ANOMALÍAS EN REDES INFORMÁTICAS

PRESENTADO POR:

- Lady Patricia Mamani Charca
- Miriam Mamani Mamani
- Azael Jholyem Mayta Quispe

Desde el momento en que una computadora se conecta a Internet, se abren ante los usuarios toda una nueva serie de posibilidades; sin embargo, éstas traen consigo toda una serie de nuevos y, en ocasiones complejos tipos de ataque o intrusiones. Dichos ataques pueden ser caracterizados como anomalías en el comportamiento usual del flujo de datos en dicha red de comunicaciones.

IDS (Sistema de detección de intrusos)

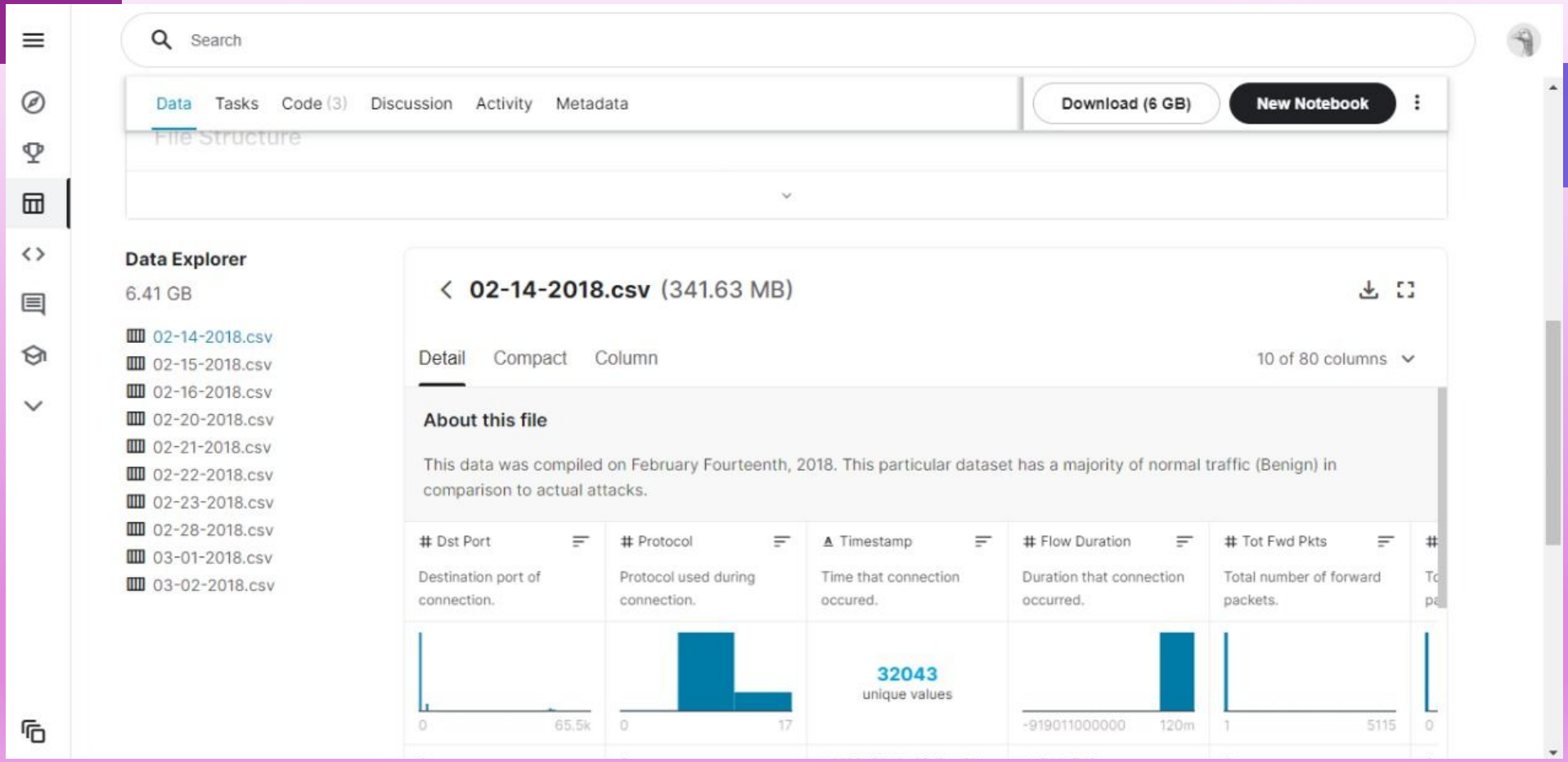
- Es una herramienta de seguridad encargada de monitorear los eventos que ocurren en un sistema informático en busca de intentos de intrusión. Se define intento de intrusión como cualquier intento de comprometer la confidencialidad, integridad, disponibilidad o evitar los mecanismos de seguridad de una computadora o red.

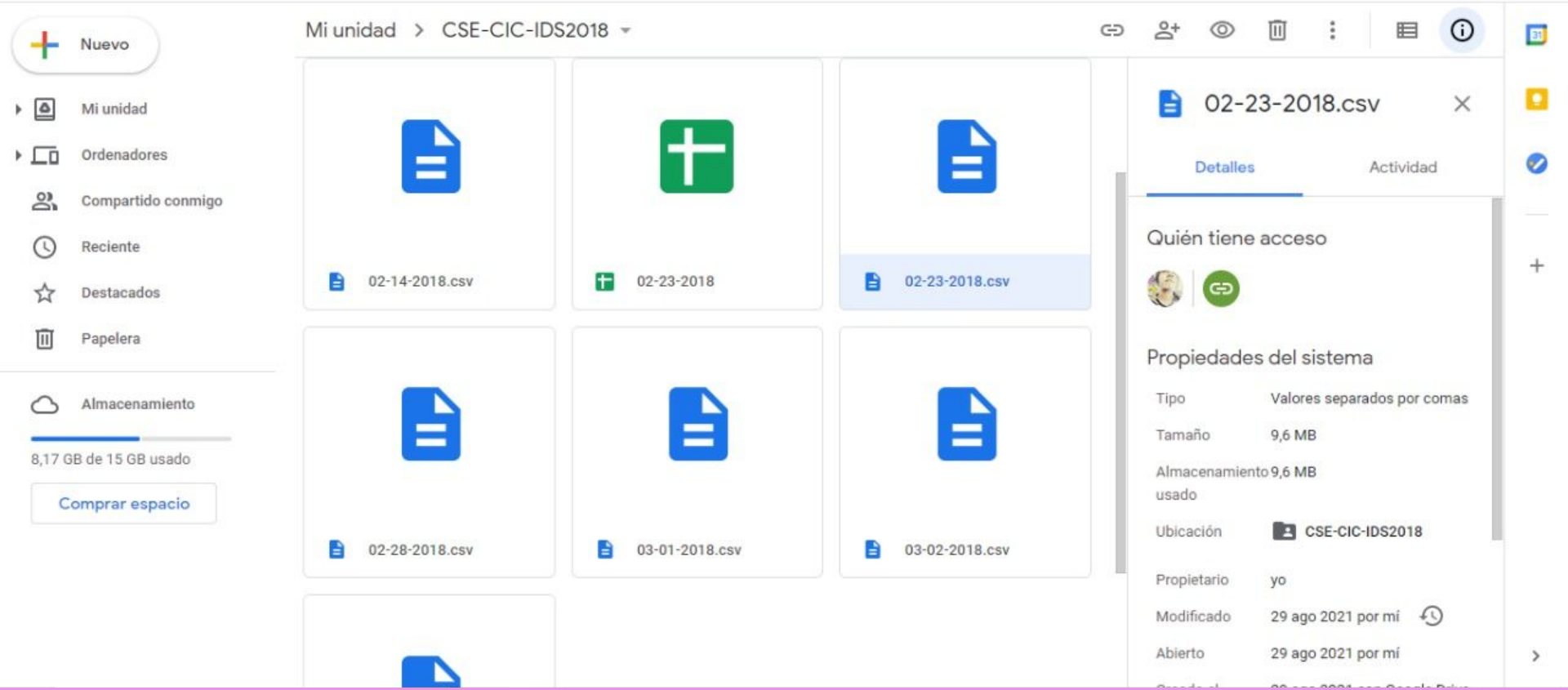
NIDS (Sistema de Detección de Intrusiones en Redes)

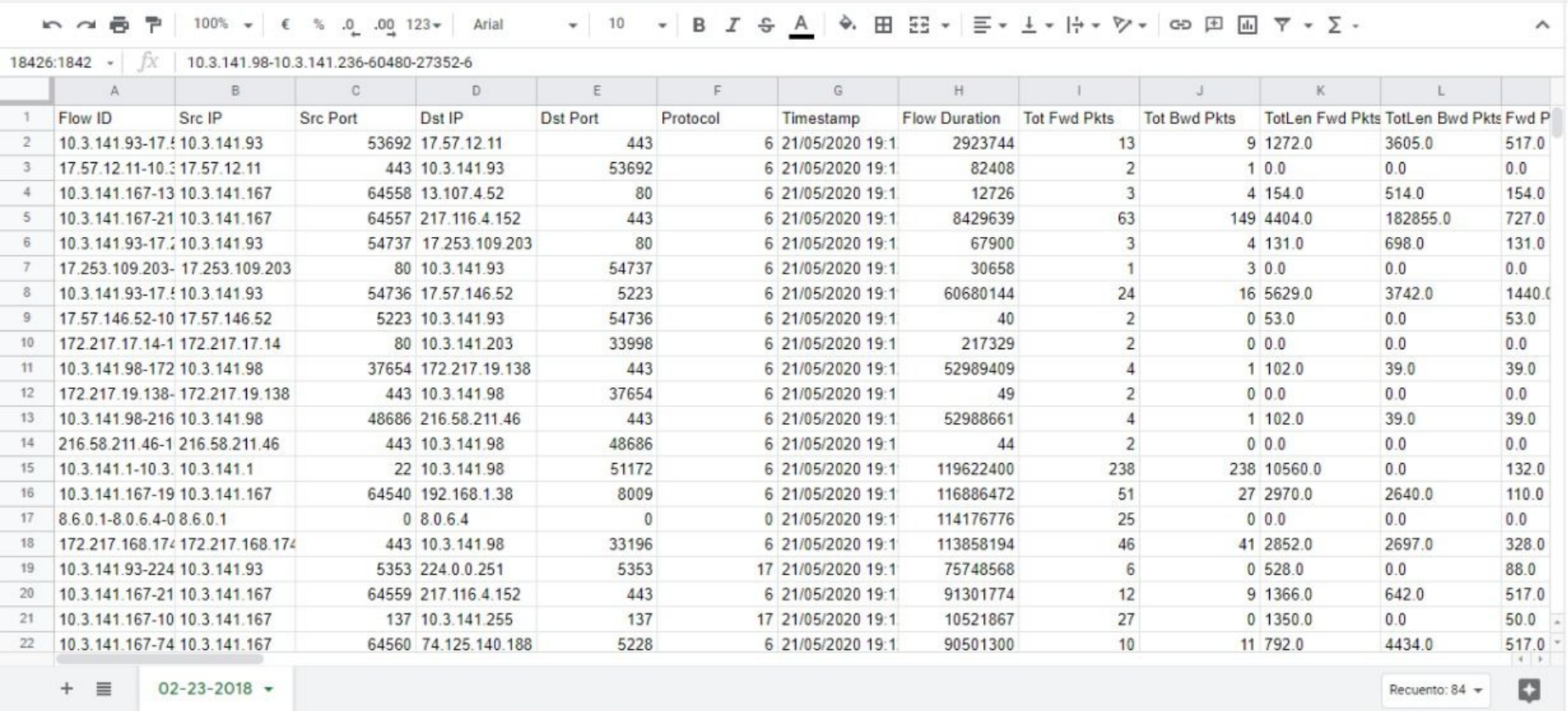
Es una herramienta capaz de detectar intentos de acceso no autorizados a una red. Busca detectar anomalías que indiquen un riesgo potencial, tales como ataques de denegación de servicio, escaneadores de puertos o intentos de entrar en un ordenador, analizando el tráfico en la red en tiempo real.

Base de Datos CSE-CIC-IDS 2018

La base de datos la elaboró la Universidad de New Brunswick a partir del año 2017 y se siguieron generando nuevas versiones posteriormente, (*IDS 2018 Intrusion CSVs (CSE-CIC-IDS2018)*, 2018) esta base de datos se creó en un entorno emulado durante un período de 5 días, contiene tráfico de red en formato de paquetes y los resultados del análisis del tráfico en flujo bidireccional, obtenidos con la herramienta CICFlowMeter y descritos por 80 característica (Versión de la base de datos del 02-23-2018).


Search
Data
Tasks
Code (3)
Discussion
Activity
Metadata
Download (6 GB)
New Notebook
File Structure
Data Explorer
6.41 GB
02-14-2018.csv
02-15-2018.csv
02-16-2018.csv
02-20-2018.csv
02-21-2018.csv
02-22-2018.csv
02-23-2018.csv
02-28-2018.csv
03-01-2018.csv
03-02-2018.csv
02-14-2018.csv (341.63 MB)
Detail
Compact
Column
10 of 80 columns
About this file
This data was compiled on February Fourteenth, 2018. This particular dataset has a majority of normal traffic (Benign) in comparison to actual attacks.
Dst Port
Destination port of connection.
0
65.5k
Protocol
Protocol used during connection.
0
17
Timestamp
Time that connection occured.
32043
unique values
Flow Duration
Duration that connection occurred.
-919011000000
120m
Tot Fwd Pkts
Total number of forward packets.
1
5115
0


Nuevo
Mi unidad
Ordenadores
Compartido conmigo
Reciente
Destacados
Papelera
Almacenamiento
8,17 GB de 15 GB usado
Comprar espacio
Mi unidad > CSE-CIC-IDS2018
02-14-2018.csv
02-23-2018
02-23-2018.csv
02-28-2018.csv
03-01-2018.csv
03-02-2018.csv
02-23-2018.csv
Detalles
Actividad
Quién tiene acceso
Propiedades del sistema
Tipo
Valores separados por comas
Tamaño
9,6 MB
Almacenamiento usado
9,6 MB
Ubicación
CSE-CIC-IDS2018
Propietario
yo
Modificado
29 ago 2021 por mí
Abierto
29 ago 2021 por mí



18426:1842 | 10.3.141.98-10.3.141.236-60480-27352-6

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Flow ID	Src IP	Src Port	Dst IP	Dst Port	Protocol	Timestamp	Flow Duration	Tot Fwd Pkts	Tot Bwd Pkts	TotLen Fwd Pkts	TotLen Bwd Pkts	Fwd P
2	10.3.141.93-17.5	10.3.141.93	53692	17.57.12.11	443	6	21/05/2020 19:1	2923744	13	9	1272.0	3605.0	517.0
3	17.57.12.11-10.3	17.57.12.11	443	10.3.141.93	53692	6	21/05/2020 19:1	82408	2	1	0.0	0.0	0.0
4	10.3.141.167-13	10.3.141.167	64558	13.107.4.52	80	6	21/05/2020 19:1	12726	3	4	154.0	514.0	154.0
5	10.3.141.167-21	10.3.141.167	64557	217.116.4.152	443	6	21/05/2020 19:1	8429639	63	149	4404.0	182855.0	727.0
6	10.3.141.93-17.2	10.3.141.93	54737	17.253.109.203	80	6	21/05/2020 19:1	67900	3	4	131.0	698.0	131.0
7	17.253.109.203-	17.253.109.203	80	10.3.141.93	54737	6	21/05/2020 19:1	30658	1	3	0.0	0.0	0.0
8	10.3.141.93-17.5	10.3.141.93	54736	17.57.146.52	5223	6	21/05/2020 19:1	60680144	24	16	5629.0	3742.0	1440.0
9	17.57.146.52-10	17.57.146.52	5223	10.3.141.93	54736	6	21/05/2020 19:1	40	2	0	53.0	0.0	53.0
10	172.217.17.14-1	172.217.17.14	80	10.3.141.203	33998	6	21/05/2020 19:1	217329	2	0	0.0	0.0	0.0
11	10.3.141.98-172	10.3.141.98	37654	172.217.19.138	443	6	21/05/2020 19:1	52989409	4	1	102.0	39.0	39.0
12	172.217.19.138-	172.217.19.138	443	10.3.141.98	37654	6	21/05/2020 19:1	49	2	0	0.0	0.0	0.0
13	10.3.141.98-216	10.3.141.98	48686	216.58.211.46	443	6	21/05/2020 19:1	52988661	4	1	102.0	39.0	39.0
14	216.58.211.46-1	216.58.211.46	443	10.3.141.98	48686	6	21/05/2020 19:1	44	2	0	0.0	0.0	0.0
15	10.3.141.1-10.3.	10.3.141.1	22	10.3.141.98	51172	6	21/05/2020 19:1	119622400	238	238	10560.0	0.0	132.0
16	10.3.141.167-19	10.3.141.167	64540	192.168.1.38	8009	6	21/05/2020 19:1	116886472	51	27	2970.0	2640.0	110.0
17	8.6.0.1-8.0.6.4-0	8.6.0.1	0	8.0.6.4	0	0	21/05/2020 19:1	114176776	25	0	0.0	0.0	0.0
18	172.217.168.174	172.217.168.174	443	10.3.141.98	33196	6	21/05/2020 19:1	113858194	46	41	2852.0	2697.0	328.0
19	10.3.141.93-224	10.3.141.93	5353	224.0.0.251	5353	17	21/05/2020 19:1	75748568	6	0	528.0	0.0	88.0
20	10.3.141.167-21	10.3.141.167	64559	217.116.4.152	443	6	21/05/2020 19:1	91301774	12	9	1366.0	642.0	517.0
21	10.3.141.167-10	10.3.141.167	137	10.3.141.255	137	17	21/05/2020 19:1	10521867	27	0	1350.0	0.0	50.0
22	10.3.141.167-74	10.3.141.167	64560	74.125.140.188	5228	6	21/05/2020 19:1	90501300	10	11	792.0	4434.0	517.0

02-23-2018

Recuento: 84

Tipos de ataques de la base de datos CSE-CIC-IDS 2018

Ataques	Instrumentos	Duración	Atacante	Víctima
Bruteforce attack	FTP - Patator SSH - Patator	1 día	Kali Linux	Ubuntu 16.4 (Web Server)
DoS attack	Hulk, GoldenEye, Stowloris, Slowhttptest	1 día	Kali Linux	Ubuntu 16.4 (Apache)
DoS attack	Heartleech	1 día	Kali Linux	Ubuntu 16.4 (Open SSL)
Web attack	*Damn Vulnerable Web App (DVWA) *In - house selenium framework (XSS and Brute - force)	2 días	Kali Linux	Ubuntu 16.4 (Web Server)
Infiltration attack	*First level:Dropbox download in a windows machine *Second Level: Nmap and postscan	2 días	Kali Linux	Windows and Macintosh
Botnet attack	* Ares (developed by Python):remote shell, file upload/download, capturing *Screenshots and key logging	1 día	Kali Linux	Windows V.7, 8.1, 10 (32-bit) and 10 (64-bit)
DDoS + PortScan	Low Orbit Ion Canon (LOIC) for UDP, TCP, or HTTP requests.	2 días	Kali Linux	Windows V.7, 8.1, 10 (32-bit) and 10 (64-bit)

Regresión Logística

Es una técnica analítica que nos permite relacionar funcionalmente una variable dicotómica, es decir, una variable que puede adoptar un número limitado de categorías con un conjunto de variables independientes o también llamadas predictoras.

Fue necesario seleccionar hiperparámetros para mejorar la precisión del modelo, se optó por determinarlos mediante la técnica de malla o grid search para obtener la mejor combinación y se dividieron los tres conjuntos de datos en un 70% para el entrenamiento y un 30% para realizar las pruebas.

Resultados

Se dividió en tres muestras definidas como Sample 2 , Sample 3 y Sample 4, las cuales estuvieron balanceadas con la misma cantidad de ataques maliciosos y benignos. Se realizaron tres modelos de Regresión Logística aplicados a cada muestra, en la siguiente tabla se muestran dichos modelos entrenados.

Nombre	Sample	% de original	GridSearch	Tiempo
LogReg1	Sample2	8%	{1,10}	1 h
LogReg2	Sample3	4%	{0.01,0.1,1,10,100,1000,10000}	30 min
LogReg3	Sample4	1%	{0.01,0.1,1,10,100,1000,10000}	15 min

Se consideraron pocos hiperparámetros debido al coste de tiempo de ejecución, sin embargo, estos mostraron buenos resultados.

Matriz de confusión

A continuación, se muestran los resultados del segundo modelo (Sample 3) LogReg2, tras realizar la prueba con su respectivo conjuntos de test.

La matriz de confusión del modelo muestra buenos resultados de verdaderos positivos y verdaderos negativos y en cuanto a los falsos positivos y falsos negativos son pocos los casos en los que tuvo fallas, el cual representa solo al 7% del conjunto de prueba.

	Positivos (%)	**Negativos (%)**
Verdaderos	48,05	45,41
Falsos	4,58	1,96

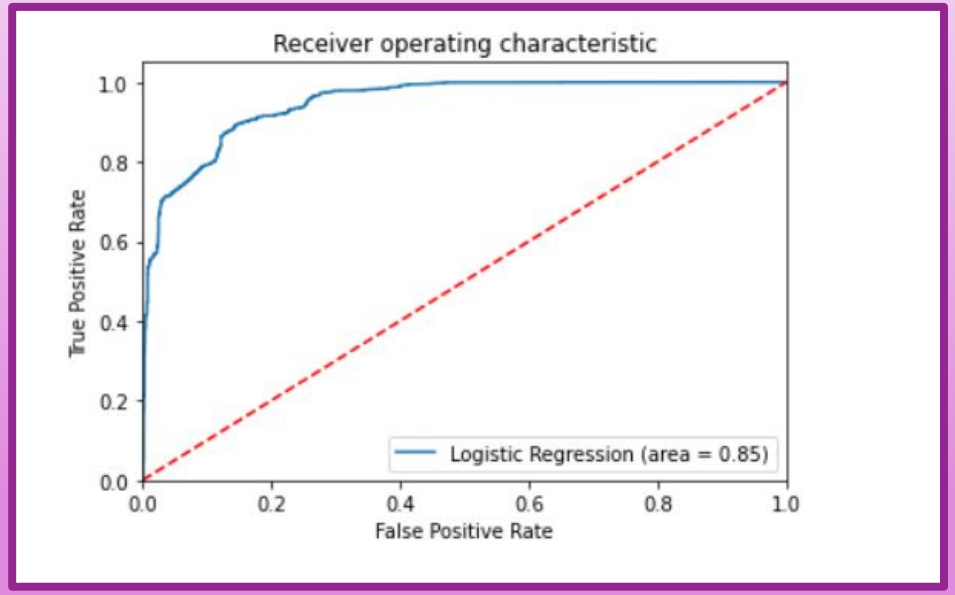
Precisión del modelo

La precisión del modelo, esta es del 96% y un recall del 91% en ataques benignos, y una precisión del 91% y recall del 96% en ataques maliciosos, este resultado muestra que el modelo es bastante bueno al realizar las predicciones.

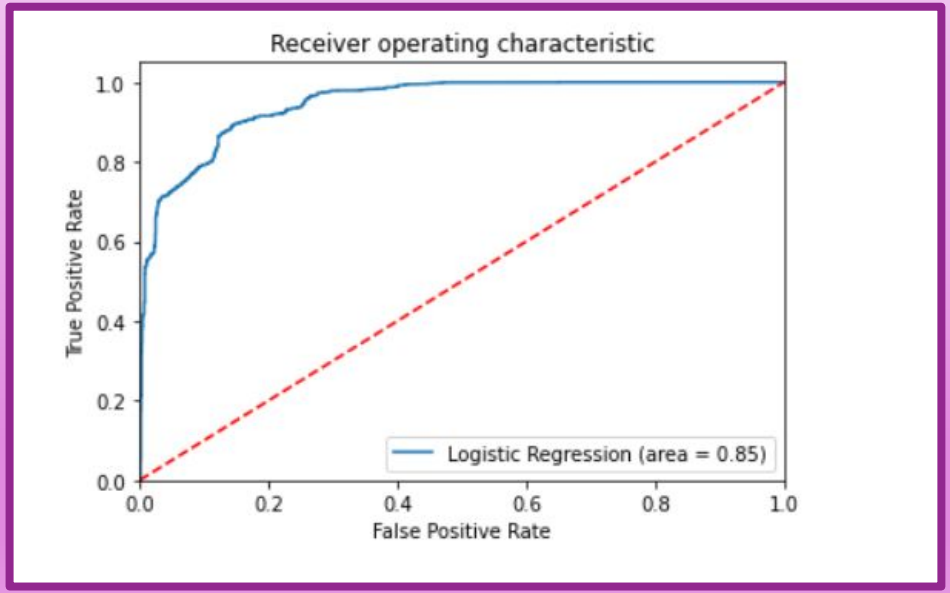
	Precisión	**Recall**
Benignos	0.96	0.91
Ataques	0.91	0.96

Curva ROC

Se utilizó la curva ROC para representar la proporción entre los verdaderos positivos frente a los falsos positivos. El área bajo la curva ROC es 0.93, este resultado confirma que el modelo se comporta adecuadamente sobre el conjunto de prueba.



Para evitar la posibilidad de que se estuviera produciendo un sobreaprendizaje del modelo, en cuanto a los datos tomados del conjunto de datos del CSE-CIC-IDS2018, se realizaron otras pruebas del modelo sobre otras muestras aleatorias del conjunto de datos original. En estas pruebas, los resultados fueron bastante buenos, similares a los del ROC de la Figura 1.



Conclusiones

Se propuso un método para la detección de intrusiones en redes informáticas mediante la aplicación de una técnica de aprendizaje automático, conocida como Regresión Logística, el método consistió en un preprocesamiento de datos, una definición de datos de entrenamiento y prueba, y el desarrollo, entrenamiento y prueba del modelo.

La precisión de detección del modelo en cuanto a ataques benignos es de 96% y de ataques maliciosos es de 91%, estos resultados de consideran bastante buenos, sin embargo, aún hay desafíos por superar, estos desafíos se centran en: gran tamaño de datos, mayor dimensionalidad y preprocesamiento de datos.

¡Gracias!