**Texto

Descripción generada automáticamente**

Estudiantes

Luis Pablo Chavarria Morales

Curso

Inteligencia en Redes de Datos [202403\_001\_BIT-31\_1]

Profesor

Carlos Andres Mendez Rodriguez

Proyecto

Detección de anomalías en redes con IA

Tabla de contenido

[Tema de Investigación 4](#_Toc185426882)

[Marco teórico 4](#_Toc185426883)

[Objetivo General 5](#_Toc185426887)

[Objetivos Específicos 5](#_Toc185426889)

[Plan de Trabajo Detallado 6](#_Toc185426894)

[Metodología de Investigación 7](#_Toc185426895)

[Recursos para la Implementación 7](#_Toc185426896)

[Implementación del proyecto 7](#_Toc185426901)

[Preprocesamiento 8](#_Toc185426904)

[Análisis de Datos: 8](#_Toc185426908)

[Exploración de la distribución de características relevantes, como 'Total Fwd Packets', para tráfico normal y de ataque. 8](#_Toc185426909)

[Filtrado de datos basado en etiquetas para separar tráfico normal (BENIGN) y ataques específicos (como 'PortScan'). 8](#_Toc185426910)

[Bibliography 10](#_Toc185426920)

### Tema de Investigación

El proyecto se enfocará en la detección de anomalías en redes con IA, con el objetivo es detección automática de patrones inusuales en el tráfico de red, con el fin de mejorar la seguridad en redes de telecomunicaciones, con el fin de identificar comportamientos anómalos que indiquen posibles amenazas, como ataques cibernéticos o accesos no autorizados.

### Marco teórico

### La detección de anomalías en el presente aprovecha la inteligencia artificial (IA) para identificar cambios inesperados en el comportamiento de un conjunto de datos. Estos cambios pueden indicar incidentes como amenazas a la seguridad.

### La detección de anomalías importante en sectores como las finanzas o la ciberseguridad. Un ejemplo seria la banca la cual es un sector que se beneficia de la detección de anomalías, ya que por este medio se pueden identificar actividades fraudulentas y patrones incoherentes, y proteger los datos. Sin la detección de anomalías, podría llevar a la pérdida de ingresos y valor de marca. Toda empresa se enfrenta a violaciones de seguridad y a la pérdida de información confidencial de los clientes. Si esto ocurre, se arriesga a perder un nivel de confianza de los clientes que puede ser irrecuperable.

### Implementar firewalls y usar soluciones de seguridad simples ya no es suficiente para proteger las redes de una anomalía o ataque de red, debido al aumento de malwares o ataques de DDoS. Por este motivo y los otros mencionados, una solución de IA puede profundizar en datos en tiempo real para identificar anomalías

### Objetivo General

### Implementar un sistema basado en inteligencia artificial para la detección de anomalías en redes de telecomunicaciones, mejorando la capacidad de identificación de posibles amenazas.

### Objetivos Específicos

### Investigar diferentes algoritmos de IA que son efectivos para la detección de anomalías en redes.

### Implementar un prototipo de sistema de detección de anomalías utilizando algoritmos de IA seleccionados, evaluando su rendimiento en un entorno simulado.

### Evaluar la efectividad del sistema de IA en relación con métodos tradicionales de detección, midiendo métricas como la precisión, la sensibilidad y la tasa de falsos positivos.

### Investigar mejoras en la metodología utilizada y en la configuración de los algoritmos para aumentar la eficacia del sistema.

### Plan de Trabajo Detallado

Para la simulación del trabajo se manejar dos opciones;

1. Selección del Dataset o Entorno de Simulación

Se utilizará un dataset de tráfico de red que contenga tanto tráfico normal como tráfico anómalo o malicioso. Basándonos en bibliografías consultadas algunas se pueden utilizar las siguientes para la simulación:

* CICIDS 2017: Contiene tráfico de red normal y varias categorías de ataques (DoS, phishing, etc.). Este es uno de los más utilizados para la evaluación de sistemas de detección.
* UNSW-NB15: Proporciona un conjunto de datos diverso con características detalladas del tráfico de red, incluyendo tráfico normal y anómalo.

1. Lenguaje de programación

* Python se utilizará por sus bibliotecas de Machine Learning e IA.

1. Google Colab

* Entorno basado en la nube: Evitas problemas relacionados con los recursos limitados de hardware local.
* GPU gratuita: Útil para entrenamiento más rápido de modelos avanzados.
* Integración directa con Google Drive: Facilita el acceso y la carga de datos (como el dataset CICIDS 2017).

### Metodología de Investigación

Buscar artículos académicos, libros y publicaciones recientes sobre técnicas de IA (aprendizaje supervisado, no supervisado, redes neuronales, etc.) aplicadas a la seguridad en redes.

Herramientas: Google Scholar, IEEE Xplore, SpringerLink, bases de datos académicas.

Configuración del entorno de pruebas (entorno simulado o dataset).

Implementación del algoritmo.

Documentar el proceso de investigación.

### Recursos para la Implementación

### Python

### Google Colab

### CICIDS 2017

### Google Drive

### Implementación del proyecto

### El dataset se encuentra almacenado en un archivo comprimido (MachineLearningCSV.zip) dentro de Google Drive.

### Se utilizó Google Colab para montar Google Drive y extraer los archivos necesarios.

### Preprocesamiento:

### Extracción de los archivos CSV del archivo comprimido.

### Limpieza y normalización de los nombres de las columnas para facilitar el análisis.

### Verificación de las etiquetas (Label) y las características relevantes para la detección de anomalías.

### Análisis de Datos:

### Exploración de la distribución de características relevantes, como 'Total Fwd Packets', para tráfico normal y de ataque.

### Filtrado de datos basado en etiquetas para separar tráfico normal (BENIGN) y ataques específicos (como 'PortScan').

### Errores

### ---------------------------------------------------------------------------

### MessageError Traceback (most recent call last)

### <ipython-input-1-d5df0069828e> in <cell line: 2>()

### 1 from google.colab import drive

### ----> 2 drive.mount('/content/drive')

### MessageError: Error: credential propagation was unsuccessful

### Este error sucede cuando Google Colab tiene problemas para acceder a tu Google Drive.

### Análisis de Resultados

A screenshot of a graph

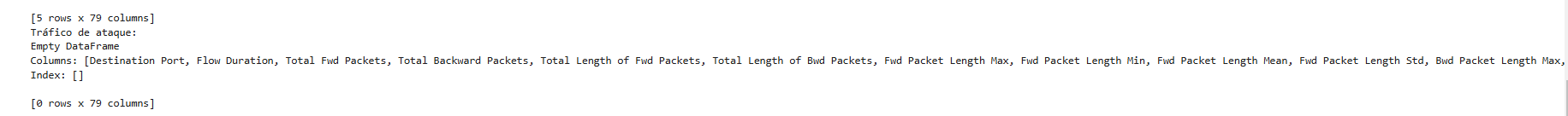
Description automatically generatedAcorde a los resultados s obtenidos muestran un desempeño sobresaliente del modelo, con una precisión (accuracy) de 99.99% en la detección de anomalías, lo que indica que el modelo es capaz de identificar correctamente tanto las conexiones normales como las potencialmente maliciosas. Durante el proceso, se identificaron valores infinitos en las características, lo que fue solucionado reemplazándolos por valores finitos mediante la técnica “replace”.

Estas métricas representan el promedio de las métricas individuales de BENIGN y PortScan, sin ponderar por el número de instancias en cada clase.

A close-up of a text

Description automatically generated

El dataset fue cargado exitosamente desde Google Drive.



Las etiquetas (Label) fueron validadas, confirmando la presencia de tráfico normal y ataques específicos.

# Bibliography

AWS. (n.d.). *Amazon web Services*. Retrieved from https://aws.amazon.com/es/what-is/anomaly-detection/

Barnard, J. (2023, Diciembre 12). Retrieved from IBM: https://www.ibm.com/es-es/topics/anomaly-detection

Iman Sharafaldin, A. H. (2018, Enero). Retrieved from https://www.unb.ca/cic/datasets/ids-2017.html

*kio.tech*. (2020). Retrieved from https://www.kio.tech/blog/transformaciondigital/inteligencia-artificial-deteccion-de-anomalia

Navalur, S. (2022, 11 12). Retrieved from Yotube: https://www.youtube.com/watch?v=pKWclJMeWMQ&t=36s