This work is under a Creative Commons
Attribution – NonComercial – Share Alike 4.0 International License

Ciberdefensa: Modelo de herramienta de asesoramiento para la ciberatribución empleando redes neuronales, validada a través de casos reales
Claudio Lopez and Alejando Molina
SADIO Electronic Journal of Informatics and Operations Research (EJS) Vol. 24 No. 1 (2025) e-ISSN 1514-6774
https://doi.org/10.24215/15146774e070 | https://revistas.unlp.edu.ar/ejs
Sociedad Argentina de Informática e Investigación Operativa | Universidad Nacional de La Plata | Buenos Aires | Argentina

Ciberdefensa: Modelo de herramienta de asesoramiento para la ciberatribución empleando redes neuronales, validada a través de casos reales

Ciberdefense: Cyber attribution advice tool model using neural networks , validated through real cases

Mg Lic Claudio Lopez

Universidad de la Defensa Nacional, Argentina

Dr Ing Alejando Molina

Universidad Tecnologica Nacional, Argentina

Resumen.

La Ciberatribución es una parte fundamental de la ciberdefensa de un Estado. La tarea de asignar un responsable de una ciberagresion (y sobre todo si este lo constituye otro Estado) es realmente complicada teniendo en cuenta el avance tecnológico de herramientas afines a los objetivos de los ciberatacantes. Esta actividad (o sea la Ciberatribucion) es fundamental para crear una verdadera disuasión que desaliente la realización de los mencionados ataques.

Por otro lado, la ciencia informática ha desarrollado de manera vertiginosa, la teoría y el empleo de cientos de herramientas de inteligencia artificial que se ven en nuestra vida diaria, y en las que se usan redes neuronales. Las redes neuronales se nutren de miles de datos y sus resultados son más que aceptables para la optimización, clasificación o predicción.

Las fuentes de ingreso de esos datos pueden ser totalmente variadas y de acuerdo con su cantidad se puede afirmar que se logrará mayor o menor precisión en la salida.

Se propone demostrar que las redes neuronales en el contexto de los procedimientos de la Ciberatribucion pueden ser empleadas con éxito como una herramienta de asesoramiento en la determinación del origen de un ataque cibernético a una infraestructura que posea una función crítica.

Si bien existen estudios referidos a este tema, la particularidad de este trabajo es la de inscribirse en el ámbito de ciberdefensa propio en un todo de acuerdo con la legislación nacional vigente
.

Palabras clave: Ciberdefensa, Ciberdisuacion, aprendizaje automático, redes neurales

Received September 2024; Accepted December 2024; Published March 2025

Abstract.

Cyberatribution is a fundamental part of the cyberdefense of a State. The task of assigning a person responsible for a cyber aggression (and especially if this constitutes another State) is really complicated taking into account the technological advance of tools related to the objectives of cyber attackers. This activity (that is, cyberatribution) is essential to create a true deterrence that discourages the realized attacks.

On the other hand, computer science has developed vertiginously, the theory and use of hundreds of artificial intelligence tools that are seen in our daily lives, and in which neural networks are used. Neuronal networks are nourished by thousands of data and their results are more than acceptable for optimization, classification or prediction.

The sources of entry of this data can be totally varied and according to their amount it can be affirmed that greater or lesser precision will be achieved at the output.

It is proposed to demonstrate that neuronal networks in the context of cyberatribution procedures can be successfully used as an advice tool in determining the origin of a cyber attack to an infrastructure that possesses a critical function.

While there are studies referring to this issue, the particularity of this work is to register in the field of own cyberdefense in a whole in accordance with current national legislation
.
Keywords: cyberdefense, cyberdisuation, automatic learning, neural networks

1. Introducción: El problema de la ciberatribucion

El Gen Larry Welch expresa que "... el ciberespacio es un dominio en, desde y a través del cual las operaciones militares producen los efectos deseados. Los objetivos militares fundamentales relativos a este dominio son esencialmente los mismos que en los otros dominios. El objetivo principal es la libertad de acción en, a través y desde el ciberespacio según sea necesario, para apoyar los objetivos de la misión." (Welch, 2011, p. 2) [5].

Para mantener la libertad de acción será necesario, entre otras cuestiones, disuadir al posible adversario de ejecutar agresiones dentro del ciberespacio. Por lo tanto, el objetivo de esa "Ciberdisuacion" será," lograr que los estados naciones agresores, reales o potenciales, perciban claramente que los costos esperados (económicos, políticos, militares, geopolíticos, de imagen, etc.) asociados a una ciberagresión, superan ampliamente a los resultados esperados de la misma" (Uzal, 2016, p. 8) [3].

Ante un Ciberataque, se hace necesario realizar una correcta determinación del origen de este a fin de determinar si el mismo se encuadra dentro de los parámetros del artículo 51 de la Carta de las Naciones Unidas que trata sobre el derecho inmanente de legítima defensa, individual o colectiva en caso de ataque armado contra un Miembro de la ONU [6].

Dado que:

a. Las armas cibernéticas a menudo se despliegan bajo un manto de anonimato, lo que dificulta averiguar quién es realmente responsable
b. Pueden ser desplegadas de manera remota empleando lugares privados o públicos y desde cualquier sitio en el mundo
c. Uno de los escenarios potenciales más desfavorables que se le puede presentar a un estado nación es recibir ciberagresiones y, por incapacidad tecnológica y/o de gestión, terminar adjudicando los desastres ocasionados por dichos ataques a accidentes imprevistos. (Uzal, 2016, p. 9) [3].
d. Cuanto más demoremos en determinar quién es el agresor, el mismo podrá eludir la acción de respuesta (Uzal, 2015, pp. 2-9) [4].

Entonces, la tarea de establecer el responsable último de estas acciones o atribuir o llevar a cabo la "Ciberatribucion" conlleva la dificultad implícita de que ésta deberá tener una alta probabilidad de éxito, a fin de que la respuesta a una ciberagresión pueda ser interpretada como legítima defensa y quedar incluida en los términos del Artículo 51(Uzal, 2015, pp. 2-9) [4].

En el plano de la legislacion argentina, recordemos que nuestra Ley de Defensa Nacional y sus Decretos modificatorios (Ley 23.554, Decreto 727/2006 y Decreto 571/2020) expresan en su Artículo 1º "Las Fuerzas Armadas, instrumento militar de la defensa nacional, serán empleadas ante agresiones de origen externo perpetradas por fuerzas armadas pertenecientes a otro/s Estado/s…", por lo tanto, se hace necesario que la ciberatribucion obtenga resultados que tengan en cuenta este aspecto [8].

El Decreto 2645/2014, Directiva de Política de Defensa Nacional, de fecha 30/12/2014, en uno de sus enunciados dice: "…Dentro de la amplia gama de operaciones cibernéticas, sólo una porción de éstas afecta específicamente el ámbito de la Defensa Nacional. En efecto, en materia de ciberdefensa existen dificultades fácticas manifiestas para determinar a priori y ab initio si la afectación se trata de una agresión militar estatal externa. Por tal motivo, resulta necesario establecer dicha calificación a posteriori actuando como respuesta inmediata el Sistema de Defensa únicamente en aquellos casos que se persiguieron objetivos bajo protección de dicho sistema, es decir que poseen la intención de alterar e impedir el funcionamiento de sus capacidades" [7].

Se puede extraer como conclusión entonces que:

- El sistema de Defensa Nacional solo reaccionará ante una agresión (ciberagresion) militar estatal externa y actuará únicamente en aquellos casos en que se afecten infraestructuras bajo su protección.
- En segundo término, la determinación del origen del ataque se hará con posterioridad, es decir, después de un análisis de ciberatribución que puede llevar tanto tiempo que, mientras se realiza, se podrán seguir sufriendo más ataques.

2. Objetivo

El objetivo de esta propuesta será definir un algoritmo que: emplee redes neuronales, sea validado por casos reales, y que posibilite ser empleado como una herramienta de asesoramiento en un proceso de ciberatribucion a fin de contribuir a tomar correctas decisiones por parte del órgano de la ciberdefensa nacional.

3. Desarrollo de la propuesta

3.1. Antecedentes

Entre los antecedentes más destacados a esta propuesta cabe mencionar los desarrollos realizados por la compañía Georgia Tech a través del proyecto Rhamnousia. Este proyecto está conectando diversos conjuntos de datos para impulsar nuevas técnicas algorítmicas de atribución que aceleran la obtención de resultados. Si bien las herramientas y técnicas que se desarrollan no apuntan directamente a los individuos responsables, la iniciativa proporciona pruebas de la participación de grupos específicos, identificables por sus métodos de ataque, errores consistentes y otras características únicas. La investigación es patrocinada por el Departamento de Defensa de EEUU, y esta dirigida por investigadores del Instituto de Tecnología de Georgia (https://www.gatech.edu/), en colaboración con otras instituciones académicas y empresas. Michael Farrell, científico jefe del Laboratorio de Seguridad de la Información y Tecnología Cibernética del Instituto de Investigación Tecnológica de Georgia (GTRI), está familiarizado con los problemas que enfrenta el gobierno de los EE. UU, debido a la incapacidad de identificar a quienes atacan los intereses de los EE.UU. en el ciberespacio. "La disuasión es prácticamente imposible si no se puede identificar al adversario", señaló. "La atribución es el eje de la disuasión en el ciberespacio, y el gobierno de los Estados Unidos necesita una forma repetible y liberable de avanzar". La investigación utiliza técnicas de ingeniería y ciencia de datos para examinar conjuntos de datos nuevos y existentes para encontrar información relevante.

3.2. Metodología

Para el desarrollo de la propuesta:

- Se empleo un modelo de aprendizaje automático supervisado.
- Se estudió la estructura de datos más conveniente para los valores de entrada. Estos valores debieron ser codificados para el correcto empleo de la red.
- Se tomaron 4 (cuatro) países simulados como valores de etiqueta para cada instancia de vectores de datos tanto de entrenamiento como de validación.
- Se construyó la red neuronal con una cierta cantidad inicial de neuronas y capas, así como cuantificadores específicos que se fueron probado y modificando hasta alcanzar un valor de evaluación aceptable.
- Como métricas cuantitativas se usaron herramientas propias del entorno de programación (ejemplo matriz de confusión).
- Se empleó el entorno de programación abierto Colaboratory de Google,

- Se usó el lenguaje Tensorflow versión 1.14, basado en Python. El mismo constituye una biblioteca de software de código abierto para aprendizaje automático que fue desarrollado por Google a fin de satisfacer las necesidades de sus sistemas.
- Se validaron los resultados de la red a través de casos reales de atribuciones resueltas y asesoramiento de expertos.

3.3. Modelado de la propuesta

Para el conjunto de datos del modelo se eligió una estrategia inicial de muestreo sobre pocas características que podían tener un poder predictivo fuerte. Esto ayudo a confirmar que el modelo funcionara según lo previsto.

Se empleó como herramienta principal el sitio https://www.kaggle.com/. En él se encontraron distintos Datasets vinculados a la ciberseguridad y a la ciberdefensa que permitieron extraer información de referencia. También se usaron datos e información de las siguientes publicaciones:

- NATIONAL CYBERSECURITY AND CYBERDEFENSE POLICY SNAPSHOTS https://observatoriociberseguridad.org/#/home - Reporte de Ciberseguridad para America Latina y el Caribe 2020 [10]
- THE GLOBAL RISKS REPORT 2021 16TH EDITION – Publicación de WORLD ECONOMIC FÓRUM [11].
- Cyber Warfare Conflict Analysis and Case Studies (2017) - Mohan B. Gazula https://ccdcoe.org/library/strategy-and-governance/ [2].
- A Guide to Cyber Attribution"(2018), OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE, US [1]
- MITRE ATT&CK MATRIX FOR ENTERPRISE (2022) [9]

Cada vector de datos representó un determinado caso de ciberataque simulado atribuido a un potencial adversario también simulado.

Las características del modelo son:

- Tipo_ataque
- Objetivo_material
- Efecto_buscado
- Carácter_de_la_agresion
- Confiabilidad_informacion
- Hecho_politico_cercano
- Complejidad_agresion
- Frecuencia_ataque
- Participación_terceros

Como se puede visualizar los aspectos técnicos (como IP de origen, IP de destino, numero de paquetes, etc.) de la ciberagresion no han sido tenidos en cuenta ya que el propósito del modelo será ofrecer un asesoramiento inicial, acerca de las probabilidades atribuidas a cada actor como potencial ejecutor de una agresión.

Fue asignado a cada atributo un rango de valores discretos a fin de facilitar el aprendizaje del modelo.

Todos estos atributos y sus valores han sido analizados y determinados por el autor. Esto no quiere decir que no puedan existir otros. Lo mismo cabe acotar para los atributos.

Estos valores fueron codificarlos a los fines de que el modelo pueda ser ejecutado en el entorno de programación correspondiente. Se emplearon valores numéricos enteros, teniendo como norma establecer un valor 0 cuando el atributo dentro del vector de datos tiene un valor desconocido. La escala de valores se iniciará en 1 (uno) como valor más bajo del rango.

En la Tabla 1 se muestra la asignación de los valores

Tabla 1. Asignación de valores numéricos a los atributos.

Atributo	Valor	Representación en el modelo	Observaciones
Tipo_ataque	Reconocimiento	1	
	Desarrollo de recursos	2	
	Acceso inicial	3	
	Ejecución	4	
	Persistencia	5	
	Descubrimiento de credenciales de acceso	6	
	Movimiento lateral	7	
	Recopilación	8	
	Comando y Control	9	
	Exfiltración	10	
	Impacto	11	
	No hay datos	0	
Objetivo_material	Personas	1	
	Infraestructura critica	2	
	Sistemas	3	
	No hay datos	0	
Efecto_buscado	Robo de información confidencial	1	
	Modificación de archivos	2	
	Negación del servicio	3	
	Destrucción del Sistema Informático	4	
	Sabotaje	5	
	No hay datos	0	

Motivación	Terrorismo	1	
	Espionaje	2	
	Política	3	
	Económica	4	
	No hay datos	0	
Carácter_de_la_agresion	Promovida por un estado	1	
	Ejecutada por un estado haciendo uso de sus FFAA	2	
	Ejecutada por un estado sin hacer uso de sus FFAA	3	
	No hay datos	0	
Confiabilidad_informacion	Muy Confiable	1	
	Confiable	2	
	No hay datos	0	
Hecho_politico_cercano	Si existió un hecho político cercano	1	
	No existió un hecho político cercano	2	
	No hay datos	0	
Complejidad_agresion	El agresor posee muy buena infraestructura y conocimientos técnicos para la ejecución del ataque	1	
	El agresor posee buena infraestructura y conocimientos técnicos para la ejecución del ataque	2	
	El agresor posee escasa infraestructura y conocimientos técnicos para la ejecución del ataque	3	
	No hay datos	0	
Frecuencia_ataque	Muy frecuente	1	
	Frecuente	2	
	Poco frecuente	0	
Participación_terceros	Hubo participación de terceros	1	
	No hubo participación de terceros	2	
	No hay datos	0	

Se tomaron 5 (cinco) etiquetas para clasificar los datos que representan a los potenciales agresores simulados y se identificaron de acuerdo a la Tabla 2.

En total el Dataset en el inicio contenía 1309 registros de los cuales el 60% (786) fueron empleados como datos de entrenamiento del modelo y el resto (523) como datos de validación.

El criterio para la asignación de las etiquetas a cada vector de datos se realizó inicialmente de manera aleatoria a fin de evitar sesgos y suponiendo de que no se hallaron patrones de comportamiento que induzca a que siempre que se presente una determinada característica la etiqueta corresponderá a un Estado particular

Tabla 2. Valores de etiquetas

Etiqueta	Correspondencia
1	ESTADO A
2	ESTADO B
3	ESTADO C
4	ESTADO D
0	INDETERMINADO

3.4. Red neuronal propuesta

En el modelo presentado se usó la clasificación de clase o etiqueta múltiple. Se empleo el criterio de regresión logística. La misma produce un decimal entre 0 y 1.0 en la salida de la red.

Se utilizó como función de activación, la función “sigmoid”, que en redes neuronales es usada para modelos de regresión logística.

Se implemento como función de salida la función softmax o función exponencial normalizada, que extiende la idea de la regresión logística a un mundo de múltiples clases. softmax asigna probabilidades decimales a cada clase. Se muestra la red de manera gráfica y esquemática, (Figura 1) con la capa de entrada, las capas ocultas y las de salida.

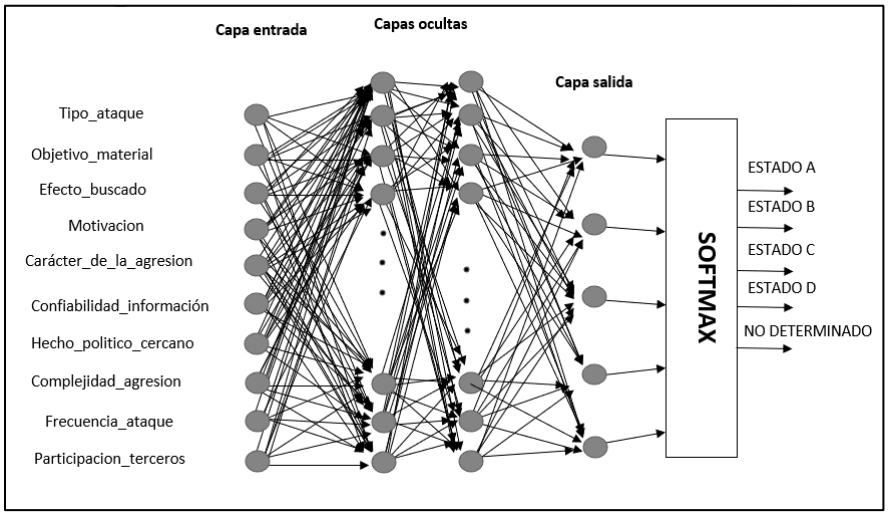

Fig.1. Esquema de red neuronal

```
model = tf.keras.Sequential()
model.add(keras.layers.Dense(10, activation='sigmoid', input_shape=(10, )))
model.add(keras.layers.Dense(20, activation='sigmoid'))
model.add(keras.layers.Dense(5, activation='softmax'))

model.compile(optimizer="adam",loss="categorical_crossentropy",metrics = ['accuracy'])
```

```
history = model.fit(xs_train,ys_train, epochs=300)
```

Fig.2. Construcción con código de la red

Se muestra en la Figura 2 el código en tensorflow de la construcción de la red empleando la clase keras. En cuanto a los argumentos de optimización y perdida fueron empleados “adam” y “categorical_crossentropy” respectivamente, adam es un método de descenso de gradiente estocástico que se basa en la estimación adaptativa de momentos de primer y segundo orden, categorical _crossentropy se utiliza para el modelo de clasificación de clases múltiples donde hay dos o más etiquetas de salida. La etiqueta de salida, si está presente en forma de número entero, se convierte en codificación categórica mediante keras. Y finalmente se tiene accuracy como métrica que se emplea para monitorizar el proceso de aprendizaje (y prueba) de la red neuronal. Una vez definido el modelo y configurado su método de aprendizaje se invoca al método fit

```
model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_3 (Dense)             (None, 10)                110

 dense_4 (Dense)             (None, 20)                220

 dense_5 (Dense)             (None, 5)                 105

=================================================================
Total params: 435
Trainable params: 435
Non-trainable params: 0
_________________________________________________________________
```

Fig.3. Arquitectura de la red

Se observa en la Figura 3 que se requieren 435 parámetros (columna Param #), que corresponden a los 110 parámetros para la primera capa, 220 para la segunda y 105 para la tercera (salida)

3.5. Resultados

Se procedió a ejecutar el modelo con distintas iteraciones o epoch . Los resultados, de manera gráfica, fueron los que muestra la Figura 4**.**

Si se evalúa el modelo con los datos de validación (“test”) como muestra la Figura 5, o sea con datos nunca vistos por este, en el ciclo de iteraciones de la Figura 4 (a) , se ve que el mismo tiene una precisión del 17 %, Es decir, el modelo es incapaz de generalizar.

Este dato indica que el modelo “no ha aprendido” y que se encontró ante una situación de “overfitting” o sobreaprendizaje

Se dedujo entonces que los conjuntos de datos de entrenamiento y validación podrían no estar correctamente construidos y se revisó la estrategia de separación de estos, hallándose una falta de correspondencia entre ambos que se apreció como una de las causales del problema. En otras palabras, los datos fueron separados solo a través de un simple corte en la lista que posibilitaba la relación 60 % - 40 % entre tipos de conjunto de datos lo que podía indicar que ambos no eran de la misma “naturaleza”.

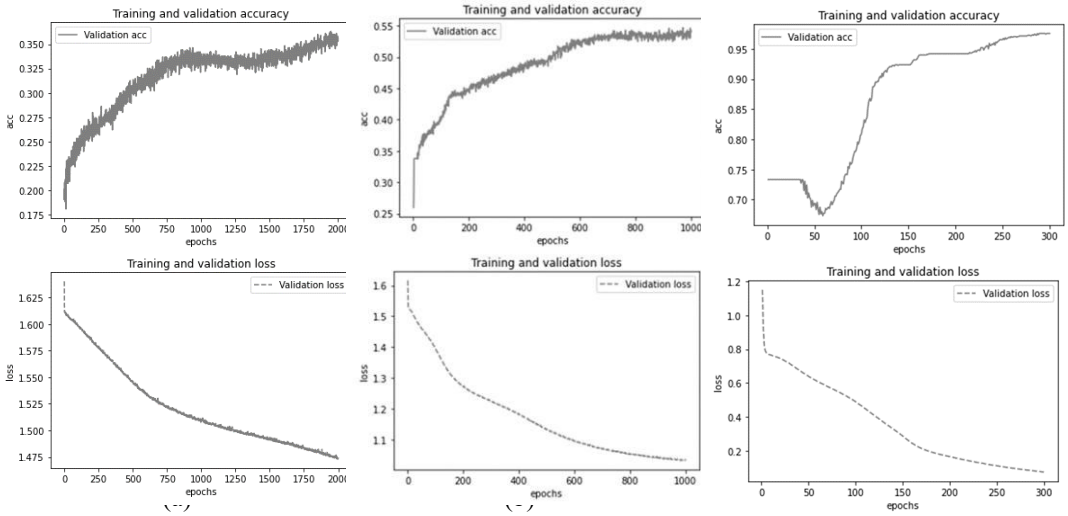

Fig 4: Resultados de los ciclos de iteraciones

```
test_loss, test_acc = model.evaluate(xs_test, ys_test)

12/12 [==============================] - 0s 2ms/step - loss: 1.7106 - accuracy: 0.1750

print('Test accuracy:', test_acc)

Test accuracy: 0.17499999701976776
```

Fig. 5: Perdida y precisión de datos de validación 1º iteración

A continuación, se elaboró otro conjunto de datos de validación teniendo en cuenta este detalle.

Posteriormente se volvió a ejecutar el modelo, pero esta vez en 1000 iteraciones a fin de detectar rápidamente cualquier problema. Se obtuvieron resultados parecidos (Figura 4 (b)).

Se mejoró la precisión (55% con datos de entrenamiento y 25% con datos de validación) pero se empeoró la perdida 3,12 con datos de validación (Figura 6).

Se volvió a revisar los conjuntos de datos y se buscó que haya alguna correspondencia en la distribución de estos y sus etiquetas en vez de incrementar la cantidad de vectores de datos. De hecho, se redujo el Dataset a 948 registros. Asimismo, se cambió la relación de los conjuntos de datos haciéndola de 70% - 30%.

Como se aprecia en la Figura 4 (c) el cambio ha sido muy grande lográndose llevar la precisión arriba del 90 % con solo 300 iteraciones. Por lo tanto, se deduce que este sería el modelo más apto a los fines que se buscan (Figura 7).

```
test_loss, test_acc = model.evaluate(xs_test, ys_test)

12/12 [==============================] - 0s 2ms/step - loss: 3.1287 - accuracy: 0.2556

print('Test accuracy:', test_acc)

Test accuracy: 0.25555557012557983
```

Fig. 6: Perdida y precisión de datos de validación 2º iteración

```
test_loss, test_acc = model.evaluate(xs_test, ys_test)

11/11 [==============================] - 0s 3ms/step - loss: 0.1082 - accuracy: 0.9634

print('Test accuracy:', test_acc)

Test accuracy: 0.9634146094322205
```

Fig. 7: Perdida y precisión de datos de validación 3º iteración

Se muestra, además, el resultado total de la clasificación del modelo por medio del método *model.predict* , y un ejemplo de clasificación en la fila 11 del conjunto de validación donde la predicción se percibe en el valor más alto de los indicados allí , en este caso sería el primero empezando por la izquierda correspondiente al valor 0 o sea país INDETERMINADO.

Se han empleado herramientas para la evaluación del modelo como lo es la matriz de confusión. La misma constituye una tabla con filas y columnas que contabilizan las predicciones en comparación con los valores reales

En la Figura 9 se observa aplicada en el modelo

Otra métrica para empleada con el objeto de medir la precisión entre las predicciones y las etiquetas fue *precisión_score*,

Si se desea probar un vector cualquiera se puede apreciar en la Figura 13 que el resultado predicho por la red corresponde al Estado B o sea el que estaría involucrado en el ataque.

3.6. Discusión

Una de las observaciones que se pueden hacer de la construcción de la red es que el principal problema no fue el código que se elaboró sino el tratamiento que se le dio a los datos.

```
predictions= model.predict(xs_test)
print(predictions)
```

```
[[9.37825501e-01 5.70566058e-02 1.94890879e-03 3.45552166e-04
  2.82340520e-03]
 [9.33681071e-01 6.03133403e-02 2.79712165e-03 1.98428682e-03
  1.22422550e-03]
 [8.71142805e-01 7.80616179e-02 1.20318625e-02 5.89643923e-06
  3.87577079e-02]
 ...
 [4.56051528e-01 1.30183995e-01 3.38017778e-03 1.50364201e-06
  4.10382777e-01]
 [4.52811569e-01 1.31552175e-01 3.46387923e-03 1.51578649e-06
  4.12170827e-01]
 [4.49889928e-01 1.33059859e-01 3.56233446e-03 1.53056590e-06
  4.13486332e-01]]
```

```
print(predictions[11])
```

```
[9.8961020e-01 4.8194937e-03 1.2330770e-03 4.3200920e-03 1.7179977e-05]
```

Fig. 8: Valores de predicción del modelo

```
confusion_matrix(ys_test.argmax(axis=1), predictions.argmax(axis=1))
```

```
array([[207,   0,   0,   0,   0],
       [  8,   0,   0,   0,   0],
       [  0,   0,  94,   0,   0],
       [  0,   0,   0,  15,   0],
       [  4,   0,   0,   0,   0]])
```

Fig. 9: Matriz de confusión del modelo

Los resultados fueron más precisos cuando, en primer lugar, se realizó una distribución más acorde entre los conjuntos de datos de entrenamiento y validación y posteriormente se estableció alguna relación entre las características y etiquetas, o sea cuando el modelo “encontró una conexión” entre ambos.

Sin embargo, se deberá tener cuidado con esto último ya que podrá llevar a un riesgo de sesgamiento de los datos, por lo que corresponderá analizar la estrategia de construcción de los Datasets de entrenamiento y validación y evitar sobreajustes.

Una de las soluciones para atender este problema sería aumentar sustancialmente la cantidad de datos involucrados en el modelo, cuestión que no asegura el éxito.

Asimismo, como se sabe, el entorno legal donde trabajaría el modelo nos obligaría a tener en cuenta sólo aquellos ciberataques atribuidos a un adversario externo, estatal y militar.

4. Validación

4.1. Análisis de casos

La cuestión de saber si el moldeo sirve para ser aplicado en un entorno real, donde un Comando de Ciberdefensa decide emplear este como una de las herramientas de asesoramiento a fin de determinar el origen de un ataque cibernético y dar una respuesta adecuada al mismo se basó en el trabajo realizado por Mohan B. Gazula en “Cyber Warfare Conflict Analysis and Case Studies” (2017), e información extraída de Internet.

Se tomaron cuatro casos como referencia para realizar un análisis de las características de cada uno y compararlas con los valores establecidos en conjunto de datos del modelo.

Los casos analizados fueron:

a. RUSIA - GEORGIA 2008

   Se refiere al conflicto armado protagonizado entre la Republica de Georgia y la Federación Rusa con el apoyo de las autoproclamadas repúblicas prorrusas de Osetia del Sur y Abjasia en agosto 2008. Si bien la guerra empezó oficialmente el 7 de agosto de 2008 en Osetia del Sur, se extendió posteriormente a otras regiones de Georgia y al mar Negro. Semanas antes se "intervino” el sitio web del presidente de Georgia, el cual sufrió un ataque DDoS de piratas informáticos asociados a Rusia.

   A nivel estratégico, los ataques rusos de reconocimiento y exploración del ciberespacio comenzaron mucho antes del inicio real del combate virtual y físico.

   La infraestructura de Internet de Georgia sufrió el 20 de julio de 2008, “bombardeos” coordinados de millones de solicitudes, conocidos como ataques distribuidos de denegación de servicio o DDoS., que sobrecargaron y cerraron efectivamente numerosos servidores.

   Según los expertos, era la primera vez que se producía un ciberataque conocido.

Las modalidades de ataque incluyeron: Desfiguración de sitios web (hacktivismo), operaciones psicológicas basadas en la web (Psyc-Ops), una feroz campaña de propaganda (PC) y, por supuesto, ataques distribuidos de denegación de servicio (DDoS). La aplicación a este conflicto del modelo se muestra en la Figura 10. Por ejemplo, el Tipo de ataque , impacto , corresponde a la codificación 11, la Motivación , política , corresponde al valor 3

b. YELLOWSTONE 1- 2014

A principios de 2014, intrusos informáticos que se cree operaban desde Irán lanzaron un ataque cibernético a los sistemas del casino Las Vegas Sands Corp. que cerró grandes secciones de la compañía a causa de este. La operación fue realizada en respuesta a los comentarios de sus dueños sobre la política exterior iraní, principalmente de su CEO Sheldon Adelson quien el 08 de enero de 2014 participando en un panel de discusión sobre "¿Existirán los judíos?" en la Universidad Yeshiva de Nueva York, dijo que pondría fin a cualquier ambición nuclear de Irán al detonar una bomba atómica en un área desértica despoblada de ese país y agregó: "Y luego dices: '¿Ves? El siguiente sera en el medio de Teherán". Piratas informáticos localizaron la información de inicio de sesión de un ingeniero de sistemas informáticos sénior de Sands con sede en Las Vegas que había pasado brevemente por Pensilvania. Esos datos permitieron a los piratas lanzar una "bomba de malware" el 10 de febrero de 2014 dirigida directamente a los sistemas informáticos en Las Vegas. La bomba de malware afectó alrededor de las tres cuartas partes de los servidores informáticos de la empresa en Las Vegas. Las computadoras no funcionaban, el correo electrónico quedo inoperante, la mayoría de los teléfonos no funcionaron y varios de los sistemas tecnológicos que ayudaron a ejecutar la operación se detuvieron. La aplicación a este conflicto del modelo se muestra en la Figura , por ejemplo, el carácter de la agresión , promovida por un estado , corresponde a la codificación 1, el hecho político cercano , no existió un hecho político cercano , corresponde al valor 2

c. CONFLICTO RUSO-UCRANIANO 2015

En diciembre de 2015 hackers rusos comenzaron con una campaña de phishing dirigida contra el personal de TI y los administradores de sistemas que trabajaban para varias empresas responsables de la distribución de electricidad en Ucrania. La campaña de phishing envió correos electrónicos a los trabajadores de tres de las empresas con un documento de Word malicioso adjunto. Cuando los trabajadores abrían el archivo adjunto, aparecía una ventana emergente que les pedía que habilitaran macros para el documento. Si cumplían, un programa llamado BlackEnergy3, infectaba las máquinas y abría una puerta trasera a los piratas informáticos. Así se recolectaron credenciales de trabajadores, algunas de ellas para las VPN que los trabajadores de la red usaban para iniciar sesión de forma remota en la red SCADA. Una vez que ingresaron a las redes SCADA, prepararon el escenario para su ataque.

Durante muchos meses antes, realizaron un amplio reconocimiento, exploraron y mapearon las redes y obtuvieron acceso a los controladores de dominio de Windows,

donde se administraban las cuentas de usuario de las redes. Luego, escribieron un firmware malicioso para reemplazar el firmware legítimo en los convertidores de serie a Ethernet en más de una docena de subestaciones

Alrededor de las 3:30 p.m. del 23 de diciembre de 2015, ingresaron a 35 redes SCADA a través de las VPN secuestradas. Pero antes de que lo hicieran, lanzaron un ataque

telefónico de denegación de servicio contra los centros de llamadas de los clientes para evitar que estos se comunicaran para informar sobre la interrupción. Después de que el ataque se completó, se usó una pieza de malware llamada KillDisk para borrar los archivos de las estaciones de operador y dejarlos inoperables también. KillDisk borra o sobrescribe datos en archivos esenciales del sistema, lo que hace que las computadoras se bloqueen. Debido a que también sobrescribe el registro de inicio maestro, las computadoras infectadas no pudieron reiniciarse.

La afectación no duró mucho tiempo, solo de una a seis horas para todas las áreas afectadas, sin embargo, los centros de control estuvieron completamente inoperativos por más de dos meses. La aplicación a este conflicto del modelo se muestra en la Figura 12.

d. THE SHAMOON-ATTACK I Y II

El 15 de agosto de 2012 la compañía Saudi Aramco, petrolera nacional de Arabia Saudita, (aunque la compañía no lo anunció oficialmente), se vio obligada a aislar su red informática. La capacidad de Saudi Aramco para suministrar el 10% del petróleo del mundo estuvo repentinamente en riesgo.

Ese 15 de agosto, una persona con acceso privilegiado a las computadoras de la compañía petrolera estatal saudita, activó el virus informático para iniciar lo que se considera uno de los actos de sabotaje informático más destructivos en una empresa. Fue un ataque a 35.000 computadoras de Aramco que inutilizó las computadoras infectadas, lo que hizo que la empresa tardara una semana en restaurar sus servicios.

El código malicioso se transmitió a través de Internet y luego procedió a moverse a través de las computadoras de la red. El virus se copiaba a sí mismo dentro de una tarea del sistema operativo Windows.

El malware empleado se trató de Disttrack un gusano/troyano/filerase capaz de infectar masivamente los equipos ubicados en una determinada red local, y una vez infectados, proceder en primer lugar a “limpiar-borrar”los archivos encontrados para, a continuación, reescribir su (Master Boot Record).

El malware siguió con su actividad hasta 2017, no dejando rastros. Aunque la inteligencia de EE. UU. señaló a Irán como el perpetrador, no hay evidencia específica para apoyar esta hipótesis. Sin embargo, se sabe que ambos países tienen un enfrentamiento ideológico y geopolítico desde 1979 cuando el gobierno islámico de Irán llamó a derrocar a todas las monarquías árabes protagonizando la llamada Guerra Fría de Medio Oriente. La aplicación a este conflicto del modelo se muestra en la Figura 13.

Atributo	Valor	Representación en el modelo	Observaciones
Tipo_ataque	Impacto	11	
Objetivo_material	Sistemas	3	
Efecto_buscado	Negación del servicio	3	
Motivación	Política	3	
Carácter_de_la_agresion	Ejecutada por un estado sin hacer uso de sus FFAA	3	
Confiabilidad_informacion	Muy Confiable	1	
Hecho_politico_cercano	Si existió un hecho político cercano	1	
Complejidad_agresion	El agresor posee muy buena infraestructura y conocimientos técnicos para la ejecución del ataque	1	
Frecuencia_ataque	Frecuente	2	
Participación_terceros	Hubo participación de terceros	1	

Fig. 10: Comparación del caso Rusia - Georgia 2008 con la estructura de datos del modelo

Atributo	Valor	Representación en el modelo	Observaciones
Tipo_ataque	Impacto	11	
Objetivo_material	Sistemas	3	
Efecto_buscado	Negación del servicio	3	
Motivación	Politica	3	
Carácter_de_la_agresion	Promovida por un estado	1	
Confiabilidad_informacion	Confiable	2	
Hecho_politico_cercano	No existió un hecho político cercano	2	
Complejidad_agresion	El agresor posee buena infraestructura y conocimientos técnicos para la ejecución del ataque	2	
Frecuencia_ataque	Poco frecuente	0	
Participación_terceros	Hubo participación de terceros	1	

Fig. 11 : Comparación del caso Yellowstone 1 con la estructura de datos del modelo

Atributo	Valor	Representación en el modelo	Observaciones
Tipo_ataque	Comando y Control	9	
Objetivo_material	Infraestructura critica	2	
Efecto_buscado	Sabotaje	5	
Motivación	Terrorismo	1	
Carácter_de_la_agresion	Promovida por un estado	1	
Confiabilidad_informacion	Muy Confiable	1	
Hecho_politico_cercano	Si existió un hecho político cercano	1	Inicio Conflicto en Ucrania 2014
Complejidad_agresion	El agresor posee muy buena infraestructura y conocimientos técnicos para la ejecución del ataque	1	
Frecuencia_ataque	Muy frecuente	1	
Participación_terceros	Hubo participación de terceros	1	

Fig. 12: Comparación del caso Ruso-Ucraniano 2015 con la estructura de datos del modelo

Atributo	Valor	Representación en el modelo	Observaciones
Tipo_ataque	Movimiento lateral	7	
Objetivo_material	Sistemas	3	
Efecto_buscado	Negación del servicio	3	
Motivación	Politica	3	
Carácter_de_la_agresion	No hay datos	0	
Confiabilidad_informacion	Confiable	2	
Hecho_politico_cercano	No existió un hecho político cercano	2	
Complejidad_agresion	El agresor posee buena infraestructura y conocimientos técnicos para la ejecución del ataque	2	
Frecuencia_ataque	Poco frecuente	0	
Participación_terceros	Hubo participación de terceros	1	

Fig. 13: Comparación del caso The Shamoon-Attack I Y II con la estructura de datos del modelo

4.2. Corolario

Como se puede observar, si se analiza cuidadosamente, se han elegido cuatro casos, dos relacionados a operaciones llevadas a cabo y atribuidas a la Federación Rusa y otras dos atribuidas a la República Islámica de Irán (si bien no está totalmente confirmado).

En los casos de atribución a la Federación Rusa se puede notar diferencias en cada caso respecto al efecto buscado durante el ataque o al objetivo material de este, pero la similitud entre ambos es que los mismos se realizaron prácticamente previo y durante el conflicto como se muestra en las Figuras 10 y 12 (en el caso del conflicto con Ucrania , los antecedentes de confrontación entre ambas naciones vienen desde la crisis de 2014 y siguen permaneciendo) o sea hubo un detonante o hecho político cercano para el inicio del ciberataque.

En el caso iraní el conflicto ha perdurado con el tiempo por cuestiones ideológicas (en el caso Yellowstone 1 el ataque fue puntual a un privado en respuesta a un acontecimiento especifico) como en el caso de Arabia Saudita (además de con Israel o Estados Unidos, enemigos declarados del régimen, desde hace muchos años), por lo que se asignó el valor 2 , (o sea **no** existió un hecho político cercano) al atributo Hecho_politico_cercano de la estructura de datos del modelo, en los dos sucesos presentados (Figuras 11 y 13). Además, no se trata de un actor que posea una infraestructura importante de ciberdefensa en comparación con las potencias.

Esto permite inferir de que es posible encontrar conductas que permitan establecer patrones que favorezcan la precisión de los resultados del modelo. También cabe aclarar que no todas las operaciones de ciberataque en la historia han podido ser atribuidas de forma fehaciente por lo que esta cuestión se ha incluido en el modelo con una etiqueta de país INDETERMINADO.

5. Conclusiones

La complejidad del ciberespacio, la gran cantidad de actores, estatales o privados, el avance de la tecnología y de los exploits, etc., hacen muy difícil suponer que se está totalmente blindado contra las intrusiones maliciosas en los sistemas. Por lo tanto, se debe hacer la suposición de que es muy probable que una infraestructura critica propia esté ya siendo víctima, por diversas motivaciones, de un ciberataque perpetrado por hackers apoyados por algún Estado.

Si bien herramientas como los honeypots o la vigilancia del comportamiento de los mencionados sistemas pueden llegar a ser muy útiles, la ciberdisuacion hoy se convierte en un eje relevante en la ciberdefensa de cualquier Estado -Nación. La “amenaza” de las consecuencias que puede tener cualquier acción contra las infraestructuras criticas propias llega a ser un factor clave. “Mostrar las armas”, o sea dar a conocer las propias capacidades, refuerza este concepto. Esto se fortifica si se está en capacidad de lograr certeza en la ciberatribución, por ello contar con instrumentos eficaces como una inteligencia de amenazas adecuada, tecnologías modernas como la Inteligencia artificial, legislación que tenga en cuenta las características estratégicas de los tiempos actuales y otros, serán imprescindibles para lograr “identificar” a los responsables de la agresión.

Por otro lado, el Aprendizaje Automático, ha cambiado el paradigma de la programación y por lo tanto de la construcción de los sistemas en la actualidad. El avance la ciencia hacia una Inteligencia Artificial General, constituye un logro incalculable, pero a la vez un riesgo todavía

no medido dado el uso perverso que se le pueda llegar a dar. Por el momento se deben aprovechar las ventajas que ofrece y recurrir a ella usándola en los casos que son factibles. La solución propuesta pudo beneficiarse de esta herramienta y obtener resultados que pueden llegar a ser muy útiles como ayuda a la decisión.

El modelo no es complicado de emplear, es altamente escalable, portable y su mantenimiento (actualización) es sencillo.

Poniendo foco en los datos nuevamente, su análisis y ponderación se hacen fundamentales para su incorporación al modelo.

Su ámbito de aplicación sería el Comando Conjunto de la Ciberdefensa, y su función principal la de constituirse en una herramienta más de asesoramiento para la determinación del origen de un ciberataque a infraestructuras críticas nacionales.

La información volcada en él será la que ese Comando Conjunto considere como valida y útil para que las salidas sean aprovechables. Debe provenir de una inteligencia de amenazas robusta que posibilite incrementar su confiabilidad a fin de que, como se expresó más arriba, se pueda llegar a resultados más certeros. Asimismo, se debe preservar la confidencialidad en la gestión de la mencionada información, a través de una política a tal efecto, ya que la misma reviste carácter de clasificada (secreta).

Teniendo en cue3nta el texto de nuestra Ley de Defensa, se hace necesario obtener resultados del modelo que tengan en cuenta este aspecto.

El uso de ejemplos de casos reales contribuyó a sostener la validez del modelo, ya que se logró “encastrar” sin problemas, su estructura de datos y hallar una correcta relación entre los valores de cada atributo con la información obtenida de cada caso. Asimismo, se encontraron patrones de comportamiento que inducen a que probablemente se hubiese obtenido un resultado positivo si se hubiere aplicado el modelo.

6. Futuras líneas de investigación

En la actualidad las capacidades de atribución son mejores que en el pasado reciente, en gran parte porque las naciones están más atentas a la posibilidad de actividad cibernética maliciosa. Por lo tanto, es más probable que recopilen datos que podrían ser útiles en la investigación de una intrusión presente o futura. Las herramientas de atribución son mejores y los analistas tienen más experiencia. Dicho de otra manera, dada la probabilidad de actividad cibernética maliciosa en el futuro, muchas naciones están más dispuestas a realizar inversiones en inteligencia y desarrollar una capacidad de investigación que valdrá la pena [5].

Por consiguiente, podrán existir diferentes maneras de empleo de la IA a través de redes neuronales que permitan construir otros tipos de herramientas como la presentada aquí. El aprendizaje por refuerzo permite a los modelos aprender a tomar decisiones óptimas a través de interacciones con un entorno. En este caso se podría analizar su eficiencia en la ciberatribucion ya que un entorno en donde la facilidad con la que se pueden plantar pistas falsas, realizar operaciones de bandera falsa y en donde es más probable que se lleven a cabo contramedidas

para despistar a los investigadores, especialmente a medida que crece lo que está en juego, hacen interesante su aplicación.
Otra línea para sondear seria la aplicación de un algoritmo de búsqueda complejo como Monte Carlo Search Tree (MCST), el cual trabaja muy bien con información imperfecta.

7. Referencias.

1. "A Guide to Cyber Attribution"(2018), OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE US, https://www.dni.gov/files/CTIIC/documents/ODNI_A_Guide_to_Cyber_Attribution.pdf
2. Carta de las Naciones Unidas
3. Decreto 2645/2014 - DIRECTIVA DE POLÍTICA DE DEFENSA NACIONAL – ACTUALIZACIÓN
4. Gazula Mohan B. , (2017), Cyber Warfare Conflict Analysis and Case Studies, Cybersecurity Interdisciplinary Systems Laboratory (CISL)- Sloan School of Management, Room E62-422 - Massachusetts Institute of Technology.
5. Lin Herbert (2016)," Attribution of Malicious Cyber Incidents", Hoover Institution - Stanford University. https://www.hoover.org/
6. Ley 23.554 de DEFENSA NACIONAL
7. MITRE ATT&CK MATRIX FOR ENTERPRISE (2022). https://attack.mitre.org/matrices/enterprise/
8. NATIONAL CYBERSECURITY AND CYBERDEFENSE POLICY SNAPSHOTS
9. THE GLOBAL RISKS REPORT 2021 16TH EDITION – Publicación de WORLD ECONOMIC FÓRUM
10. Uzal Roberto, (2016) Ciber Disuasión. Un capítulo particularmente sensitivo de la Ciberdefensa, BOLETÍN DEL ISIAE Instituto de Seguridad Internacional y Asuntos Estratégicos Número 64. 8-18
11. Uzal Roberto (2015), El Problema de la Ciber Atribución: Aportes para una estrategia de Ciber Defensa BOLETÍN DEL ISIAE Número 61, 2 – 9
12. Welch Larry D. Usaf (Ret.), s.f. "Cyberspace – the Fifth Operational Domain". Institute for Defense Analysis https://www.ida.org/-/media/feature/publications/2/20/2011-cyberspace---the-fifth - operational-domain/2011-cyberspace---the-fifth-operational-domain.ashx
13. Sedkaoui S., (2018) Statistical and Computational Needs for Big Data Challenges, En Al Mazari (Ed.), "Big Data Analytics in HIV/AIDS Research", (pp. 21-53). Hershey, PA: IGI Global, 2018, doi:10.4018/978-1-5225-3203-3.ch002
14. Shandilya Shishir K. , Wagner Neal, Nagar Atulya K. (2020) Advances in Cyber Security Analytics and Decision Systems, EAI/Springer Innovations in Communication and Computing
15. Segal, A., Grigsby, A., (2018) New Entries in the CFR Cyber Operations Tracker: Q1 2018" Council on Foreign Relations, Abril 23, https://www.cfr.org/blog/new-entries-cfr-cyber-operations-tracker-q1-2018
16. Scott Rose, Borchert Oliver, Stu Mitchell, Connelly Sean, (2020) "Zero Trust Architecture", National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

17. Tsagourias Nicholas , Farrell Michael D, (2020), Cyber attribution: technical and legal approaches and challenges [Archivo PDF] https://sites.tufts.edu/
18. Trobec, R., Slivnik, B., Bulić, P., Robič, B. (2018) Introduction to Parallel Computing: From Algorithms to Programming on State-of-the-Art Platforms. [Archivo PDF] https://e6.ijs.si/~roman/files/Book_jul2018/book/book.pdf
19. Xing Fang , Maochao Xu, Shouhuai Xu , Peng Zhao (2019) A deep learning framework for predicting cyber attacks rates, EURASIP Journal on Information Security. [Archivo PDF]. https://jis-eurasipjournals.springeropen.com/
20. Yannakogeorgos Panayotis A. (2016), Strategies for Resolving the Cyber Attribution Challenge, Air Force Research Institute Air University Press Maxwell Air Force Base, Third Edition. https://media.defense.gov/2017/May/11/2001745613/-1/ 1/0/CPP_0001_YANNAKOGEORGOS_CYBER_TTRIBUTION_CHALLENGE.PDF
21. 12 Tips to 2022 , Seminar Presentations (2020), IPSS INTERNATIONAL PROGRAM ON CYBERSECURITY STUDIES, EUROPEAN CENTER FOR SECURITY STUDIES