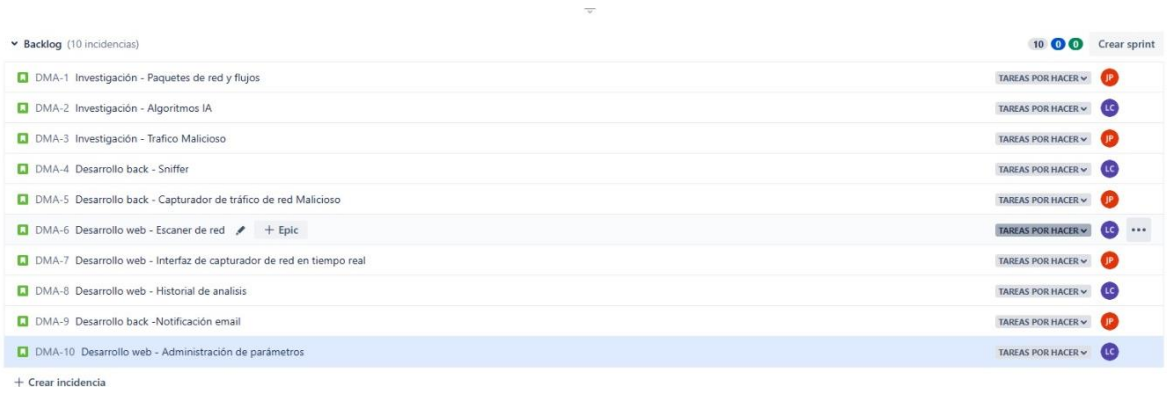
Un modelo de aprendizaje automático para análisis de vulnerabilidades de tráfico malicioso incrementa la detección de ataques cibernéticos.

Product Backlogs

El producto backlog está conformado por una pila de historias de usuarios, las cuales se han de gestionar a medida que avance el proyecto, para ello se implementó el modelo SCRUM a través de la herramienta JIRA como se puede revisar la Figura 3

Figura 3

Backlog



A continuación, se lista las historias de usuarios recolectadas:

1) H1. Yo como investigador de laboratorio requiero que se haga un análisis de características de paquetes de red que se pueden emplear para alimentar algoritmos de inteligencia artificial
2) H2. Yo como investigador de laboratorio requiero que se haga un análisis comparativo de algoritmos de inteligencia artificial basado en los paquetes de redes para aplicar el más optimo en un análisis de red

3) H3. Yo como investigador de laboratorio requiero que se analice y aplique ataques de prueba para recolectar generar datasets
4) H4. Yo como operador requiero que se capturen los paquetes de red para aplicar algoritmos de inteligencia artificial sobre los mismos
5) H5. Yo como operador requiero que se analice el tráfico de red con algoritmos de inteligencia artificial para capturar datos de tráfico malicioso
6) H6. Yo como operador requiero una opción para escanear el tráfico malicioso de una interfaz de red para tomar acciones necesarias
7) H7. Yo como operador requiero una opción para escanear visualizar el tráfico malicioso en tiempo real para tomar acciones necesarias
8) H8. Yo como operador requiero una opción de visualizar el historial de análisis para tener constancia de dichos ataques
9) H9. Yo como operador requiero una opción notificación de alertas vía mail para actuar y proteger la red
10) H10. Yo como operador requiero una opción de configuración de datos del sistema para tener control sobre el mismo

Cada una de ellas cuenta con su respectivo puntaje y tiempo determinado, de acuerdo a la figura 4

Figura 4

Pocker planning

Orden de trabajo/Módulo	Actividad	Menor Valor (puntos historia)	Mayor Valor(puntos historia)	Valor más Frecuente(puntos historia)	Complejidad	Fuente estimación	Justificación para la solución
CERT - OT001-R1	Yo como investigador del espe-cert requiero que se haga una analisis de caracteristicas de paquetes de red que se pueden emplear para alimentar algoritmos de inteligencia artificial	21	21	21	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R2	Yo como investigador del espe-cert requiero que se haga una anlisis compartivo de algoritmos de inteligencia artificial basado en los paquedes de redes para aplicar el mas optimo en un analisis de red	21.00	21.00	21.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R3	Yo como investigador del espe-cert requiero que se analice y aplique ataques de prueba para recolectar generar datasets	21.00	21.00	21.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R4	Yo como operador requiero requiero que se capturen los paquetes de red para aplicar algoritmos de ia sobre los mismos	21.00	21.00	21.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R5	Yo como operador requiero que se analice el trafico de red con algoritmos de inteligencia	21.00	21.00	21.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R6	Yo como operador requiero una opción para escanear el trafico malicioso de una interfaz de red para tomar acciones necesarias	21.00	21.00	21.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R7	Yo como operador requiero una opción para escanear visualizar el tráfico malicioso en tiempo real para tomar acciones necesarias	13.00	13.00	13.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R8	Yo como operador requiero una opción de visualizar el historial de analisis para tener constancia de dichos ataques	8.00	8.00	8.00	Alta	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R9	Yo como operador requiero una opción notificación de alertas via mail para actuar y proteger la red	8.00	8.00	8.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo
CERT - OT001-R10	Yo como operador requiero una opción de configuración de datos del sistema para tener control sobre el mismo	5.00	5.00	5.00	Media	Juicio Experto	En base a consenso de estimación de planing poker de equipo

Sprints

De acuerdo a la metodología Scrum los sprints según (Roche, 2022) son pequeños rangos de tiempo no más allá de un mes, en la cual los desarrolladores trabajan en la construcción de diferentes Funcionalidades. Por lo que en este proyecto se optó por realizar sprints de 2 semanas, los cuales de acuerdo al tiempo estipulado para las historias de acuerdo a la figura 5 y debido a que se trabaja con dos recursos, el seguimiento de los mismos se lo puede revisar en la figura 6.

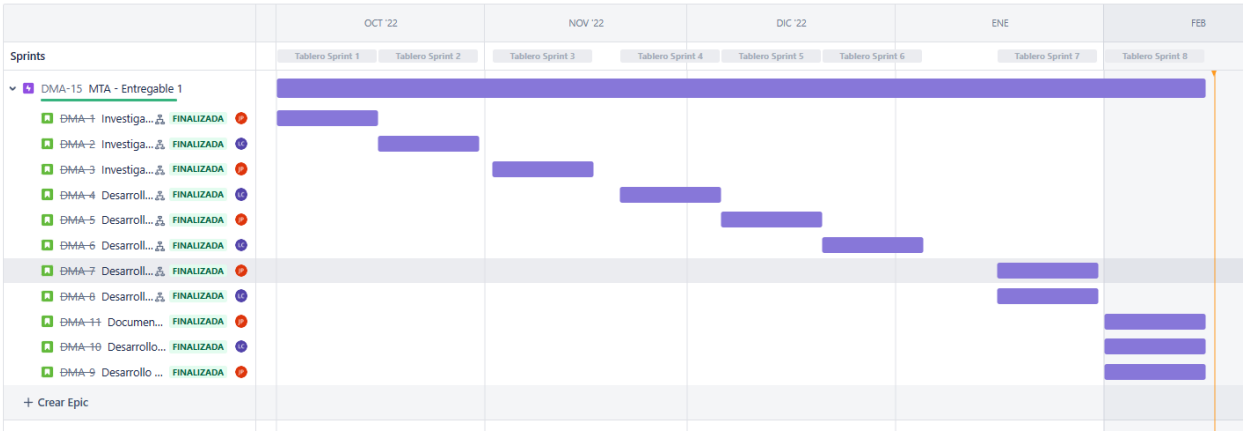
Figura 5

Estimación PERT

			0.5	1.4 (+)Equipo en formación							
Orden de trabajo / Módulo	Actividad	Puntos Historia	Conversión estandar a Días	Factor ajuste equipo	Optimista (-25% del tiempo Mas Probable)	Más probable	Pesimista (+75% del tiempo Mas Probable	Estimación Pert	Desviació	Pert + 2 Sigmas	Estimación Utilizada
CERT - OT001-R1	Yo como investigador del espe-cert requiero que se haga una analisis de caract	21.00	10.50	14.70	11.03	14.70	25.73	**15.93**	2.45	20.83	**21.00**
CERT - OT001-R2	Yo como investigador del espe-cert requiero que se haga una anlisis compartivo de algoritmos de inteligencia artificial basado en los paquedes de redes para aplicar el mas optimo en un analisis de red	21.00	10.50	14.70	11.03	14.70	25.73	**15.93**	2.45	20.83	**21.00**
CERT - OT001-R3	Yo como investigador del espe-cert requiero que se analice y aplique ataques	21.00	10.50	14.70	11.03	14.70	25.73	**15.93**	2.45	20.83	**21.00**
CERT - OT001-R4	Yo como operador requiero requiero que se capturen los paquetes de red para aplicar algoritmos de ia sobre los mismos	21.00	10.50	14.70	11.03	14.70	25.73	**15.93**	2.45	20.83	**21.00**
CERT - OT001-R5	Yo como operador requiero que se analice el trafico de red con algoritmos de i	21.00	10.50	14.70	11.03	14.70	25.73	**15.93**	2.45	20.83	**21.00**
CERT - OT001-R6	Yo como operador requiero una opción para escanear el trafico malicioso de una interfaz de red para tomar acciones necesarias	21.00	10.50	14.70	11.03	14.70	25.73	**15.93**	2.45	20.83	**21.00**
CERT - OT001-R7	Yo como operador requiero una opción para escanear visualizar el tráfico malicioso en tiempo real para tomar acciones necesarias	13.00	6.50	9.10	6.83	9.10	15.93	**9.86**	1.52	12.89	**13.00**
CERT - OT001-R8	Yo como operador requiero una opción de visualizar el historial de analisis para tener constancia de dichos ataques	8.00	4.00	5.60	4.20	5.60	9.80	**6.07**	0.93	7.93	**8.00**
CERT - OT001-R9	Yo como operador requiero una opción notificación de alertas via mail para actuar y proteger la red	8.00	4.00	5.60	4.20	5.60	9.80	**6.07**	0.93	7.93	**8.00**
CERT - OT001-R10	Yo como operador requiero una opción de configuración de datos del sistema para tener control sobre el mismo	5.00	2.50	3.50	2.63	3.50	6.13	**3.79**	0.58	4.96	**5.00**
0	0	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	**0.00**
0	0	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	**0.00**

Figura 6

SPRINTS



Requerimientos

Requerimientos funcionales

a) Escáner de red

Tabla 2

Requerimiento funcional Escaneo de red

Características	**Descripción**
Id Requerimiento:	RF001
Sistema:	MT Analyzer
Requerimiento:	La aplicación permitirá al usuario realizar un escaneo de paquetes al seleccionar una interfaz de red.
Tipo de requerimiento:	Nuevo
Dependencias	No aplica
Precondición:	No aplica
Actor:	Operario del Observatorio
Incidencias enlazadas:	DMA-1, DMA-2, DMA-3, DMA-4, DMA-5, DMA-6, DMA-7
Secuencia Normal:	1. Seleccionar interfaz de red 2. Comenzar escaneo 3. Finalizar escaneo 4. Visualización de resultados

Características	**Descripción**
Excepciones:	No aplica
Reglas de negocio	1. Se debe listar las interfaces de redes disponibles y al seleccionar una de ellas, permitirá iniciar el escaneo de la misma. 2. Al escanear se debe mostrar la información del estado del escaneo, asimismo si encuentra algún patrón de tráfico malicioso, mostrará sus datos en tiempo real. 3. Debe existir la opción para detener el escaneo y al seleccionarla mostrará un resumen del conteo de paquetes hecho por el escáner.
Frecuencia de uso	Alta

b) Envío de alerta por email

Tabla 3

Requerimiento funcional Envío de alerta por mail

Características	**Descripción**
Id Requerimiento:	RF002
Sistema:	MT Analizer
Requerimiento:	La aplicación alertará vía mail si existe algún indicio de tráfico malicioso
Tipo de requerimiento:	Nuevo
Dependencias	No aplica
Precondición:	Haber comenzado un escaneo previo
Actor:	Mt Analizer
Incidencias enlazadas:	DMA-9
Secuencia Normal:	1. Comenzar escáner 2. Detección de tráfico malicioso 3. Envío de email de alerta
Excepciones:	No se enviará el mail si no se ha configurado previamente un usuario a quien enviar
Reglas de negocio	1. Se debe parametrizar los datos del usuario remitente 2. Se debe enviar el email de alerta solo cuando el escáner encuentre un indicio de tráfico malicioso en tiempo real
Frecuencia de uso	Media

a) Historial de análisis

Tabla 4

Requerimiento funcional Historial de análisis

Características	**Descripción**
Id Requerimiento:	RF003
Sistema:	MT Analizer
Requerimiento:	La aplicación tendrá una sección donde presentará un historial de las ejecuciones de escaneos con sus resultados
Tipo de requerimiento:	Nuevo
Dependencias	No aplica
Precondición:	Haber comenzado o finalizado un escaneo previo
Actor:	Mt Analizer
Incidencias Enlazadas:	DMA-8
Secuencia Normal:	1. Comenzar escáner 2. Finalizar Escáner 3. Revisar el historial de escaneos
Excepciones:	No aplica
Reglas de negocio	1. Se debe listar un historial de escaneos con sus respectivos resultados 2. El historial debe tener la opción para revisar un listado de flujos que demostraron indicios de tráfico malicioso
Frecuencia de uso	Media

b) Administración de parámetros

Tabla 5

Requerimiento funcional Administración parámetros

Características	**Descripción**
Id Requerimiento:	RF003
Sistema:	MT Analizer
Requerimiento:	La aplicación tendrá una sección permitirá editar los parámetros del sistema
Tipo de requerimiento:	Nuevo
Dependencias	No aplica
Precondición:	No aplica
Actor:	Mt Analizer
Incidencias enlazadas:	DMA-10
Secuencia Normal:	1. Selecciona Configuración 2. Editar parámetro y guardar
Excepciones:	No aplica
Reglas de negocio	1. Se debe listar el listado de parámetros del sistema 2. Debe existir la opción para editar un parámetro y guardarlo
Frecuencia de uso	Baja

Requerimientos no funcionales

a) RNF1. La aplicación tendrá una disponibilidad de 24/7

b) RNF2. La aplicación será intuitiva y fácil de usar

c) RNF3. La aplicación tendrá una interfaz web responsiva

d) RNF4. Toda funcionalidad del sistema debe responder al usuario en menos de 5 segundos.

Requerimientos de funcionalidad de la aplicación

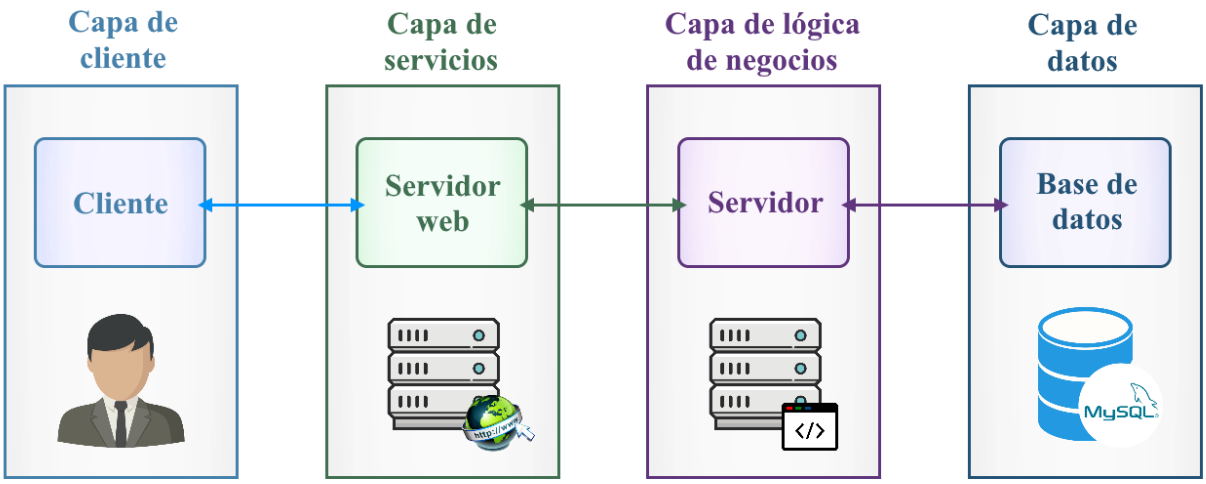
a) RFA1: Cualquier actualización o modificación realizada a la aplicación será notificada a los respectivos dueños y desarrolladores.
b) RFA2: Para modificar, agregar o eliminar una funcionalidad, se necesitará la autorización previa de los propietarios de la aplicación y un acuerdo con los desarrolladores para aumentar el alcance del proyecto.
c) RFA3: MTAnalyzer no será responsable del uso indebido de la aplicación.

Arquitectura

El presente proyecto contiene una arquitectura de software separada por capas, que divide roles y responsabilidades de forma jerárquica como se muestra en la Figura 7. La capa de cliente presenta el sistema al usuario, le comunica y captura la información del mismo, también es conocida como interfaz gráfica. La capa de servicios conformado por el servidor de aplicaciones, proporciona funcionalidades que pueden ser tanto de seguridad como de desempeños específicos propios de la aplicación (Flask, celery y apache) (IBM, Capa de servicios, 2021).

Figura 7

Arquitectura por capas



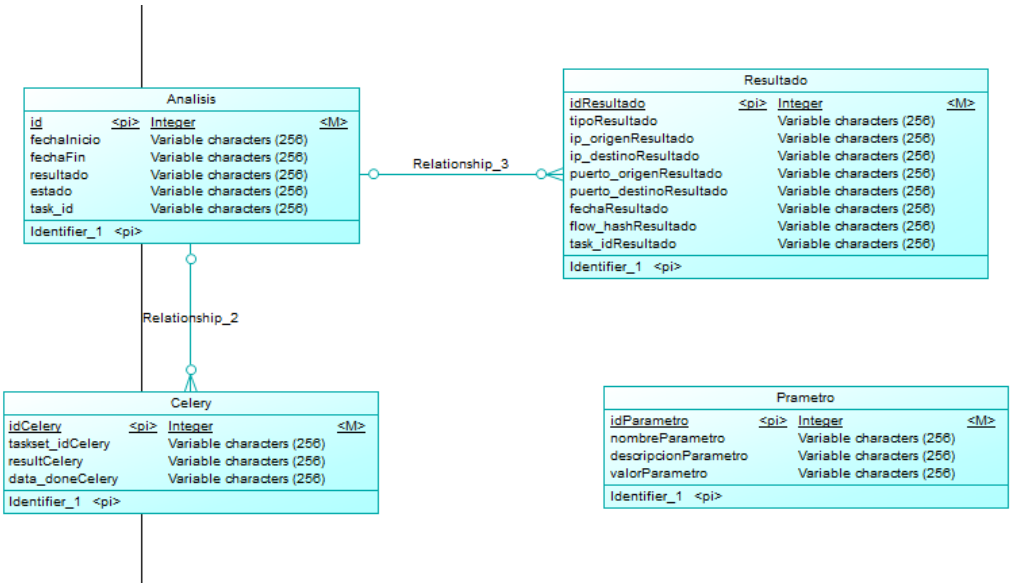
La capa lógica de negocios sirve como intermediario para el intercambio de datos entre la de cliente, servicios y datos, al manejar la lógica, cálculos, reglas de negocio y la administración (Gestión con la base de datos), dentro de la aplicación (Cuofano, 2022). Para finalizar, la capa de datos es donde residen los registros, se encarga de acceder a los mismos, al recibir solicitudes de almacenamiento o recuperación de información desde la capa de lógica de negocio.

Capa de datos

Modelo conceptual. Según (elConspirador, 2013), ayuda a representar un problema de manera gráfica, como se muestra en la figura 8 para la elaboración del modelo conceptual se incorporó cuatro tablas que ayudan al manejo de los datos que se utilizaran a lo largo de la aplicación. Como son: el almacenar los metadatos generados por escaneos realizados y además la configuración del sistema a través de parámetros, como el correo electrónico.

Figura 8

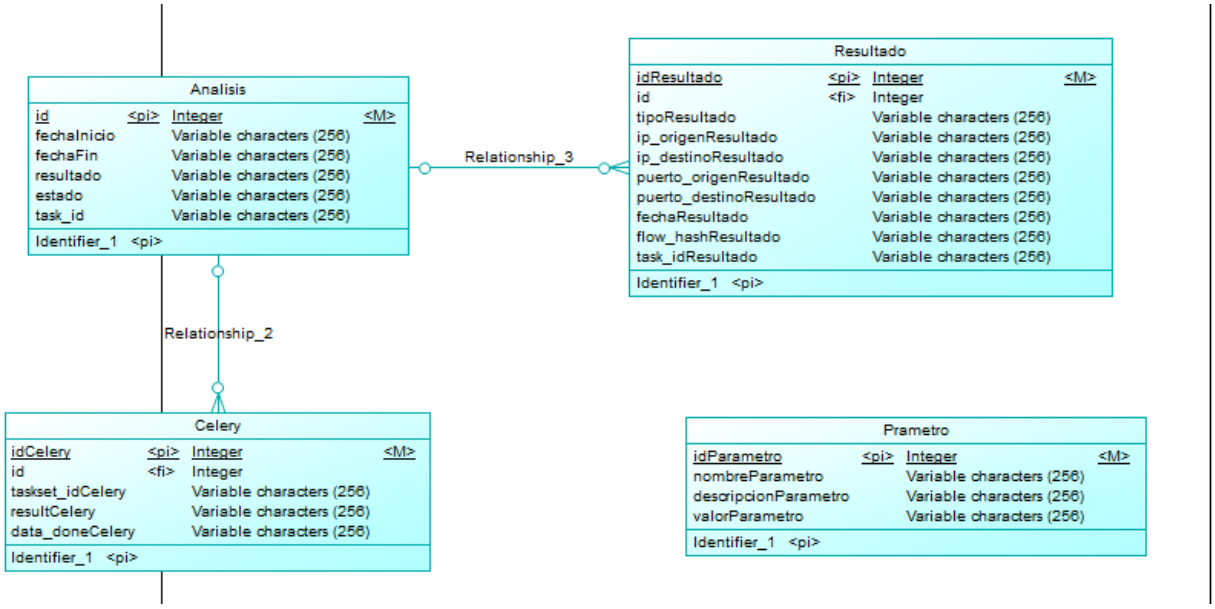
Modelo conceptual de la base de datos Autoría propia



Modelo lógico. Según (TIBCO, 2020), ayuda a tener una mejor relación entre todas las tablas, siendo un escalón superior frente al modelo conceptual, como se muestra en la figura 9 las relaciones entre las tablas son más estructuradas al incluir claves foráneas detallando si poseen una relación de uno a uno, uno a muchos, etc.

Figura 9

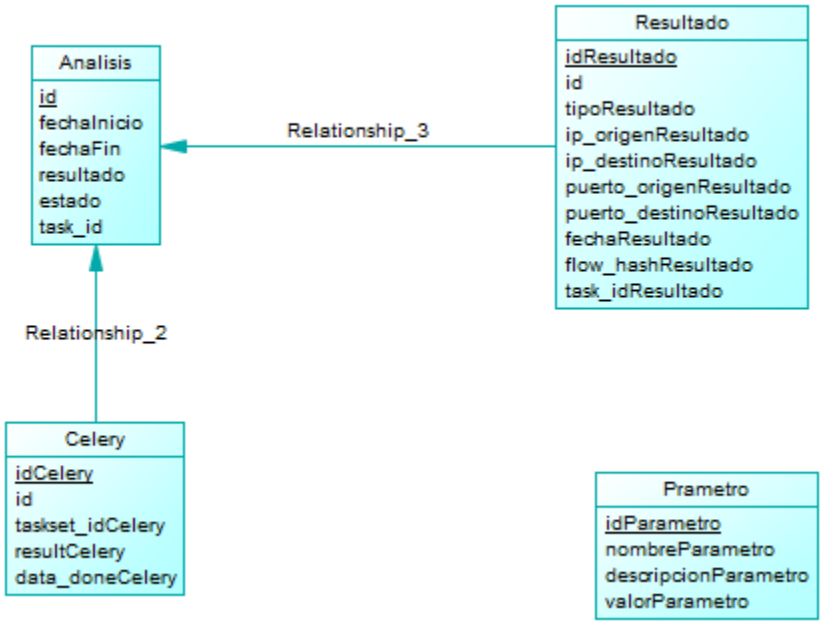
Modelo lógico de la base de datos Autoría propia



Modelo físico. Según (KeepCoding, 2023), es una representación gráfica de la estructura que se debe generar en el gestor de base de datos. Como se muestra en la figura 10 sirve para tener un esquema de cómo se debe construir la base de datos, para que la misma se complemente con los procesos a implementar.

Figura 10

Modelo físico de la base de datos Autoría propia



Casos de Uso

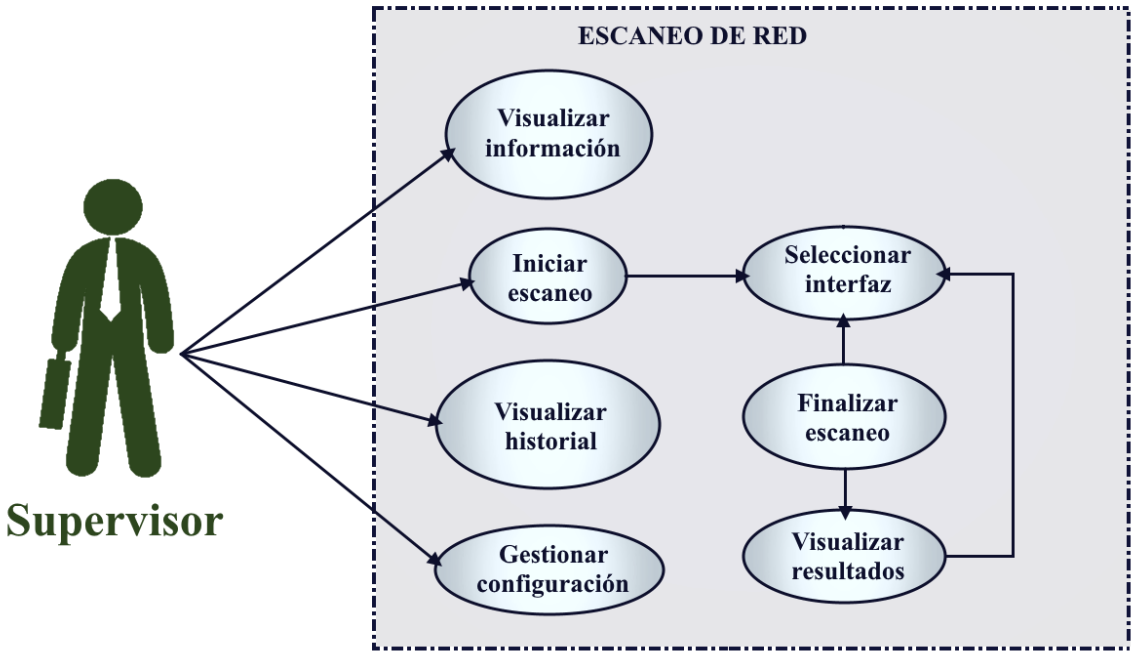
De acuerdo con (Pressman R. S., 2010), los casos de uso describen el comportamiento del sistema en distintas condiciones en las que una aplicación responde a una petición de alguno de sus participantes. Muestra cómo interactúa un usuario final (Con una serie de roles) con el sistema en circunstancias determinadas. A continuación, se propone los distintos casos de usos utilizados en la elaboración de la aplicación web y los algoritmos.

Escaneo de red

El usuario cuenta con un listado de opciones como se muestra en la figura 11, la cual le permitirá iniciar un escaneo de red, visualizar el historial de escaneos previos, configurar parámetros y revisar la información de la aplicación.

Figura 11

Caso de uso Escaneo de puertos Autoría propia

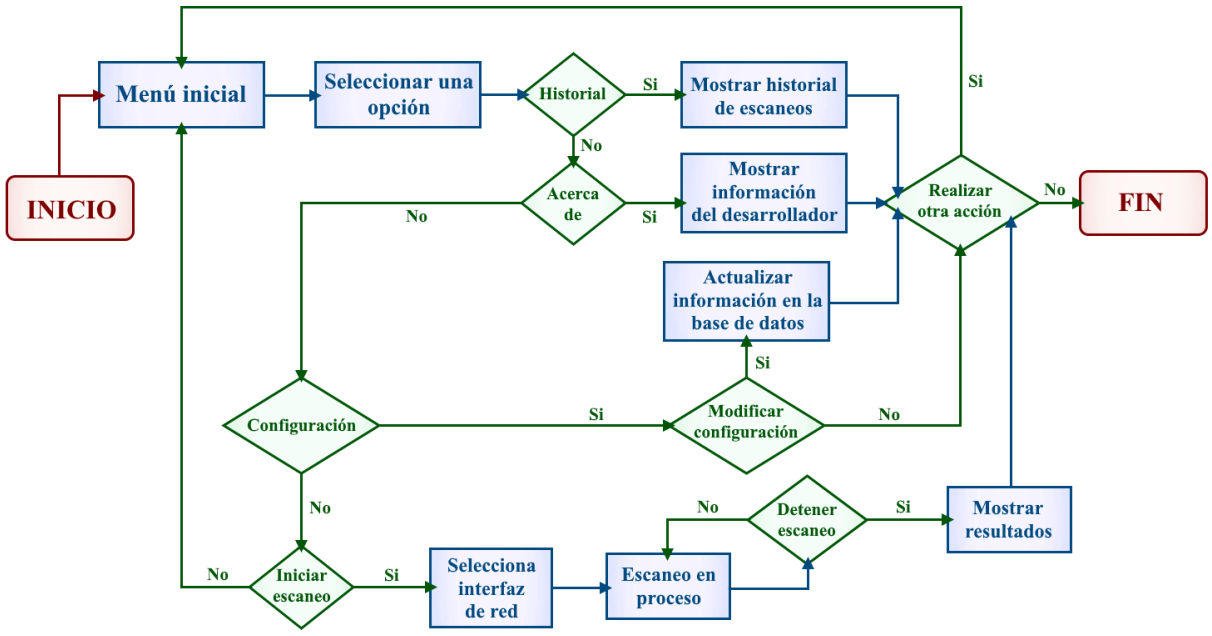


Flujograma del sistema

Según (Ucha, 2022) el flujograma es una representación gráfica de situaciones, hechos, movimientos y relaciones de todo tipo a partir de símbolos que organiza y resume una actividad, en una secuencia de pasos y/o trazados alternativos. El flujograma de la aplicación web desarrollado se representa en la figura 12 conformado por un menú de opciones para la navegabilidad por parte del usuario. Este mismo dispone de un escaneo de red el cual lista las interfaces y permite seleccionar una de ellas para iniciar el análisis, conjuntamente posee con un historial de escaneos que enumera los datos de escaneo previos. También cuenta con la opción de configuración que permite editar la información relevante del aplicativo, por último, dispone de una interfaz para la información del proyecto y sus involucrados.

Figura 12

Flujograma del sistema



Datos empleados

Recopilación

La recopilación de la data para la aplicación se lo hizo a partir de dos grupos, el primero conformado por el Dataset CICIDS2017, que representa el 50% de la información empleada para los algoritmos, este cuenta con registros de ataques realizados y tráfico de red benigno, se encuentra organizado de acuerdo a los días que fueron recolectados. Asimismo, la otra mitad restante son datos obtenidos por el equipo de investigación a través de pruebas internas en un entorno controlado. (Lashkari, 2020).

Preparación

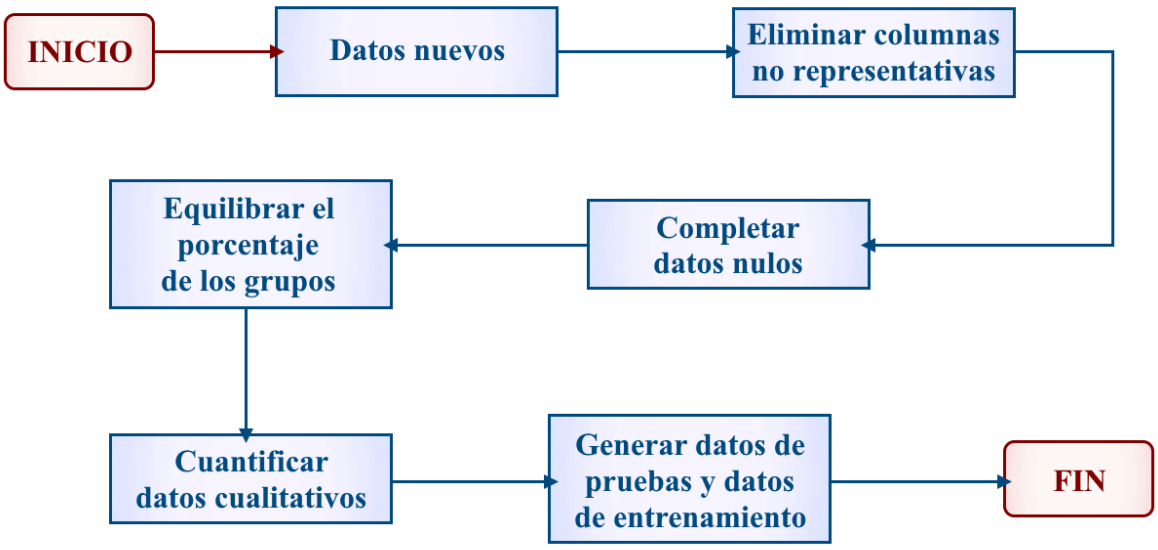
Para la preparación de los datos se empleó un proceso de rediseño de datos como se puede visualizar en la figura 14, el cual permite completar la información inconsistente que provoca errores en el resultado esperado. Además, entrega a los algoritmos información adecuada para el procesamiento y desarrollo. A continuación, se detalla de mejor manera cada paso empleado:

1) Datos nuevos: Hace referencia a un nuevo conjunto de datos
2) Eliminar columnas no representativas: Se elimina las características que no aportan valor al algoritmo como por ejemplo la IP que es un dato cualitativo que no ayuda a reconocer un patrón claro en el proceso de clasificación.
3) Completar datos nulos: Se completa los datos para no tener inconvenientes con las entradas del algoritmo, en este paso se tiene dos opciones, eliminar los registros vacíos o llenar con la media de toda la columna.
4) Equilibrar el porcentaje de los grupos: Si se cuenta con un grupo mayor de lo normal en una de las salidas, se puede llegar a converger rápidamente el algoritmo sin reconocer las demás salidas a clasificar. Para este proceso se elimina un porcentaje de los grupos mayoritarios para obtener un equilibrio en el Dataset

5) Cuantificar datos cualitativos: Se proporciona un valor numérico a los datos cuantitativos, debido a que los algoritmos no reconocen este tipo de valores.
6) Generar datos de pruebas y datos de entrenamiento: Paso final, el cual secciona los datos aleatoriamente, para el entrenamiento y comprobación de los algoritmos a diseñar.

Figura 13

Flujograma preparación de los datos Autoría propia

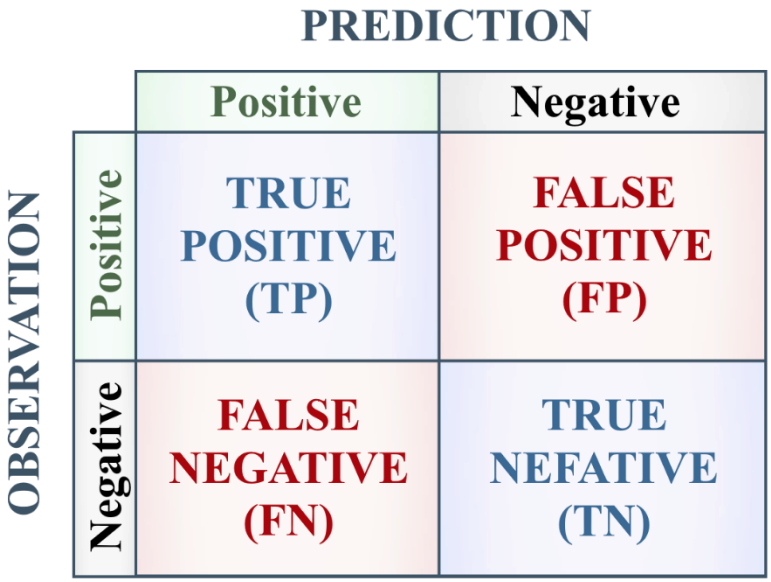


Métricas. Las métricas ayudan a interpretar el desempeño de un algoritmo, se puede emplear diferentes herramientas para generarlas, una de ellas es la matriz de confusión.

La matriz de confusión. De acuerdo con (Shin, 2020) es una herramienta que permite medir la eficacia de un modelo de clasificación construido sobre un sistema de aprendizaje automático. Esta matriz Como se muestra en la figura 15 se presenta siempre en forma de tabla, de manera que en cada columna aparece el número de predicciones de cada clase, mientras que cada fila muestra el número real de instancias de cada clase. La matriz relaciona las predicciones realizadas por un algoritmo de aprendizaje supervisado y los resultados correctos que debería haber mostrado. Así puede medirse el desempeño, determinando qué tipo de errores y aciertos posee cada modelo, a la hora de pasar por un proceso de aprendizaje sobre datos propuestos. (Núñez, 2018).

Figura 14

Matriz de confusión Autoría propia



Estructura. Positivo (P), es la observación es positiva (por ejemplo, es un gato), Negativo (N), es la observación es negativa (por ejemplo, no es un gato), Verdadero Positivo (TP), es el resultado donde el elemento predicho como positivo es correcto, Verdadero Negativo (TN), es el resultado donde el elemento predicho como negativo es correcto, Falso Positivo (FP), es llamado error de tipo 1, resultado cuando el modelo predice incorrectamente la clase positiva cuando en realidad es negativa, Falso Negativo (FN), es llamado error de tipo 2, resultado cuando el modelo predice incorrectamente la clase negativa cuando en realidad es positiva.

Recuperación/Sensibilidad. La recuperación o sensibilidad, es la tasa de verdaderos positivos (TPR) dividido entre la sumatoria de los verdaderos positivos (TP) y el total de falsos negativos (Mariposa, 2023).

$$\boldsymbol{R} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP} + \boldsymbol{FN})}$$

Un valor alto significa que hay menos falsos negativos, lo que conlleva un beneficio para el algoritmo.

Precisión. La precisión es la medida de los verdaderos positivos dividida entre la suma de los verdaderos positivos y los falsos positivos (Mariposa, 2023).

$$\boldsymbol{Pr} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP} + \boldsymbol{FP})}$$

Ejemplo, si una persona no tiene inflamación, no obstante, el modelo lo muestra y el médico le receta ciertos antibióticos. Esto puede provocar efectos adicionales en el paciente.

Especificidad. La especificidad es el total de verdaderos negativos dividido entre la suma de los falsos positivos y los verdaderos negativos. (Mariposa, 2023).

$$\boldsymbol{S} = \frac{\boldsymbol{TN}}{(\boldsymbol{TN} + \boldsymbol{FP})}$$

Es una medida que cuantifica el desempeño del clasificador al identificar valores negativos.

Exactitud. La precisión es la suma entre los verdaderos positivos con los verdaderos negativos dividida entre la suma entre los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos existentes (Mariposa, 2023).

$$\boldsymbol{Precisión\ A} = \frac{(\boldsymbol{TP} + \boldsymbol{TN})}{(\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN})}$$

Ejemplo, si se encuentra correctamente 15 valores positivos y 5 negativos de una muestra de 30, la precisión de su modelo será 20/30.

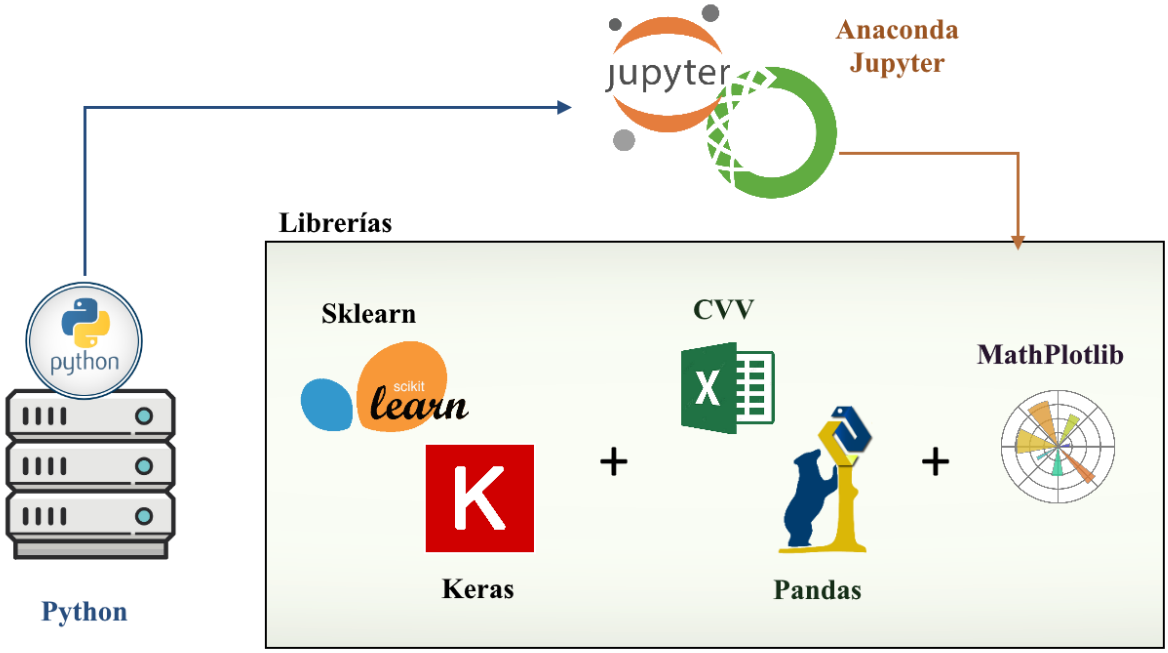
Algoritmos de Inteligencia Artificial

Tecnologías implementadas

Para el desarrollo de los algoritmos de inteligencia artificial se empleó diferentes tecnologías como se observa en la figura 16. fundamentadas en Python que es un lenguaje de programación de alto nivel, orientado a objetos multifuncional (Castro, 2021). Keras y Sklearn son librerías que permiten diseñar la arquitectura (empleando Inteligencia artificial), mientras que Pandas se usa para el manejo de datos y MathPlotlib para la visualización mediante diagramas. Todo esto gracias a las interfaces web como Jupyter notebook de anaconda que permite la ejecución y documentación del código.

Figura 15

Arquitectura Desarrollo de los algoritmos de aprendizaje Autoría propia

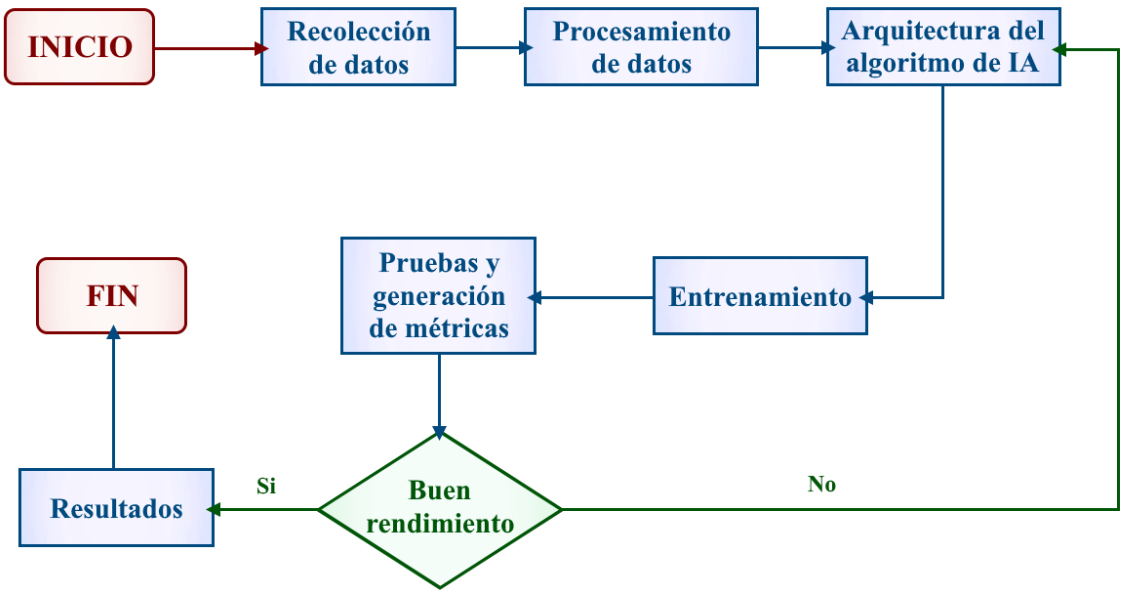


Desarrollo de algoritmos

Cada algoritmo cuenta con diferente arquitectura, no obstante, su desarrollo contiene un conjunto de pasos similares con el objetivo de alcanzar el mejor desempeño, en la figura 17 se visualiza el proceso general para el desarrollo de cada uno de los algoritmos.

Figura 16

Flujograma desarrollo de los algoritmos Autoría propia



Arquitectura de los distintos algoritmos empleados

Random Forest. Es un algoritmo de aprendizaje supervisado como se muestra en la figura 18, que combina la salida de múltiples árboles de decisión, para llegar a un único resultado y seleccionar el más óptimo (IBM, What is random forest?, 2022). El diseño de este algoritmo como se muestra en la figura 19, detalla las variables de entradas como el número de árboles a generar y un estado para manejar la aleatoriedad de los datos. Como salidas se obtiene la clasificación de los diferentes tipos ataques.

Figura 17

Arquitectura Random Forest Autoría propia

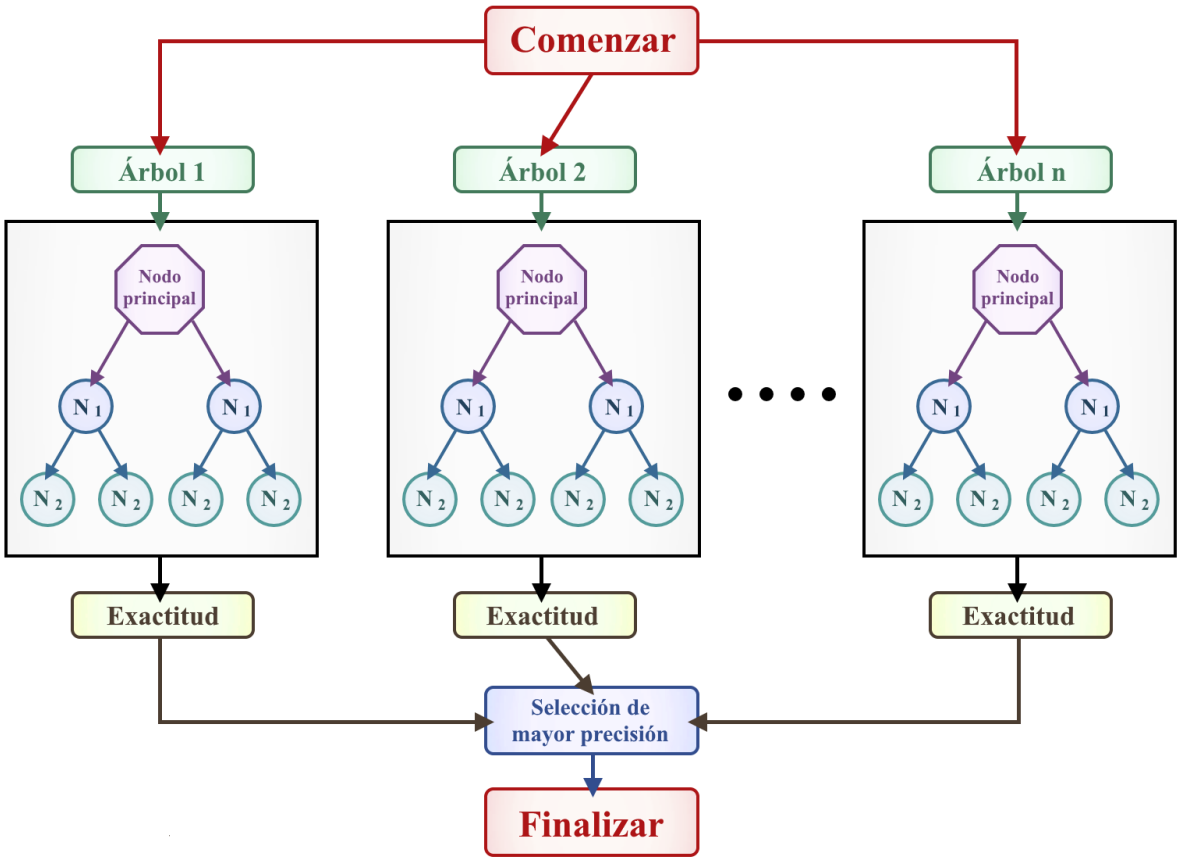
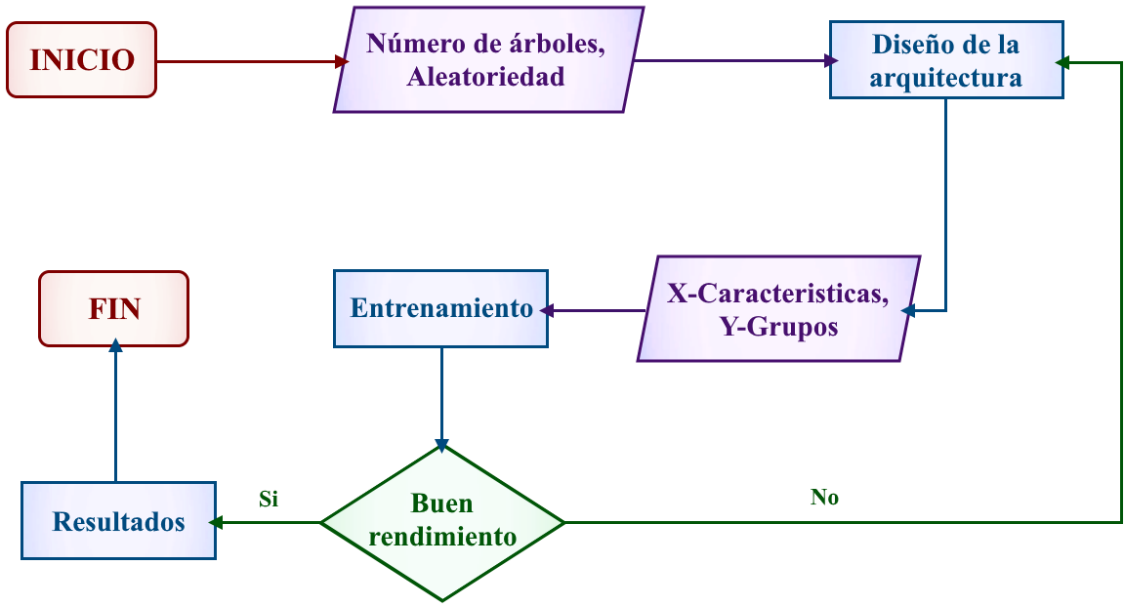


Figura 18

Flujograma Random Forest Autoría propia



Decision Tree. Es un algoritmo de aprendizaje supervisado no paramétrico como se muestra en la figura 20, de uso para tareas de clasificación como de regresión. Tiene una estructura de árbol jerárquica, que consta de un nodo padre, ramas, nodos hijos y nodos hoja hasta llegar a un resultado (Los nodos hoja representan todos los resultados posibles dentro del conjunto de datos) (IBM, ¿Qué es un árbol de decisión?, 2022). En el diseño de este algoritmo se detalla la profundidad del árbol a generar. Como se observa en la figura 21, el árbol comienza desde la característica más relevante y se desglosa a medida que continúa. Igualmente, se visualiza los diferentes puntos de inflexión que se generan a partir de las características de los paquetes de red.

Figura 19

Arquitectura Decision Tree Autoría propia

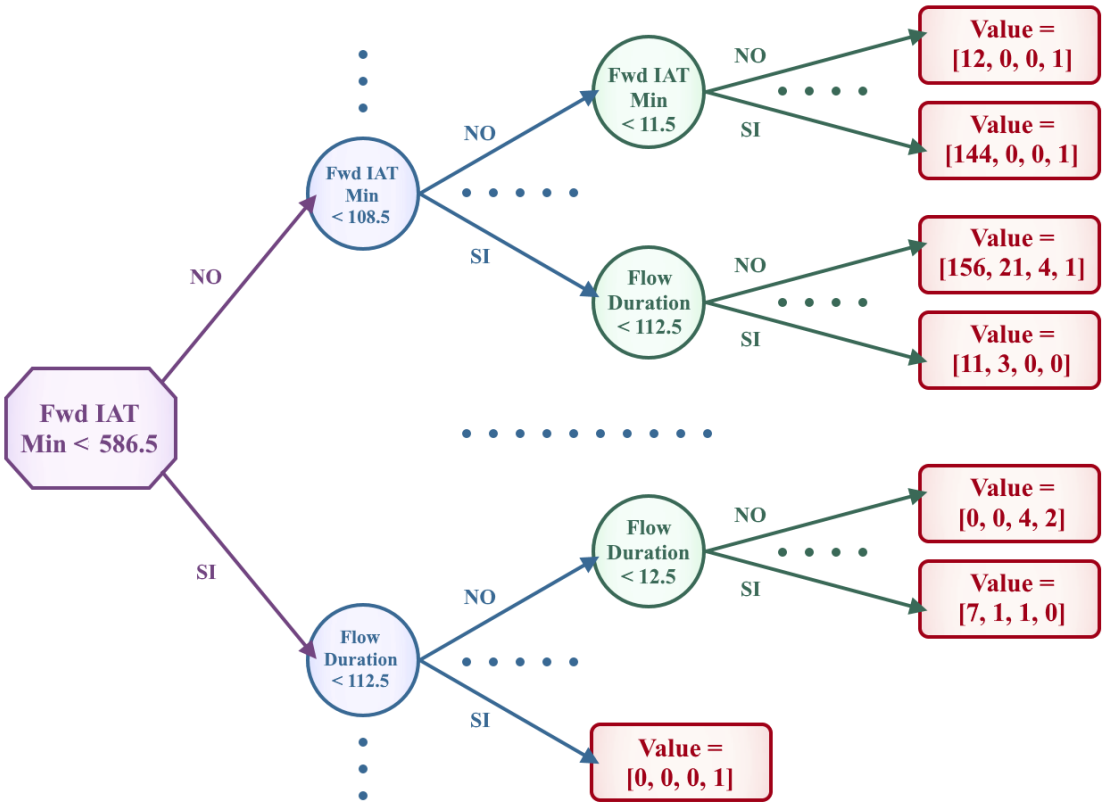
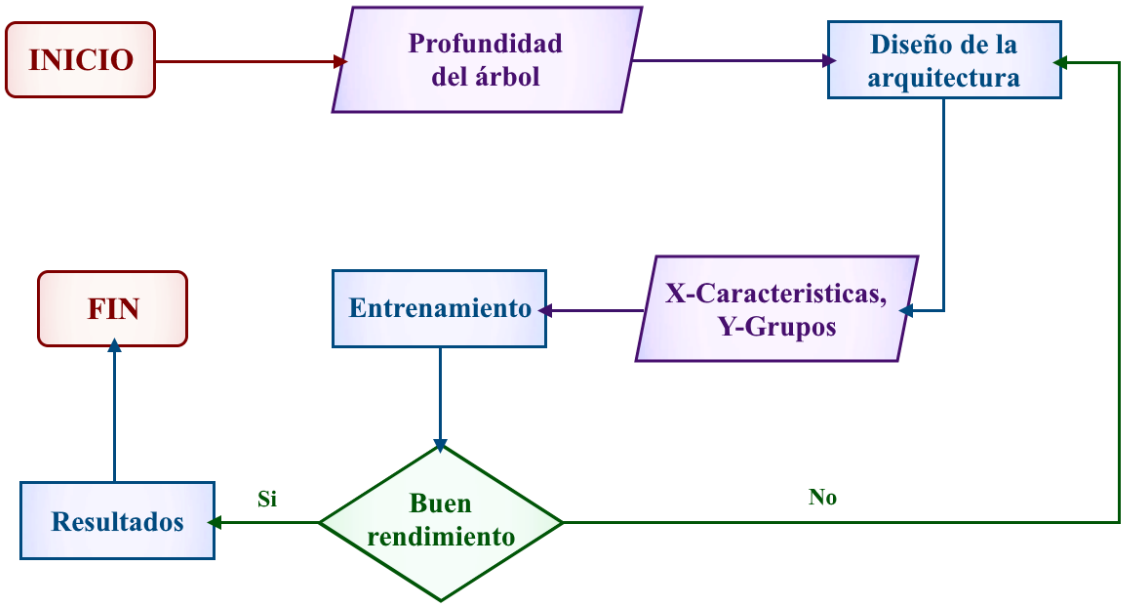


Figura 20

Flujograma Decision Tree Autoría propia



K-MEANS. Es un algoritmo aprendizaje automático no supervisado como se muestra en la figura 22. Consiste en la agrupación de datos con características similares, se pueden manejar de forma computacionalmente eficiente aplicando la función de distancia para emparejar instancias con centros de clúster, uno de los objetivos principales del algoritmo es descubrir cuál es el mejor agrupamiento de los datos (Education, Clústeres de k-medias, 2021). En el diseño del algoritmo como se observa en la figura 23, se detalla el número de grupos a identificar que para este caso son los diferentes tipos de ataques, estos se convierten en la variable dependiente, mientras que las variables independientes son todas las características de los flujos de red.

Figura 21

Arquitectura K-MEANS Autoría propia

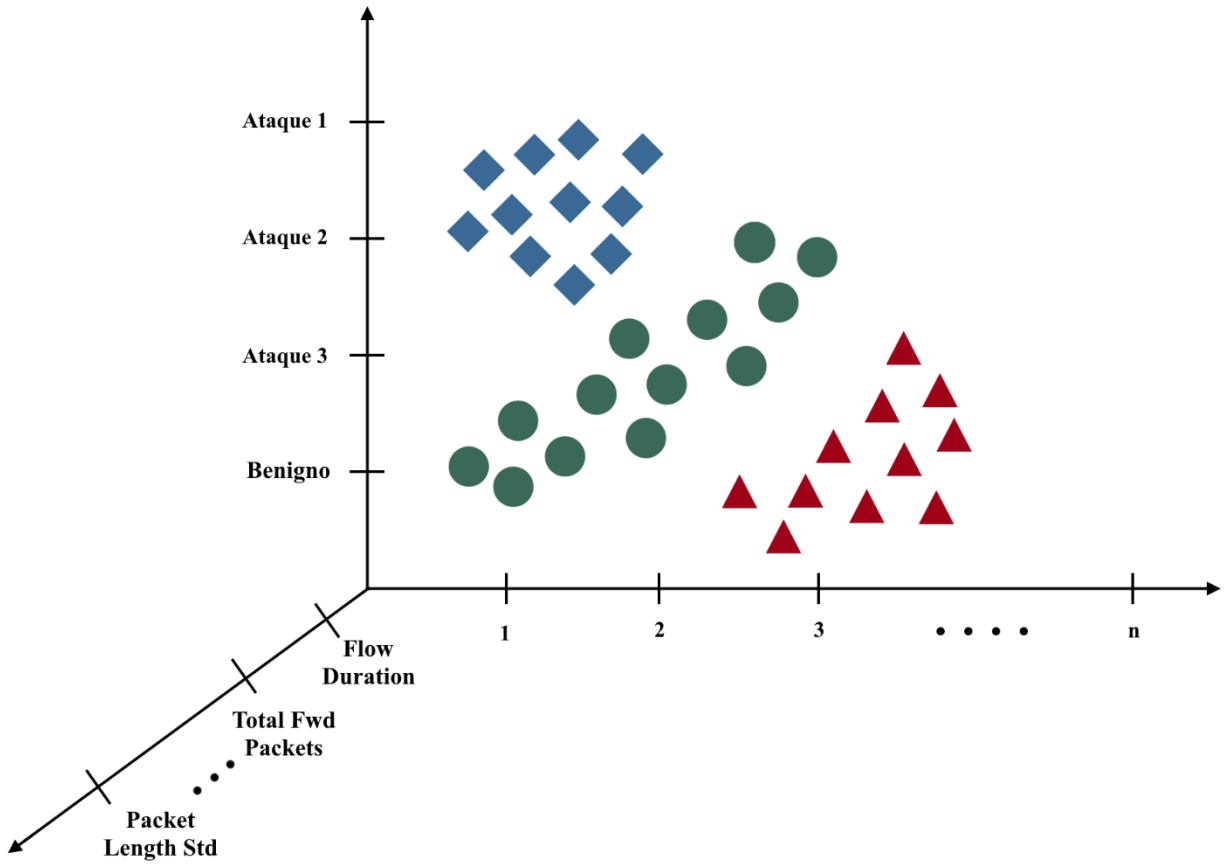
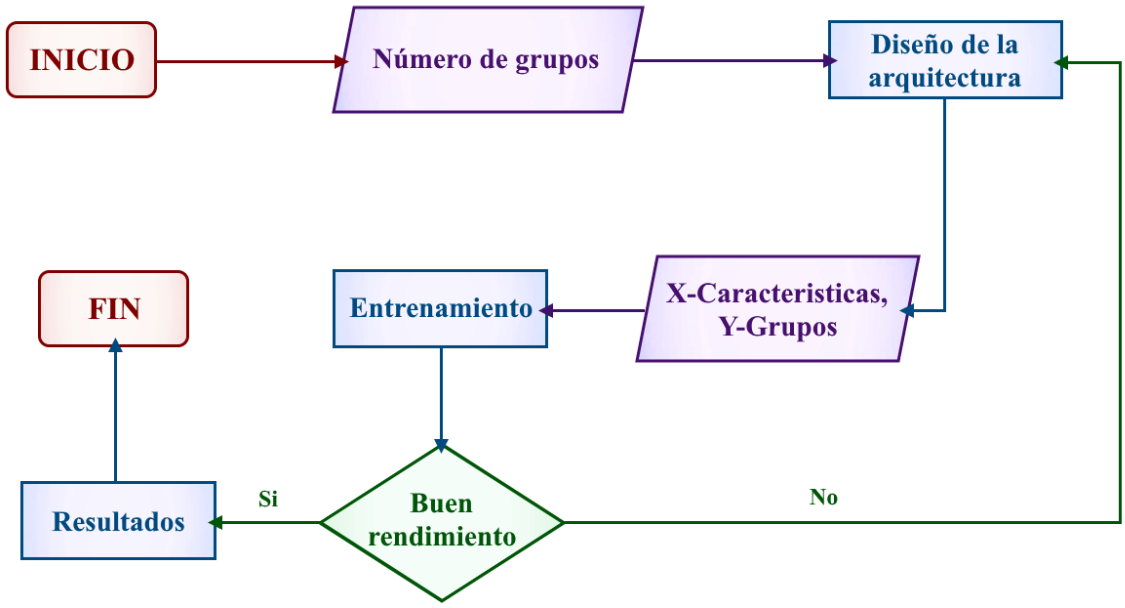


Figura 22

Flujograma K-Means Autoría propia



Neuronal networking. Conocidas como redes neuronales artificiales (ANN) están formadas por capas de nodos, que contienen una capa de entrada, una o varias capas ocultas y una capa de salida como se muestra en la figura 24. Realizan una tarea repetitiva que ayuda a mejorar de manera gradual el resultado lo que permite el aprendizaje progresivo (Education, 2020). En el diseño de este algoritmo como se observa en la figura 25 se detalla el número de capas, el número de neuronas por capa, y la función de activación de cada neurona. También, la primera capa se compone por el número de características del Dataset, y la capa de salida tendrá el número de grupos a clasificar.

Figura 23

Arquitectura Neuronal networking Autoría propia

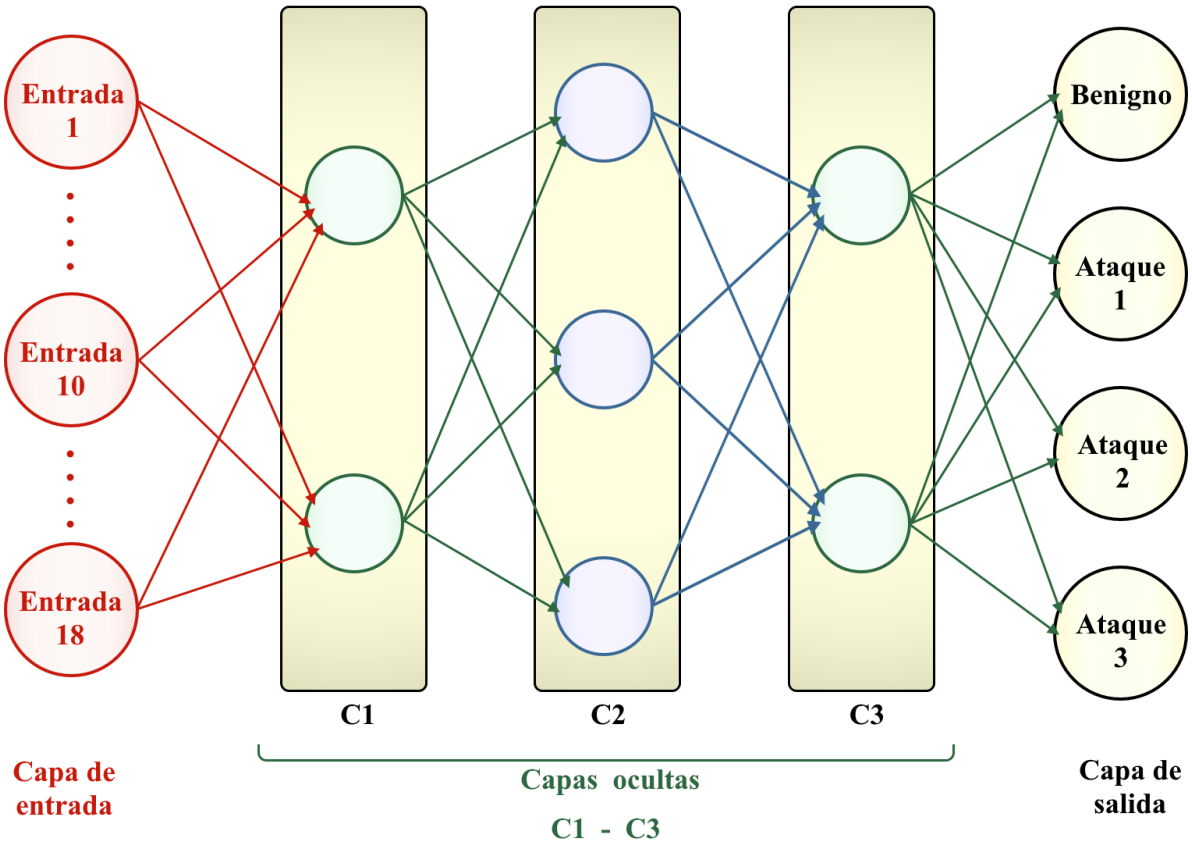
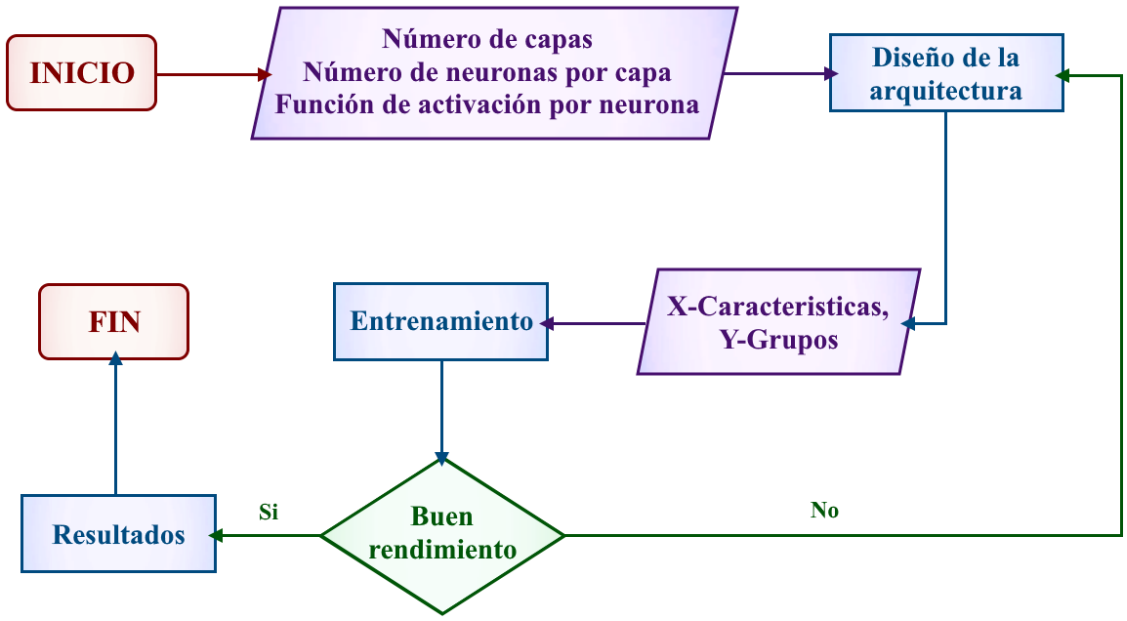


Figura 24

Flujograma Neuronal networking Autoría propia



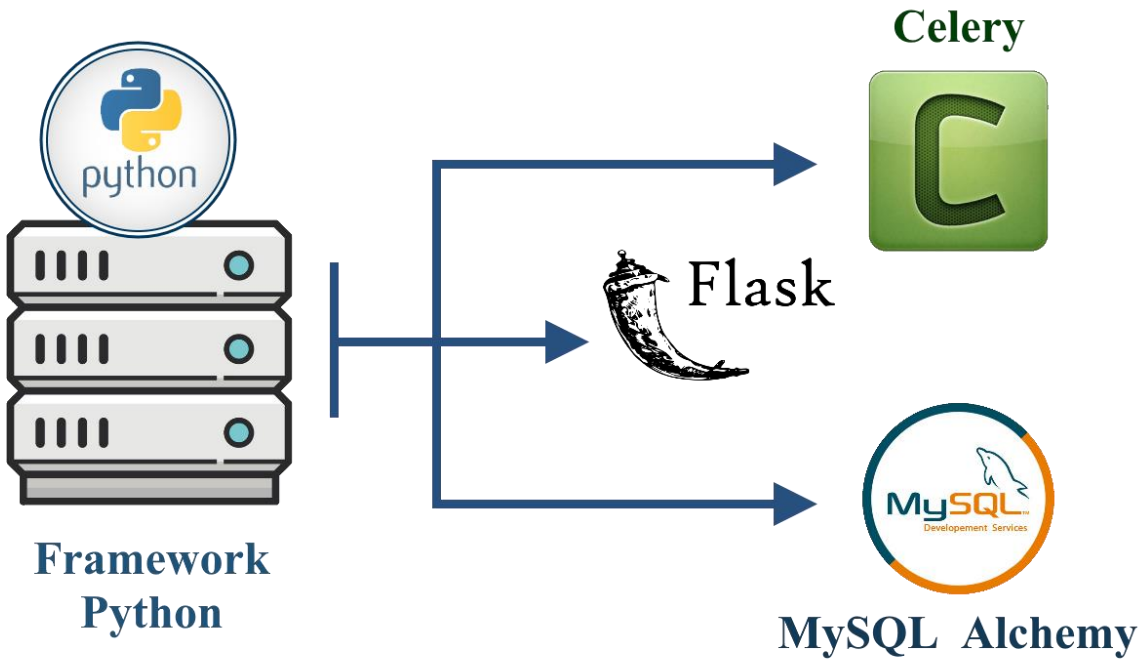
Aplicación web

Desarrollo

Para el desarrollo de la aplicación web se empleó la arquitectura como se puede observar en la figura 26 conformado por Python que es un lenguaje de programación de alto nivel, orientado a objetos, diseñado, principalmente para el desarrollo web y de aplicaciones informáticas, junto a Flask que es un micro-framework de Python que permite crear aplicaciones web con pocos pasos y puede llegar a escalar en grandes arquitecturas. (Castro, 2021), a partir de allí se implementa diferentes librerías como Celery que es una biblioteca de Python de código abierto que se utiliza para la ejecución de tareas paralelas, el cual permite ejecutar trabajos de forma asíncrona para no bloquear el flujo normal del programa (Canelo, 2022). Para finalizar el sistema usa MySQL en su capa de datos que es un sistema de gestión de bases de datos relacionales (RDBMS) de código abierto (Castro, 2021).

Figura *25*

Arquitectura Desarrollo de la aplicación web



Codificación

El código empleado se encuentra disponible en GitHub a través del siguiente enlace https://github.com/Skairlex/ESPE-DMT.git

Pruebas

Para la realización de las pruebas funcionales correspondientes, Según (Narvaez, 2023) en base a la normativa ISO 9001, cuyo apartado 9.1.2 menciona que la satisfacción del cliente se define como el resultado de comparar las expectativas del mismo, con respecto a los productos y servicios entregados. Se puede emplear las encuestas de satisfacción de cliente que consiste en un cuestionario con varias preguntas (Opción múltiple y abiertas) relacionadas al producto entregado, que miden su calidad mediante una escala de valores.

Para validar la encuesta, se requiere que profesionales con experiencia en el CERT académico evalúen la interfaz y usabilidad del producto mediante las siguientes preguntas como se observa en la tabla 7:

Tabla 6

Encuesta a expertos CERT-Académico

ENCUESTA ESPE-CERT					
	Mucho	**No mucho**	**Poco**	**Muy poco**	**Para nada**
¿El diseño de la aplicación es de su agrado?					
En cuanto al diseño, ¿Qué modificaciones propone?					
	Mucho	**No mucho**	**Poco**	**Muy poco**	**Para nada**
¿La aplicación cumple con sus expectativas?					
¿Qué funcionalidades le gustaría agregar o modificar?					
¿Recomendaría esta aplicación a sus conocidos?		**Si**			**No**

Pruebas de software

Según (Pressman R. S., 2010) las pruebas son un conjunto de actividades que se pueden planificar por adelantado y llevar a cabo sistemáticamente. Al trabajar junto a la norma ISO 29119 según (Tuya, 2009) que especifica procesos de prueba que pueden utilizarse para gobernar, gestionar e implantar pruebas de software en cualquier proyecto, o actividades de prueba. La normativa recomienda los siguientes elementos a tomar en cuenta, identificador único, objetivo, prioridad, precondiciones, entradas, resultados esperados y resultados obtenidos, planteando los distintos casos de prueba como se muestra en las tablas 8, 9 y 10.

Menú inicial

Tabla 7

Caso de pruebas menú inicial

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
1	INICIO_001	Verificar maquetación de la pantalla	N/A	N/A	Iniciar la aplicación web	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace)	N/A	Abierto	Alta

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						Cuerpo de la página web con las opciones a) Ilustración representativa (Imagen) b) Información del sistema (Texto) c) Botón iniciar escaneo (Botón) **Pie de página web con las opciones** Marquesina (Texto)			

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
2	INICIO_002	Verificar enlace ACERCA DE	N/A	N/A	Dar click en el enlace ACERCA DE, en el menú principal	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace) h) Botón iniciar escaneo (Enlace)	N/A	Abierto	Baja

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						Cuerpo de la página web con las opciones a) Ilustración igual al logo (Imagen) b) Información del director de tesis y los desarrolladores (Texto) **Pie de página web con las opciones** Marquesina (Texto)			

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
3	INICIO_003	Verificar enlace HISTORIAL	N/A	N/A	Dar click en el enlace HISTORIAL, en el menú principal	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace) h) Botón iniciar escaneo (Enlace)	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						Cuerpo de la página web con las opciones Tabla de datos con la siguiente información a) Id (Identificador) b) Descripción c) Interfaz d) Estado **Pie de página web con las opciones** Marquesina (Texto)			

No	Identificador	Objetivo	Precondición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
4	INICIO_004	Verificar enlace ESCANEAR	N/A	N/A	Dar click en el enlace ESCANEAR, en el menú principal	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace) h) Botón iniciar escaneo (Enlace)	N/A	Abierto	Media

No	Identificador	Objetivo	Precondición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						Cuerpo de la página web con las opciones a) Presione iniciar para comenzar con el escaneo (Texto) b) Iniciar (Botón) Información del escaneo en proceso (Texto) **Pie de página web con las opciones** Marquesina (Texto)			

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
5	INICIO_005	Verificar enlace CONFIGURACION	N/A	N/A	Dar click en el enlace CONFIGURACION, en el menú principal	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace) h) Botón iniciar escaneo (Enlace)	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						Cuerpo de la página web con las opciones a) Configuración (Texto) b) Tabla con las siguientes columnas a. Nombre (Texto) b. Descripción (Texto) c. Valor (Texto) d. Acción (Enlace) **Pie de página web con las opciones** Marquesina (Texto)			

Configuración mail

Tabla 8

Caso de pruebas Configuración e-mail

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
1	CONF_001	Verificar maqueta ción de la pantalla	N/A	N/A	Dar click en el enlace CONFIGU RACION, ubicado en la cabecera de la página web.	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace)	N/A	Abierto	Alta

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						Cuerpo de la página web con las opciones a) Configuración (Texto) b) Tabla con las siguientes columnas a. Nombre (Texto) b. Descripción (Texto) c. Valor (Texto) d. Acción (Enlace) **Pie de página web con las opciones** Marquesina (Texto)			

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
2	CONF_002	Mostrar formulario para editar mail	Presionar el enlace EDITAR ubicado en la tabla	N/A	Dar click en el enlace CONFIGURACION, ubicado en la cabecera de la página web. Dar click en el botón modificar Dar click en el enlace ubicado en la tabla en la sección de acción	**Cuerpo de la página web con las opciones** a) Configuración (Texto) b) Editar parámetro (Texto) c) Nombre (Texto) d) Input Nombre (Campo editable bloqueado) e) Descripción (Texto) f) Input Descripción (Campo editable) g) Mail (Texto) h) Input Mail (Campo editable) i) Registrar (Botón) j) Regresar (Botón)	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
3	CONF_003	Editar mail, dejando los campos vacíos	Presionar el enlace EDITAR ubicado en la tabla	Campo 1: Descripción (Mandar vacío) Campo 2: Mail (Mandar vacío)	Dar click en el enlace CONFIGURACION, ubicado en la cabecera de la página web. Dar click en el botón modificar Dar click en el enlace ubicado en la tabla en la sección de acción Dar click en el botón registrar	**Cuerpo de la página web con las opciones** a) Configuración (Texto) b) Editar parámetro (Texto) c) Nombre (Texto) d) Input Nombre (Campo editable bloqueado) e) Descripción (Texto) f) Input Descripción (Campo editable) g) Mail (Texto) h) Input Mail (Campo editable) i) Registrar (Botón) Regresar (Botón)	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
4	CONF_004	Editar mail, dejando lleno solo el campo 1	Presionar el enlace EDITAR ubicado en la tabla	Campo 1: Descripción (Mandar lleno) Campo 2: Mail (Mandar vacío)	Dar click en el enlace CONFIGURACION, ubicado en la cabecera de la página web. Dar click en el botón modificar Dar click en el botón registrar	**Cuerpo de la página web con las opciones** a) Configuración (Texto) b) Editar parámetro (Texto) c) Nombre (Texto) d) Input Nombre (Campo editable bloqueado) e) Descripción (Texto) f) Input Descripción (Campo editable) g) Mail (Texto) h) Input Mail (Campo editable) i) Registrar (Botón) j) Regresar (Botón)	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
5	CONF_005	Editar mail, dejando los campos llenos	Presionar el enlace EDITAR ubicado en la tabla	Campo 1: Descripción (Mandar vacío) Campo 2: Mail (Mandar lleno)	Dar click en el enlace CONFIGURACION, ubicado en la cabecera de la página web. Dar click en el botón modificar Dar click en el botón registrar	**Cuerpo de la página web con las opciones** a) Configuración (Texto) b) Editar parámetro (Texto) c) Nombre (Texto) d) Input Nombre (Campo editable bloqueado) e) Descripción (Texto) f) Input Descripción (Campo editable) g) Mail (Texto) h) Input Mail (Campo editable) i) Registrar (Botón) j) Regresar (Botón)	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
6	CONF_006	Editar mail, dejando los campos llenos	Presionar el enlace EDITAR ubicado en la tabla	Campo 1: Descripción (Mandar lleno) Campo 2: Mail (Mandar lleno)	Dar click en el enlace CONFIGURACION, ubicado en la cabecera de la página web. Dar click en el botón modificar Dar click en el botón registrar	**Cuerpo de la página web con las opciones** a) Configuración (Texto) b) Editar parámetro (Texto) c) Nombre (Texto) d) Input Nombre (Campo editable bloqueado) e) Descripción (Texto) f) Input Descripción (Campo editable) g) Mail (Texto) h) Input Mail (Campo editable) i) Registrar (Botón) **j)** Regresar (Botón)	N/A	Abierto	Media

Escanear red

Tabla 9

Caso de pruebas Escaneo de red

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
1	ESCN_001	Verificar maquetación de la pantalla	N/A	N/A	Dar click en el enlace ESCANEO, ubicado en la cabecera de la página web.	Muestra una pantalla con la información: **Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace)	N/A	Abierto	Media

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
						g) Inicio (Enlace) h) Botón iniciar escaneo (Enlace) **Cuerpo de la página web con las opciones** a) Presione iniciar para comenzar con el escaneo (Texto) b) Iniciar (Botón) c) Información del escaneo en proceso (Texto) **Pie de página web con las opciones** Marquesina (Texto)			

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
2	ESCN_002	Seleccionar interfaz de red	Iniciar escaneo de red	N/A	Dar click en el enlace ESCANEO, ubicado en la cabecera de la página web. Pulsar el botón de iniciar escaneo	**Cuerpo de la página web con las opciones** a) Seleccione una interfaz de red (Texto) b) Listar en una tabla las interfaces de red encontradas.	N/A	Abierto	Media

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
3	ESCN_003	Escaneo en proceso	Seleccionar una interfaz de red	N/A	Dar click en el enlace ESCANEO, ubicado en la cabecera de la página web. Pulsar el botón de iniciar escaneo Seleccionar una interfaz de red del listado	**Cuerpo de la página web con las opciones** a) Escaneo en proceso (Texto) b) Tabla mostrando los ataques encontrados a. Identificador b. IP atacante c. Interfaz de red d. IP del equipo e. Tipo de ataque f. Fecha c) Resultados (Botón)	N/A	Abierto	Alta

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
4	ESCN_004	Mostrar resultados	Detener escaneo de red	N/A	Dar click en el enlace ESCANEO, ubicado en la cabecera de la página web. Pulsar el botón de iniciar escaneo	**Cuerpo de la página web con las opciones** a) Resultados (Texto) b) Fecha de análisis c) Numero de paquetes analizados	N/A	Abierto	Media

No	Identificador	Descripción	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
					Seleccionar una interfaz de red del listado	a) Tabla mostrando los ataques encontrados a. Identificador b. IP atacante c. Interfaz de red d. IP del equipo e. Cantidad benignos f. Cantidad Ataques de fuerza bruta g. Cantidad DOS h. Cantidad XSS i. Cantidad SQLInyection			

Capitulo IV

Resultados

Implementación de los Algoritmos empleados

Para la implementación de los algoritmos se empleó la herramienta JUPYTER en base a las siguientes métricas para cada algoritmo descrito en la metodología

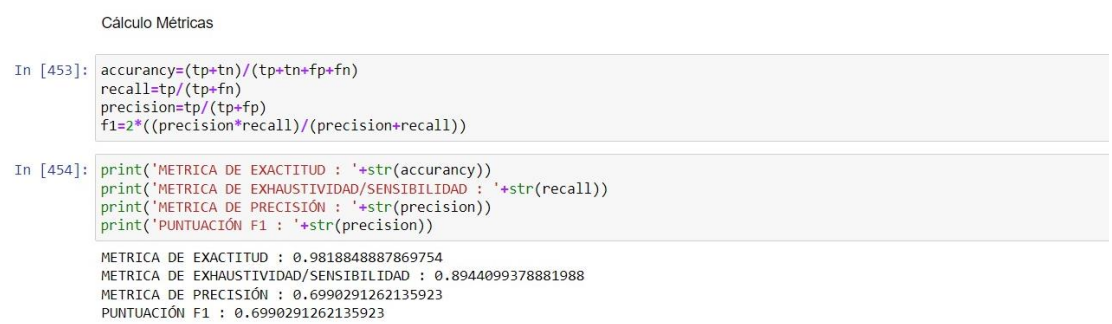
1) Exactitud
2) Sensibilidad
3) Precisión
4) Puntuación

Implementación algoritmo Random Forest

De acuerdo a la codificación realizada, se puede visualizar en la figura 27 los resultados obtenidos del algoritmo con arquitectura de Random Forest.

Figura 26

Implementación algoritmo Random Forest



Implementación algoritmo Decision Tree

De acuerdo a la codificación realizada, se puede visualizar en la figura 28 los resultados obtenidos del algoritmo con arquitectura de Decision Tree

Figura 27

Implementación algoritmo Decision Tree

```
In [57]: accurancy=(tp+tn)/(tp+tn+fp+fn)
         recall=tp/(tp+fn)
         precision=tp/(tp+fp)
         f1=2*((precision*recall)/(precision+recall))

In [58]: print('METRICA DE EXACTITUD : '+str(accurancy))
         print('METRICA DE EXHAUSTIVIDAD/SENSIBILIDAD : '+str(recall))
         print('METRICA DE PRECISIÓN : '+str(precision))
         print('PUNTUACIÓN F1 : '+str(precision))

         METRICA DE EXACTITUD : 0.9803921568627451
         METRICA DE EXHAUSTIVIDAD/SENSIBILIDAD : 0.860248447204969
         METRICA DE PRECISIÓN : 0.6873449131513648
         PUNTUACIÓN F1 : 0.6873449131513648
```

Implementación algoritmo K-means

De acuerdo a la codificación realizada, se puede visualizar en la figura 29 los resultados obtenidos del algoritmo con arquitectura de K-means.

Figura 28

Implementación algoritmo K-means

```
In [227]: accurancy=(tp+tn)/(tp+tn+fp+fn)
          recall=tp/(tp+fn)
          precision=tp/(tp+fp)
          f1=2*((precision*recall)/(precision+recall))

In [228]: print('METRICA DE EXACTITUD : '+str(accurancy))
          print('METRICA DE EXHAUSTIVIDAD/SENSIBILIDAD : '+str(recall))
          print('METRICA DE PRECISIÓN : '+str(precision))
          print('PUNTUACIÓN F1 : '+str(precision))

          METRICA DE EXACTITUD : 0.8292626992317395
          METRICA DE EXHAUSTIVIDAD/SENSIBILIDAD : 0.03881278538812785
          METRICA DE PRECISIÓN : 0.01566820276497696
          PUNTUACIÓN F1 : 0.01566820276497696
```

Implementación algoritmo Neuronal networking

De acuerdo a la codificación realizada, se puede visualizar en la figura 30 los resultados obtenidos del algoritmo con arquitectura de Neuronal networking.

Figura 29

Implementación algoritmo Neuronal networking

Matriz comparativa

Sobre la base de los resultados se puede concluir que el modelo de aprendizaje más óptimo es el “Random Forest” con un valor de Exactitud del 0,9818 (Promedio 0,8185) y el menos eficiente siendo el K-means con un valor de Exactitud del 0,8292 (Promedio 0,2248) como se observa en la tabla 11, asimismo de visualizar en las figuras 28, 29, 30, 31, 32 el desempeño en general de todos los algoritmos, donde los valores de las abscisas pertenecen el porcentaje de efectividad para cada métrica medida en porcentajes.

Tabla 10

Matriz comparativa de Algoritmos vs Métricas

Modelo de aprendizaje	Exactitud	Sensibilidad	Precisión	Puntuación
Decision Tree	0,9803	0,8602	0,6873	0,6873
Random Forest	0,9818	0,8944	0,6990	0,6990
K-means	0,8292	0,0388	0,0156	0,0156
Neuronal networking	0,9450	0,8135	0,3753	0,3753

Figura 30

Comparativo Algoritmos Exactitud

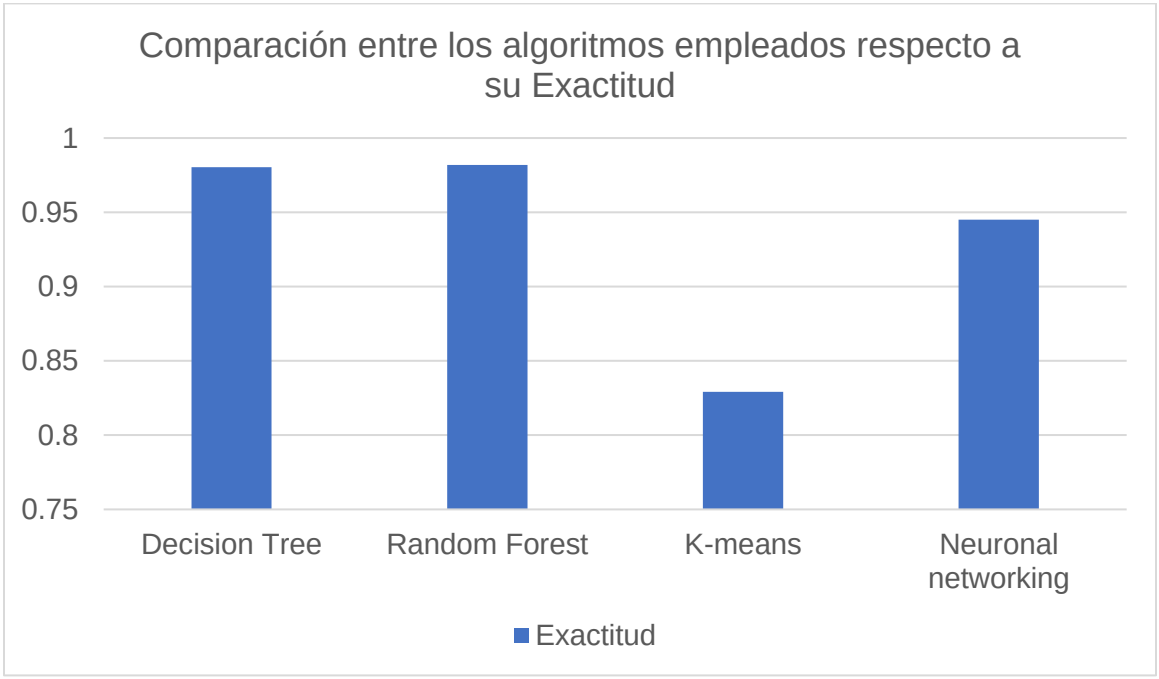


Figura 31

Comparativo Algoritmos Sensibilidad

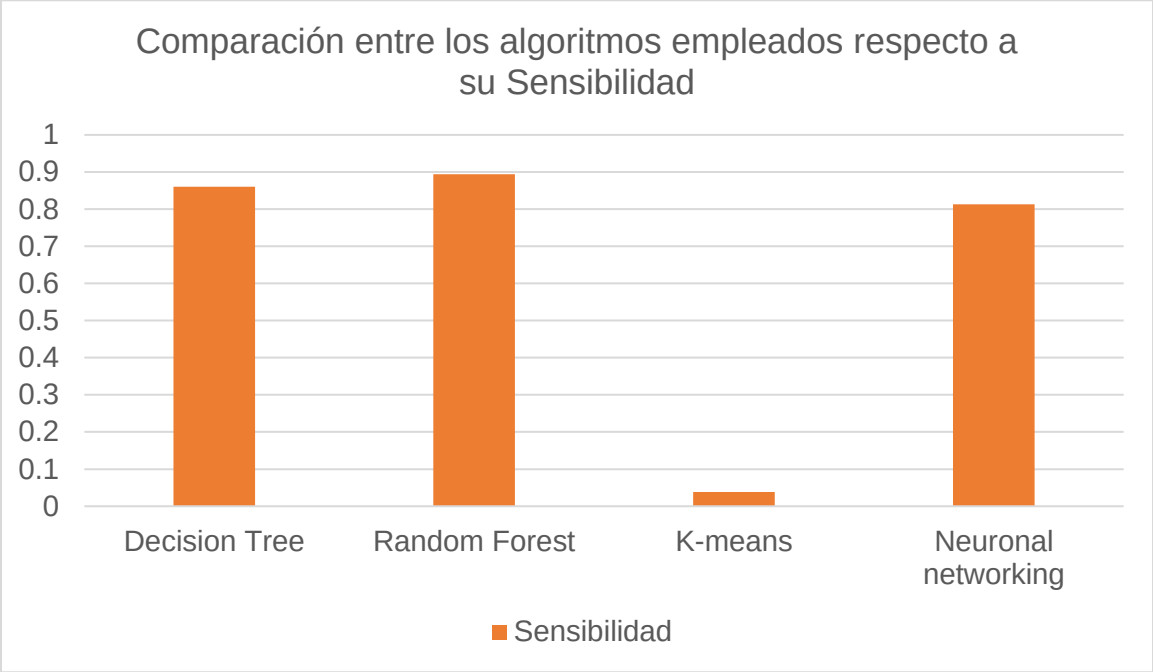


Figura 32

Comparativo Algoritmos Precisión

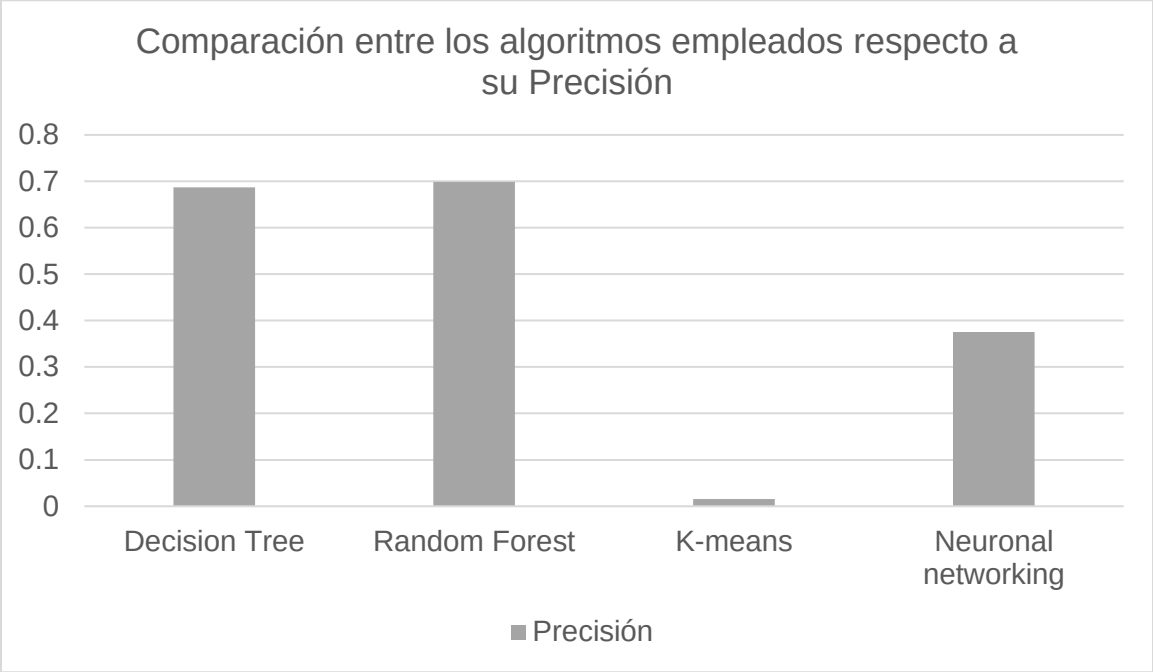


Figura 33

Comparativo Algoritmos Puntuación

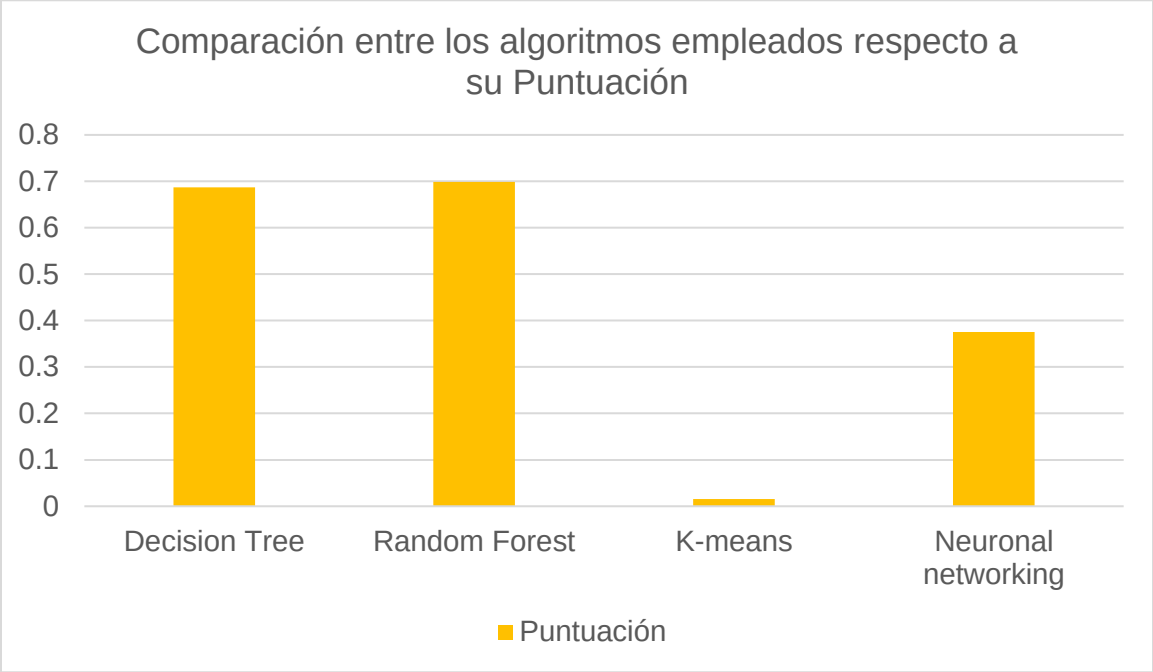
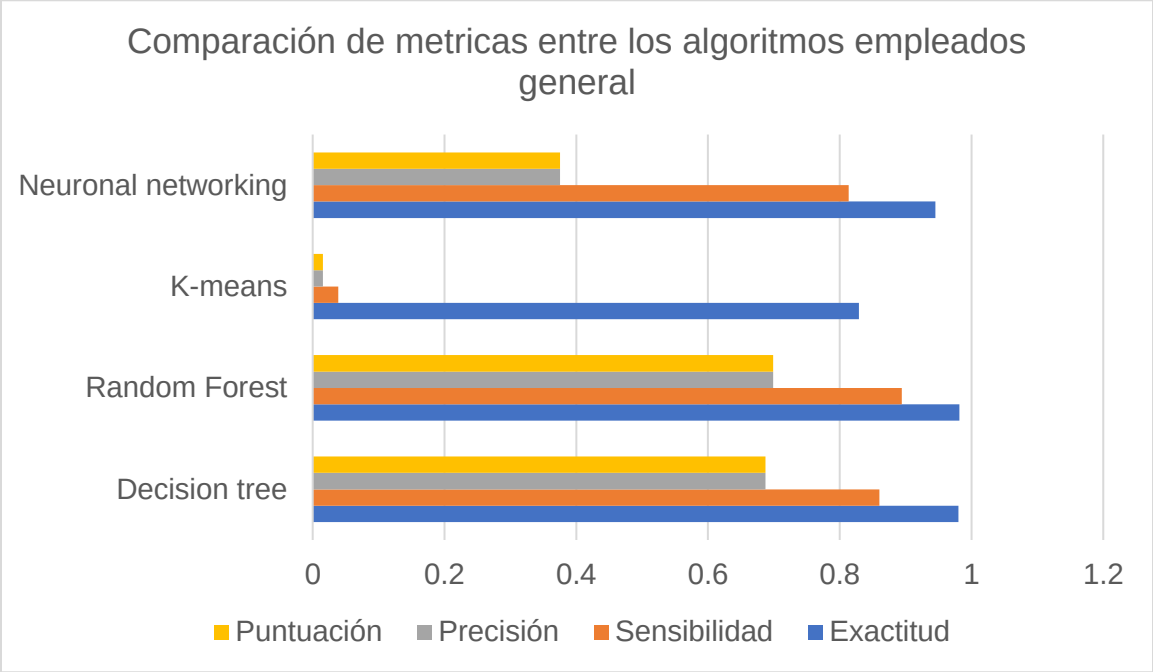


Figura 34

Comparativo Algoritmos en general



Implementación de la Aplicación web

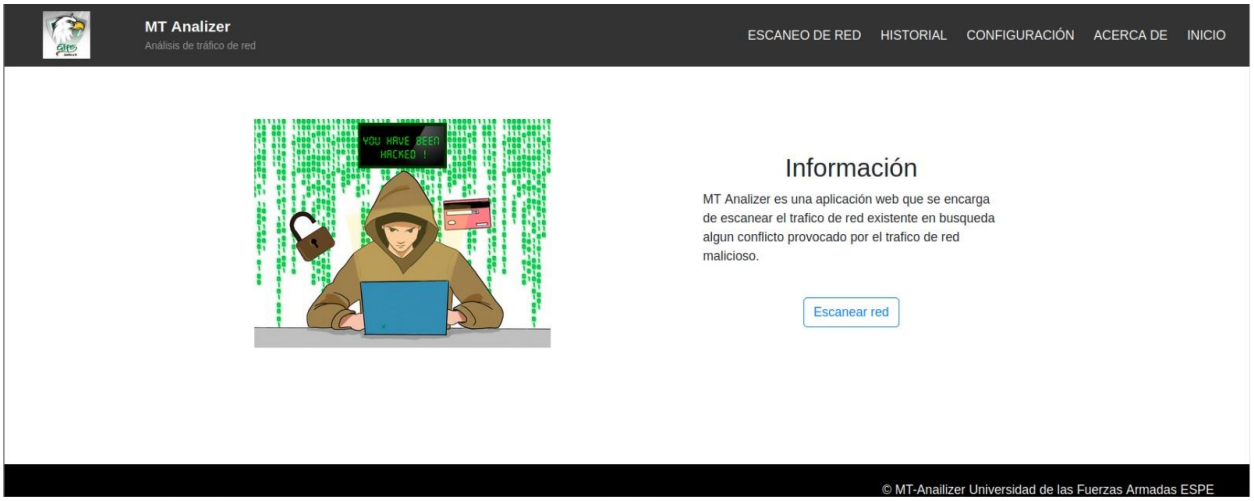
La aplicación web, cuenta con las siguientes interfaces para su funcionamiento.

Interfaz Menú de inicio

Detalla la información de la aplicación con su respectivo menú navegable, como se muestra en la figura 36.

Figura 35

Interfaz Menú de inicio

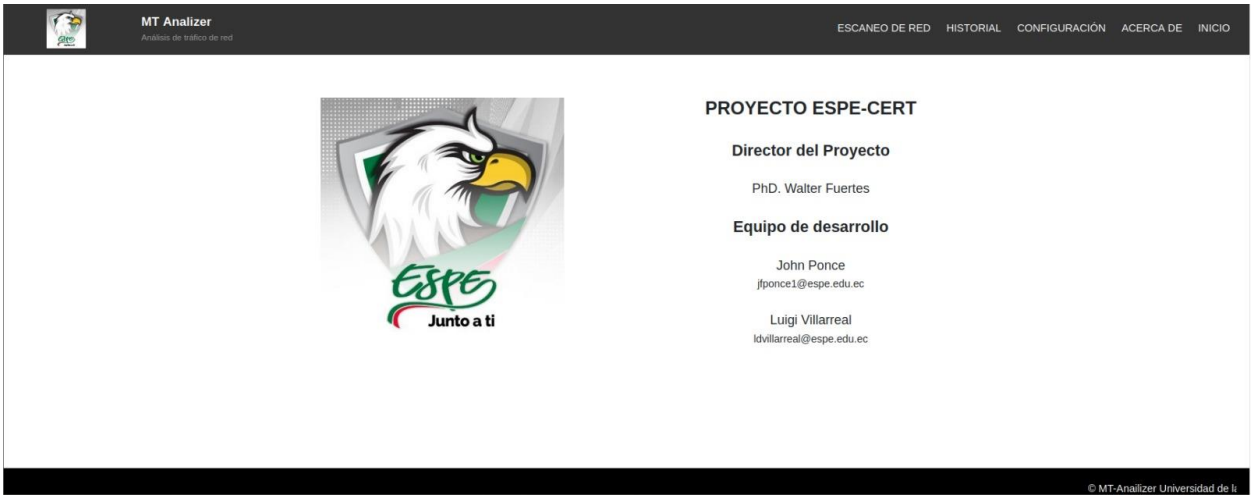


Interfaz Información desarrolladores

Muestra la información del director de proyecto junto a los desarrolladores de la aplicación, como se muestra en la figura 37.

Figura 36

Interfaz Información director y desarrolladores

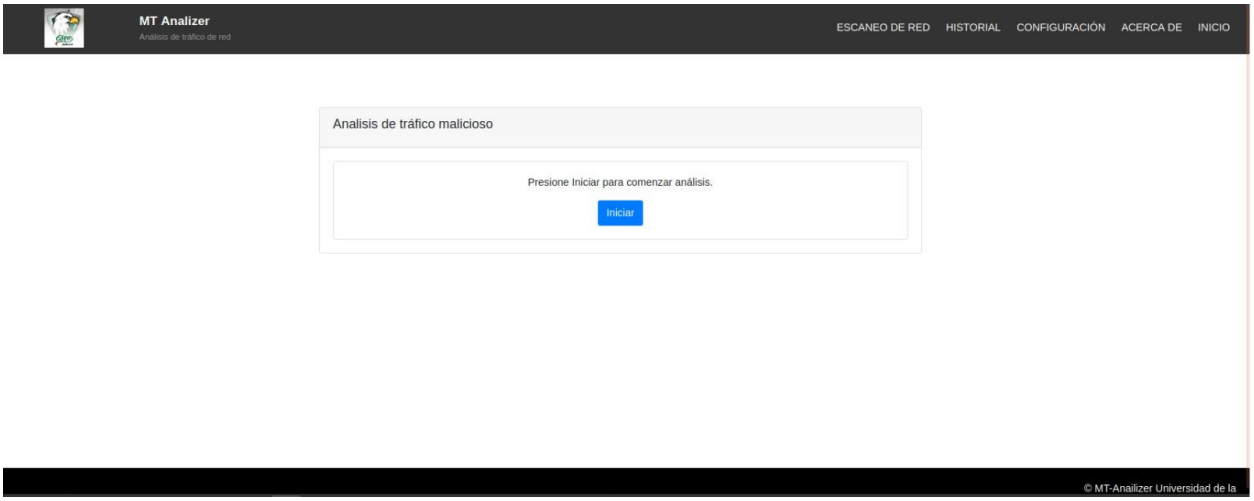


Interfaz Escaneo de puertos

Permite realizar el escaneo de interfaces al presionar INICIAR, como se muestra en la figura 38.

Figura 37

Interfaz Escaneo de puertos

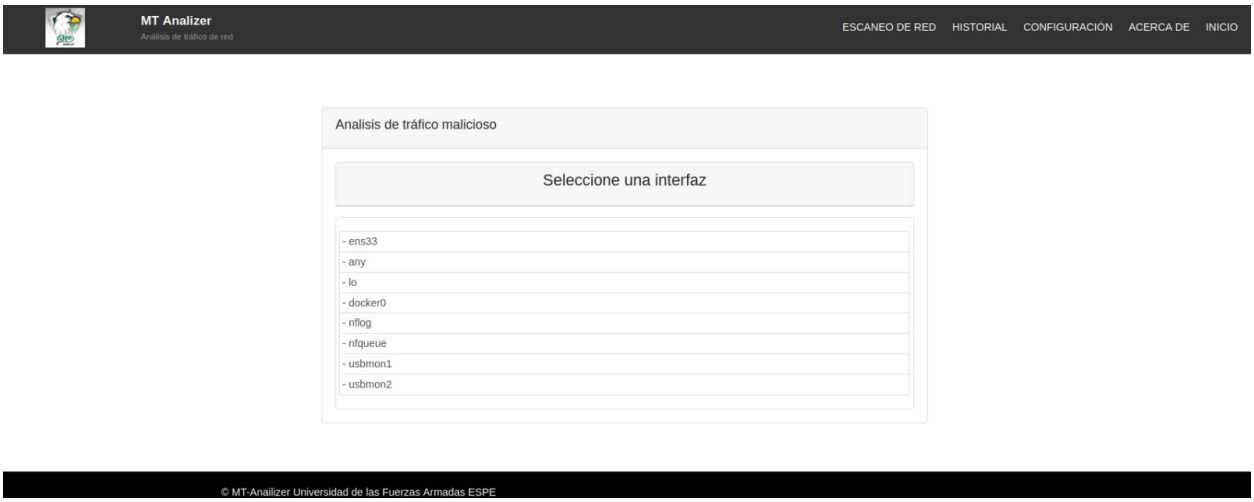


Interfaz Seleccionar interfaz

Se selecciona la interfaz analizar, como se muestra en la figura 39.

Figura 38

Interfaz Seleccionar interfaz

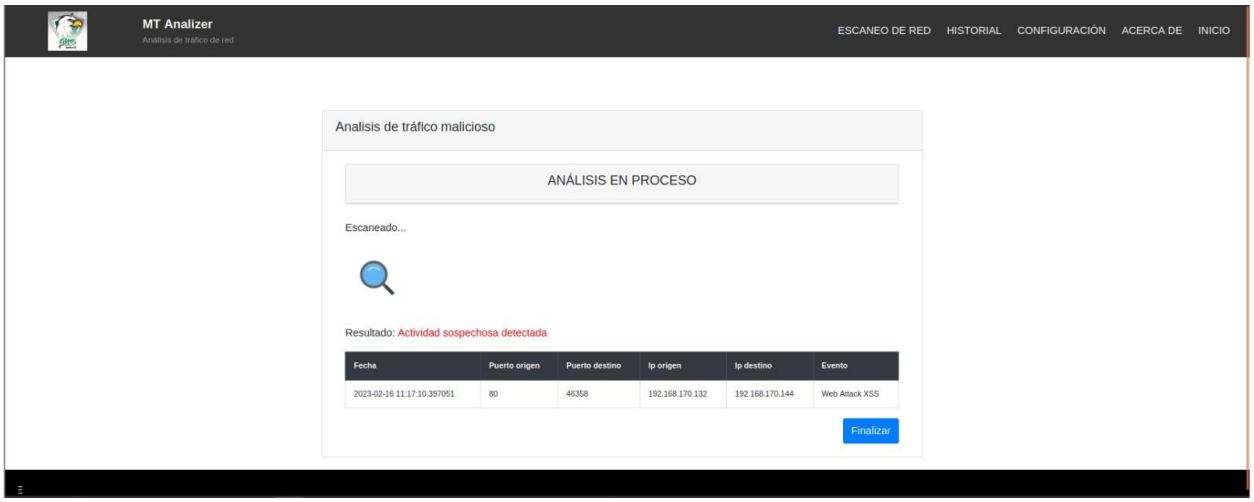


Interfaz Escaneo en proceso

La aplicación se encarga de escanear el tráfico de dicha interfaz en segundo plano, para finalizar el escaneo se debe pulsar el botón de FINALIZAR, como se muestra en la figura 40.

Figura 39

Interfaz Escaneo en proceso

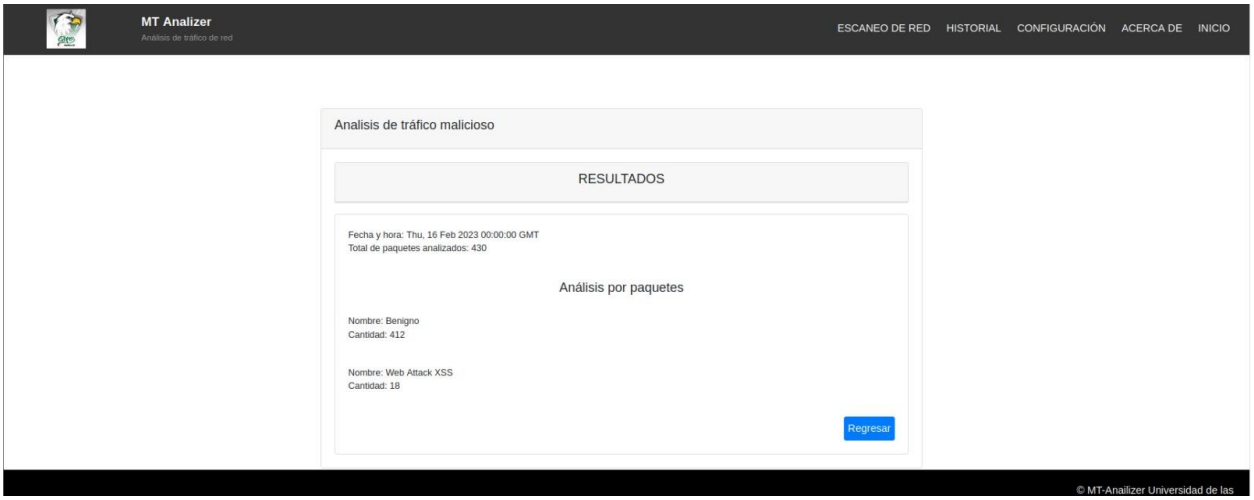


Interfaz Mostrar resultados

Se detalla toda la información encontrada del análisis en la interfaz antes seleccionada, como se muestra en la figura 41.

Figura 40

Interfaz Mostrar resultados



Interfaz Mostrar historial

Se muestra en una tabla a breves rasgos la información del escaneo realizado previamente. Al presionar el enlace ubicado en acciones se muestra una tabla mucho más detallada con la información de todos los ataques encontrados en dicho escaneo, como se muestra en la figura 42 y 43.

Figura 41

Interfaz Mostrar historial 1/2

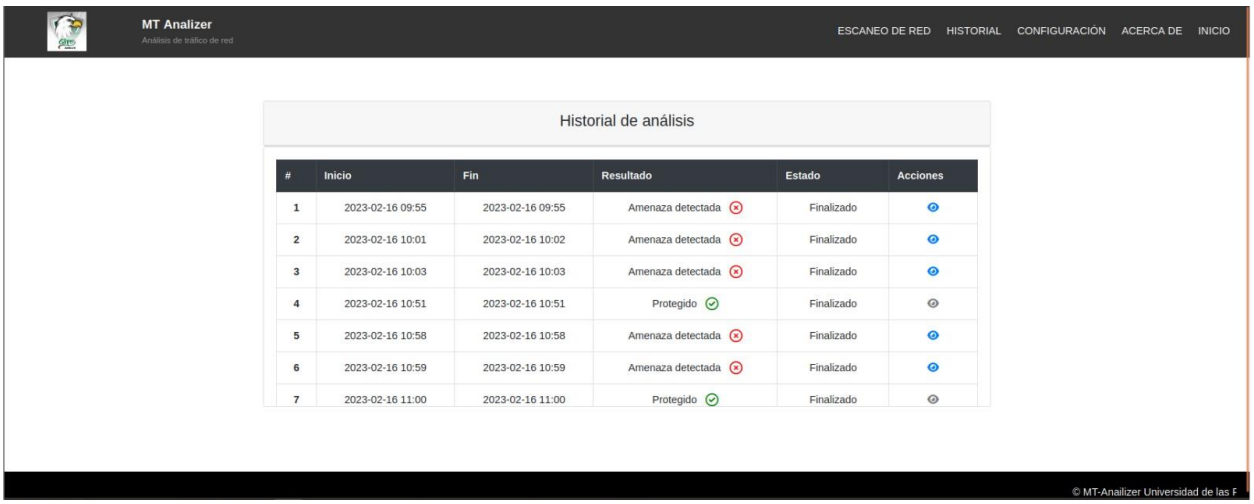
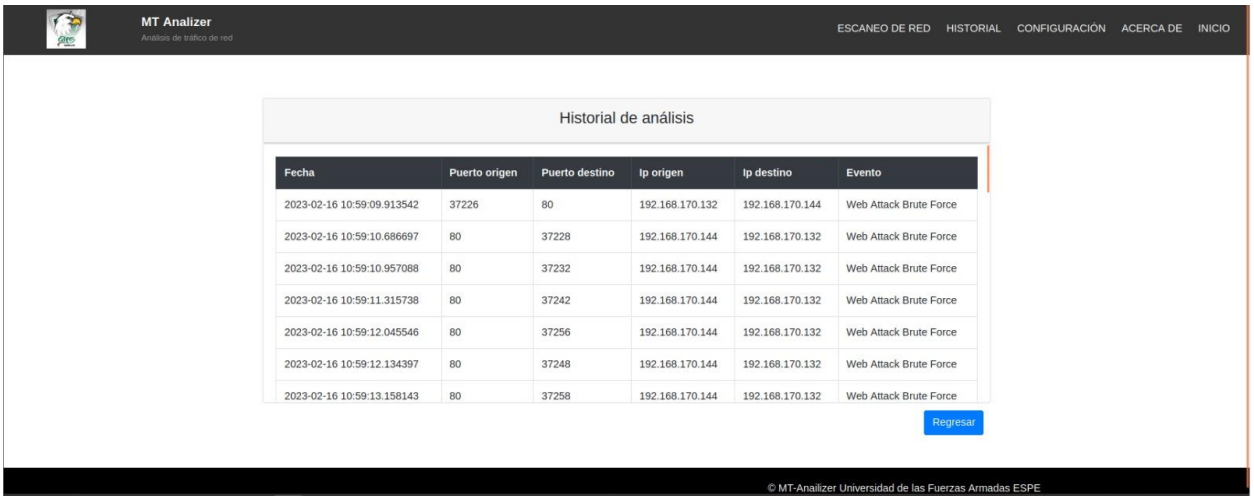


Figura 42

Interfaz Mostrar historial 2/2



Interfaz Configuración

Se muestra en una tabla con la información del mail donde se desea notificar en caso encuentre alguna anomalía al realizar un escaneo de red. Al presionar en editar se puede modificar el mail receptor de las notificaciones y la descripción del mismo, como se muestra en la figura 44 y 45.

Figura 43

Interfaz Configuración 1/2

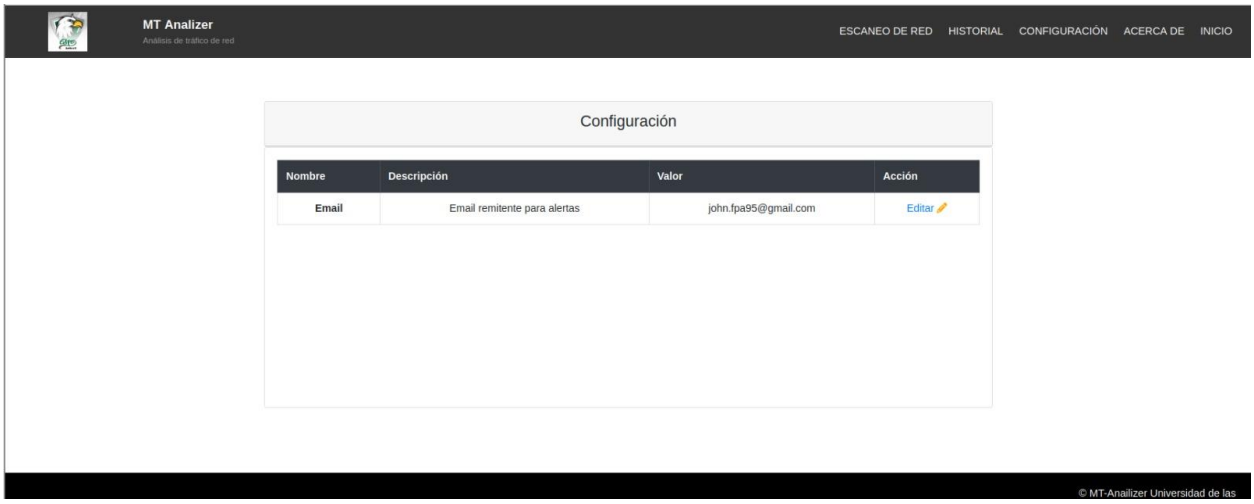
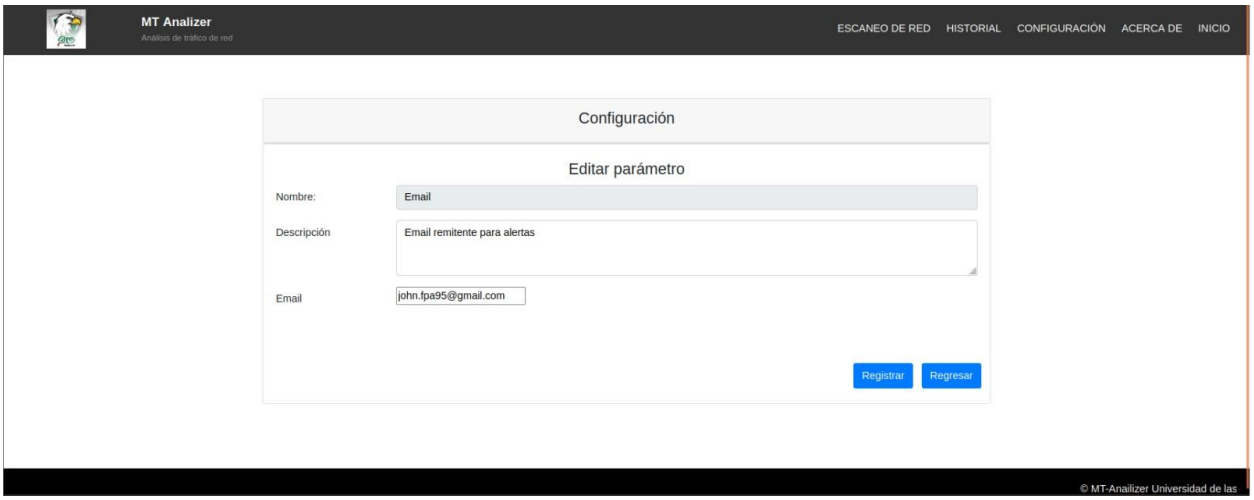


Figura 44

Interfaz Configuración 2/2



Encuesta de satisfacción al cliente

La cantidad de personas encuestadas fueron 7, tabulando las siguientes preguntas de manera cuantitativa.

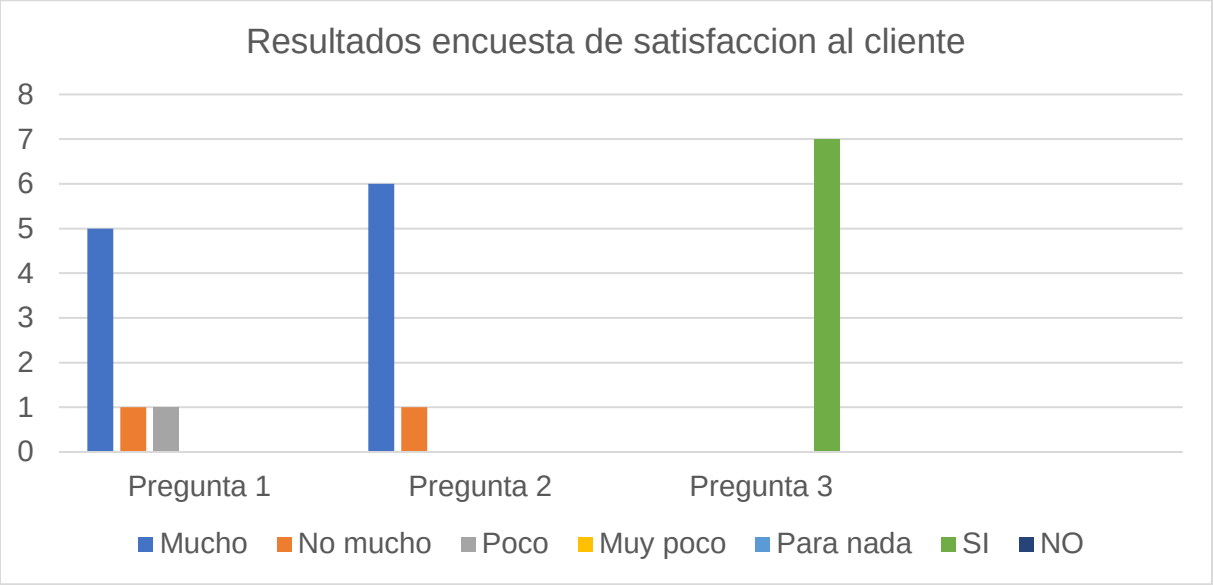
a) Pregunta 1: ¿El diseño de la aplicación es de su agrado?

b) Pregunta 2: ¿La aplicación cumple con sus expectativas?

c) Pregunta 3: ¿Recomendaría esta aplicación a sus conocidos?

Figura 45

Resultados encuesta de satisfacción al cliente



Mientras que las respuestas a las preguntas abiertas fueron las siguientes.

d) Pregunta 1: En cuanto al diseño, ¿Qué modificaciones propone?

1) Ninguna.
2) Al seleccionar la interfaz aparte del nombre de la red se debería mostrar la IP.
3) En la pantalla de resultados usar tablas para mostrar los detalles de amenazas encontradas para mantener el mismo diseño que el historial.
4) Ninguna.
5) Ninguna.
6) Hay varias secciones en la pantalla donde no se está tomando todo el espacio dejando grandes vacíos que podrían ser rellenados por instrucciones de cómo usar las distintas funcionalidades de la aplicación web.
7) Cuando termina de escanear y se pulsa el botón de FINALIZAR los resultados finales se deberían mostrar en una tabla de manera más organizada.

e) Pregunta 2: ¿Qué funcionalidades le gustaría agregar o modificar?

1) Ninguna.
2) Al momento de que el escaneo se encuentra en curso y detecta una amenaza, se debería implementar un botón que permita deshabilitar dicha interfaz para impedir que sigan atacando a la misma.
3) Agregar más opciones para notificar las amenazas como puede ser notificar por WhatsApp.
4) Ninguna.
5) Ninguna.
6) Implementar un enlace a esta encuesta directamente desde la aplicación.
7) En la sección de configuración se bebería agregar más de un mail para remitir las notificaciones.

Proyecto

Para acceder al proyecto completo entrar al siguiente enlace

https://github.com/Skairlex/ESPE-DMT.git

Manual técnico

Para acceder al proyecto completo entrar al siguiente enlace

https://docs.google.com/document/d/1eT-9W1O7KQh4LrlR0OPFgEJhi6217Jgp2ZRwO9qXb2k/edit?usp=share_link

Manual de usuario

Para acceder al proyecto completo entrar al siguiente enlace

https://docs.google.com/document/d/1FKVIRUpF-3RyGWRqhDOcO6nNItAjqaVi/edit?usp=sharing&ouid=108994427415641171638&rtpof=true&sd=true

Capitulo V

Conclusiones y Recomendaciones

Conclusiones

a) La precisión de los algoritmos depende de la conformación y procesamiento del dataset por lo que fue necesario realizar diversas pruebas e iteraciones hasta conseguir el mejor resultado. Al evaluar los algoritmos de inteligencia artificial, se determinó que el más óptimo de los mismos, para la clasificación de flujos de paquetes de red, de acuerdo a las métricas planteadas, es Random Forest, con un valor de Exactutud del 98,18% y el menos óptimo fue el K-means con un porcentaje de 82,92%.

b) Python demostró ser un lenguaje de programación completo que permitió desarrollar la funcionalidad para capturar el tráfico de red, algoritmos de inteligencia artificial e implementar la aplicación web con sus servicios. Asimismo, SCRUM y sus herramientas complementarias, como el planning pocker y el análisis PERT, ayudó a estimar tiempos con sus respectivos puntajes, los cuales fueron necesarios para cumplir con las funcionalidades propuestas. También se aplicó la arquitectura por capas, lo que permitió dividir la aplicación por módulos, esto a su vez ayuda a la organización de los diferentes componentes que conforman el proyecto desarrollado. La metodología Design Science Research, permitió generar un artefacto para el análisis del tráfico de red, mediante los modelos generados, por los diferentes algoritmos de inteligencia artificial. Al mismo tiempo, ayudo a generar la información relevante, como es el comparativo de los mismos para la clasificación de los distintos tipos de flujos de red.

c) Para realizar las pruebas funcionales y no funcionales, fue necesario aplicar las normativas ISO 9001 para la usabilidad y la ISO 29119 para pruebas de software. Lo que permitió trabajar con herramientas como las encuestas, las cuales se conforman por cinco preguntas. Tres de ellas permitieron medir la usabilidad a través de puntajes y

las restantes permitieron conocer la opinión de la aplicación lo que incluye mejoras que podrán ser analizadas y si es necesario se desarrollarán para una futura versión. Además, se aplicó los casos de pruebas para comparar los resultados esperados con los obtenidos. Lo que permitió trabajar en las funcionalidades desarrolladas hasta obtener resultados favorables y completar todos los ítems de la matriz propuesta. Gracias a estas pruebas se aprobó la instalación del aplicativo en el equipo de administración del CERT académico.

d) El manual técnico y de usuario fueron entregados al CERT académico, para su uso y divulgación de ser necesario. Asimismo, pueden ser accedidos a través del link ubicado en la sección de los resultados (Manual técnico y Manual de usuario), los cuales detallan la estructura y funcionamiento de la aplicación diseñada específicamente para usuarios nuevos.

Recomendaciones

a) El dataset debe ser debidamente procesado, de acuerdo a la sección de Preparación, para generar resultados que no proporcionen incoherencias en la predicción de nuevos datos. Igualmente, los algoritmos deben ser comprobados mediante una parte de la información para evitar el sobre entrenamiento.

b) Al trabajar con Python, se debe buscar frameworks y librerías que faciliten la codificación del aplicativo y ayuden a completar el desarrollo en un menor tiempo. Igualmente, se recomienda el manejo de JIRA u otro software similar que permita gestionar la administración de historias y tareas para completarlas en el tiempo propuesto. Igualmente es necesario contar con una arquitectura que ayude a organizar los componentes del sistema, porque esto facilita realizar cambios en el aplicativo.

c) Para las pruebas funcionales y no funcionales, se recomienda buscar una normativa que se adapte mejor al producto creado y que documente como realizar los procesos necesarios, con el fin de proporcionar calidad en el entregable final.
d) Se recomienda revisar los manuales con los clientes, con el fin de comprobar que la información sea clara para los mismos.

Bibliografía

Aguirre Ponce, A. A. (2017). Ciberseguridad en infraestructura críticas de información. *Buenos Aires: Universidad de Buenos Aires. Facultad de Ciencias Económicas.*

Andrade, R. O., Fuertes, W., Cadena, S., Cadena, A., & Tello-Oquendo, L. (2019). Information Security Management in University Campus Using Cognitive Security. International Journal of Computer Science and Security (IJCSS), *124-134.*

Andrade, R., & Fuertes, W. (2013). Diseño y dimensionamiento de un equipo de un equipo de respuesta ante incidentes de seguridad informática (CSIRT). Caso de estudio: ESPE. *Sangolquí.*

aprendeIA*. (2022). Obtenido de aprendeIA: https://aprendeia.com/algoritmo-kmeans-clustering-machine-learning/*

arimetrics*. (2022). Obtenido de arimetrics: https://www.arimetrics.com/glosario-digital/malware*

arrobasystem. (2022). ¿Qué es Seguridad Informática? *Obtenido de @arrobasystem: https://arrobasystem.com/pages/seguridad-informatica*

Belcic, I. (2019, Septiembre 28). ¿Qué es el malware? *Obtenido de AVAST: https://www.avast.com/es-es/c-malware*

Bo, H., Atsutoshi, K., Kazunori, K., Kenji, T., Daniel, D., Ola, S., . . . Akihiro, N. (2019). Alchemy: Stochastic Feature Regeneration for Malicious Network Traffic Classification. IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), *346 - 351.*

Cabezón, V. B. (2023). CERT - UAM*. Obtenido de CERT - UAM: https://www.uam.es/uam/vida-uam/cert*

Canelo, M. M. (2022). CELERY*. Obtenido de https://profile.es/blog/que-es-celery/#:~:text=Celery%20es%20una%20biblioteca%20de,la%20ejecución%20normal%20del%20programa.*

Castro, M. V. (2021). FLASK API REST*. Obtenido de https://misovirtual.virtual.uniandes.edu.co/codelabs/tutorial-flask-api-rest/index.html?index=..%2F..index#2*

CCN-CERT. (2023, 01 30). DEFENSA FRENTE A LAS CIBERAMENAZAS*. Obtenido de https://www.ccn-cert.cni.es/sobre-nosotros/servicios-ccn-cert.html*

CERT Ecuador/CC*. (2023). Obtenido de CERT Ecuador/CC: https://sites.google.com/site/certecuadorcc/*

CHANG, J. E. (2020). Análisis de ataques cibernéticos hacia el Ecuador. Editora Adjunta*, 01-18.*

Cuofano, G. (2022). What Is A Business Logic Layer? *Obtenido de https://fourweekmba.com/es/capa-de-lógica-empresarial/*

Digital Guide IONOS*. (2020, 03 10). Obtenido de Digital Guide IONOS: https://www.ionos.es/digitalguide/online-marketing/marketing-para-motores-de-busqueda/que-es-una-neural-network/*

Dresch, A., Lacerda, D. P., & Antunes, J. A. (2015). Design science research. Springer*, 067-102.*

Education, I. C. (2020, Agosto 17). Redes neuronales*. Obtenido de https://www.ibm.com/es-es/cloud/learn/neural-networks*

Education, I. C. (2021). Clústeres de k-medias*. Obtenido de https://www.ibm.com/docs/es/db2/11.5?topic=building-k-means-clustering*

elConspirador. (2013, Diciembre 21). Qué es y para qué sirve un modelo conceptual*. Obtenido de https://www.elconspirador.com/2013/12/21/que-es-y-para-que-sirve-un-modelo-conceptual/*

Estefanía, M., & Chunata, C. (2019). Análisis de las metodologías enisa y apcert para la creación del centro de respuesta a incidentes informáticos (csirt). caso práctico: prototipo de un CSIRT en la Universidad Nacional de Chimborazo. *Chimborazo: Universidad Nacional de Chimborazo,2019.*

Europeo, P. (2020, Agosto 27). ¿Qué es la inteligencia artificial y cómo se usa? *Obtenido de europa.eu: https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa*

Figueiras, S. (2021, Septiembre 20). ¿CONOCES JUPYTER NOTEBOOK? *Obtenido de https://www.ceupe.mx/blog/conoces-jupyter-notebook.html*

Fuertes, W., Reyes, F., Valladares, P., Tapia, F., Toulkeridis, T., & Pérez, E. (2017). An integral model to provide reactive and proactive services in an academic csirt based on business intelligence. *Quito: Multidisciplinary Digital Publishing Institute.*

Guacho Morocho, D. D. (2014). Diseño de un Sistema de Seguridad de la Información para EcuCERT. *Quito: Quito: EPN, 2014.*

Guayara Rubio, A. (2018). Guía de recomendación buenas prácticas basadas en las metodologías SDL, CBYC para desarrollo de software seguro apoyado en lineamientos de la fundación OWASP y estándares de la división CERT del instituto SEI de la Universidad Carnegie Mellon. *Bogota: Universidad Nacional Abierta ya Distancia UNAD.*

Heba, S., & Zouheir, T. (2019). Enhancing firewall filter performance using neural networks. 15th International Wireless Communications & Mobile Computing Conference (IWCMC), *1853 - 1859.*

Heras, J. M. (2020, 09 18). IArtificial.net*. Obtenido de IArtificial.net: https://www.iartificial.net/random-forest-bosque-aleatorio/*

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. Design research in information systems*, 9-22.*

Householder, A. D., Wassermann, G., Manion, A., & King, C. (2017). The cert guide to coordinated vulnerability disclosure. Carnegie-Mellon Univ Pittsburgh Pa Pittsburgh United States*, 000-121.*

IBM. (2020). Bosque*. Obtenido de https://www.ibm.com/docs/es/iis/11.5?topic=components-services-tier*

IBM. (2021, Febrero 28). Capa de servicios*. Obtenido de https://www.ibm.com/docs/es/iis/11.5?topic=components-services-tier*

IBM. (2022). ¿Qué es Machine Learning? *Obtenido de https://www.ibm.com/mx-es/analytics/machine-learning*

IBM. (2022). ¿Qué es un árbol de decisión? *Obtenido de https://www.ibm.com/es-es/topics/decision-trees*

IBM. (2022). What is random forest? *Obtenido de https://www.ibm.com/topics/random-forest*

IBM, ¿Qué es un arbol de decicion? *(2022). Obtenido de IBM: https://www.ibm.com/es-es/topics/decision-trees#:~:text=Un%20árbol%20de%20decisión%20es,nodos%20internos%20y%20nodos%20hoja.*

incibe-cert*. (2023). Obtenido de incibe-cert: https://www.incibe-cert.es/servicios-operadores*

Kaspersky*. (2023). Obtenido de Kaspersky: https://latam.kaspersky.com/resource-center/preemptive-safety/what-is-malware-and-how-to-protect-against-it*

KeepCoding, R. (2023, Enero 13). ¿Qué es un modelo de datos físicos? *Obtenido de https://keepcoding.io/blog/modelo-de-datos-fisicos/*

Kirsch, J. H. (2018). IBM*. Obtenido de https://www.ibm.com/mx-es/analytics/machine-learning*

Lashkari, A. H. (2020). CYBERSECURITY DATASETS AND OPEN SOURCE PROJECTS*. Obtenido de https://www.cs.unb.ca/~alashkar/Data-sets.asp*

Lima, I. R., de Castro Freire, T., & Costa, H. A. (2012). Adapting and using scrum in a software research and development laboratory. *16-23: Revista de Sistemas de Informação da FSMA.*

Mariposa, P. D. (2023, Enero 27). ¿Qué es una matriz de confusión en el aprendizaje automático? *Obtenido de https://geekflare.com/es/confusion-matrix-in-machine-learning/*

Martínez, C. V. (2023). UNAM. *Obtenido de UNAM: http://www.ii.unam.mx/es-mx/AlmacenDigital/CapsulasTI/Paginas/seguridadinformatica.aspx*

Martín-Navarro, A., Sancho, M. P., & Medina-Garrido, J. A. (2018). BPMS para la gestión: una revisión sistemática de la literatura. Journal Article*, 1988-4621.*

Medina, M. C. (2022, Mayo 27). La ciberseguridad en el futuro tecnológico tras el Covid19*. Obtenido de ENAE: https://www.enae.es/blog/la-ciberseguridad-en-el-futuro-tecnologico-tras-el-covid19?gclid=Cj0KCQiAorKfBhC0ARIsAHDzslspuMFLQXLmQR3YOU6k62mhiAHHf94VGWjfCeZSerPNN80QkKtX1K4aApLVEALw_wcB&_adin=11551547647*

Meng, Q., Pang, X., Zheng, Y., Jiang, G., & Tian, X. (2021). Development and Optimization of Software Defined Networking Anomaly Detection Architecture by GRU-CNN under Deep Learning. 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 828 - 834.*

Narvaez, M. (2023). ¿Cómo medir la satisfacción del cliente según ISO 9001? *Obtenido de QuestionPro: https://www.questionpro.com/blog/es/como-medir-la-satisfaccion-del-cliente-segun-iso-9001/*

Navid, P. S., & Hamed, M. (2020). Providing a hybrid approach for detecting malicious traffic on the computer networks using convolutional neural networks. 28th Iranian Conference on Electrical Engineering (ICEE)*, 1 - 6.*

NETAPP*. (2022). Obtenido de https://www.netapp.com/es/artificial-intelligence/what-is-artificial-intelligence/*

Núñez, G. S. (2018, Mayo 10). Detrás de la magia del aprendizaje automático (machine learning)*. Obtenido de https://medium.com/@torenunez/detrás-de-la-magia-del-aprendizaje-automático-machine-learning-18c528f1901d*

Pressman, R. S. (2005). Software engineering: a practitioner's approach. *Palgrave macmillan.*

Pressman, R. S. (2010). Ingeniería de software un enfoque práctico. En R. S. Pressman, Ingeniería de software un enfoque práctico *(págs. 113-115). Madrid: Septima.*

Roche, J. (2022). Las 5 ceremonias Scrum: claves para la gestión de procesos*. Obtenido de https://www2.deloitte.com/es/es/pages/technology/articles/ceremonias-scrum.html*

Rodas Neira, S. N., & Villalva Ayala, L. A. (2019). Análisis del impacto de las metodologías ágiles en el desarrollo de software. *Ecuador: UNIVERSIDAD ESTATAL DE MILAGRO.*

Sahoo, K. S., Tripathy, B. K., Naik, K., Ramasubbareddy, S., Balusamy, B., Khari, M., & Burgos, D. (2020). An evolutionary SVM model for DDOS attack detection in software defined networks. IEEE Access*, 2169-3536.*

Shin, T. (2020, Mayo 20). Comprensión de la Matriz de Confusión y Cómo Implementarla en Python*. Obtenido de https://www.datasource.ai/es/data-science-articles/comprension-de-la-matriz-de-confusion-y-como-implementarla-en-python*

Tanczer, L. M., Brass, I., & Carr, M. (2018). CSIRT s and Global Cybersecurity: How Technical Experts Support Science Diplomacy. Global Policy*, 1758-5880.*

Tebes, G., Peppino, D., Rivera, M. B., Becker, P., Papa, F., & Olsina, L. (2020). Especificación del Proceso de Design Science Research: Caso Aplicado a una Ontología de Testing de Software. *Buenos Aires: Universidad Nacional de La Matanza UNLaM.*

TIBCO. (2020). ¿Qué es un modelo de datos lógico? *Obtenido de https://www.tibco.com/es/reference-center/what-is-a-logical-data-model*

Tuya, J. (2009). El futuro estándar ISO/IEC 29119 - Software Testing. *Madrid: REICIS.*

Ucha, F. (2022, Marzo). Definición de Flujograma*. Obtenido de https://www.definicionabc.com/general/flujograma.php*

Unir. (2021, 06 15). ¿Qué es la seguridad informática y cuáles son sus tipos? *Obtenido de https://ecuador.unir.net/actualidad-unir/que-es-seguridad-informatica/*

Unir. (2021, 06 15). ¿Qué es la seguridad informática y cuáles son sus tipos? *Obtenido de https://ecuador.unir.net/actualidad-unir/que-es-seguridad-informatica/*

Vicente, C., & Rafael, L. (2020). Implementación de un equipo de respuesta a incidentes de seguridad informática (CSIRT) en la Fiscalía General del Estado. *Quito: Universidad Internacional SEK.*

Vivanco Toala, D., & Chilán, G. (2019). DESARROLLO DE UN OBSERVATORIO TECNOLÓGICO ENFOCADO A LA SEGURIDAD DE LA INFORMACIÓN PARA INSTITUCIONES DE EDUCACIÓN SUPERIOR (IES). *Quito: INSTITUCIONES DE EDUCACIÓN SUPERIOR (IES).*

Yuwei, S., Hiroshi, E., & Hideya, O. (2020). Adaptive intrusion detection in the networking of large-scale lans with segmented federated learning. IEEE Open Journal of the Communications Society, *102 - 112.*