Proyecto Final

DOCENTE	CARRERA	CURSO
PhD(c) Vicente Machaca Arceda	Maestría en Ciencias de la Computación	Algoritmos y Estructura de Datos

1. Integrantes

- Grupo N° 2
- Integrantes:
  - EDER ALONSO AMPUERO ATAMARI
  - HOWARD FERNANDO ARANZAMENDI MORALES
  - JOSE EDISON PEREZ MAMANI
  - HENRRY IVAN ARIAS MAMANI

2. Repositorio GitHub

URL Github: Repositorio Proyecto Final

3. Video de Exposición en YouTube

URL YouTube: Video de Exposición en YouTube

4. Marco Teórico

4.1. Detección de Publicidad Engañosa

4.1.1. Descripción del Proyecto

El impacto de la pandemia en la sociedad ha sido palpable en muchos ámbitos. Especialmente se ha notado un crecimiento exponencial del teletrabajo y del uso masivo de medios de comunicación digitales tanto a nivel personal como profesional (videoconferencias, chats, herramientas colaborativas, etc.). Aun así, el correo electrónico sigue ocupando su posición de liderazgo como el medio de comunicación más usado, situándose el tráfico global diario en 2020 por encima de los 300 billones. Por tanto, es, sin duda, el vector de ataque preferido por los ciberdelincuentes, provocando pérdidas billonarias, por ejemplo, en ataques de tipo business email compromise.

La identificación y defensa frente a 'emails' maliciosos es uno de los mayores retos de la ciberseguridad empresarial

La identificación y defensa frente a emails maliciosos continúa siendo uno de los mayores retos de la ciberseguridad empresarial hoy día. No obstante, el informe sobre Ciberamanezas y tendencias 2020 del CCN-CERT arroja una serie de conclusiones importantes que marcarán la pauta en las estrategias de ciberseguridad de las empresas. Entre otras, menciona que los cibercriminales utilizan activamente técnicas de Inteligencia Artificial (IA) para perpetrar sus ataques, lo que incrementa el riesgo exponencialmente y recrudece las amenazas.

Otro documento de ENISA sobre ataques de phishing aporta un dato muy revelador: la ciberdelincuencia basada en la ingeniería social o hacking humano se ha amplificado un 600 por ciento durante la pandemia. Hay claramente, por tanto, un crecimiento significativo y alarmante.

Pero el informe también aporta otro dato curioso, y es que el 30 por ciento de los ataques de phishing se producen en lunes. Esto pone de manifiesto que la ciberdelincuencia es una industria perfectamente organizada y planificada que no deja nada al azar, aplicando incluso técnicas de influencia propias del marketing para maximizar sus resultados.

Las empresas, concluye el dosier, necesitan la implementación de técnicas de IA para defenderse en igualdad de condiciones.

4.1.2. Problema y Solución

En el mundo de la ciberseguridad, la IA tiene una doble vertiente: es a la vez problema y solución, ya que se emplea tanto para el ataque como para la defensa. Gracias a los avances tecnológicos que han propiciado un aumento enorme en la velocidad de procesamiento, soluciones basadas en Machine Learning o Deep Learning son altamente eficaces en sus propósitos.

Toda solución de ciberseguridad basada en IA debe contemplar desde el principio aspectos tan importantes como la hoja de ruta a seguir, qué objetivos queremos alcanzar y en qué ámbito queremos aplicarla. Un ejemplo destacado es su utilización para mejorar las capacidades a nivel defensivo de un Blue Team, de un centro de operaciones de seguridad (SOC) o de un centro de respuesta a incidentes con objeto de anticiparnos y detectar de forma temprana correos electrónicos maliciosos.

En el mundo de la ciberseguridad, la Inteligencia Artificial tiene una doble vertiente: es a la vez problema y solución, ya que se emplea tanto para el ataque como para la defensa

Sin embargo, para entender este ecosistema tecnológico, necesitamos tener claras algunas definiciones básicas: Inteligencia Artificial como la capacidad de una máquina para realizar tareas cognitivas humanas como percibir, aprender, razonar y resolver problemas. Dentro de este ámbito, existen dos técnicas que funcionan con paradigmas y modelos distintos pero complementarios: Machine Learning y Deep Learning. Machine Learning como solución matemática (estadística) que permite detectar patrones y hacer predicciones en base a un conocimiento previo. Y Deep Learning con base en una red neuronal artificial que le permite tener un alto grado de comprensión de los problemas para resolverlos de forma autónoma.

Con estas definiciones en mente, la utilización de IA para capacitar los centros de operaciones de seguridad o similares tiene un caso de uso muy interesante para automatizar el análisis basado en el comportamiento de grandes cantidades de correos electrónicos en el menor tiempo posible, de cara a la predicción y anticipación. Pero no hablamos de un escaneo al estilo tradicional; para eso ya existen herramientas comerciales que se encargan de hacerlo y nos protegen razonablemente bien. Se trata de complementar este tipo de defensa con el análisis basado en el comportamiento para poder defendernos frente a lo desconocido.

5. Programa

5.1. Descripcion del Programa

5.2. Partes del Programa

5.2.1. Interfaz html

En la imagen **??** se evalua los resultados después de crear el archivo *sketch.js*

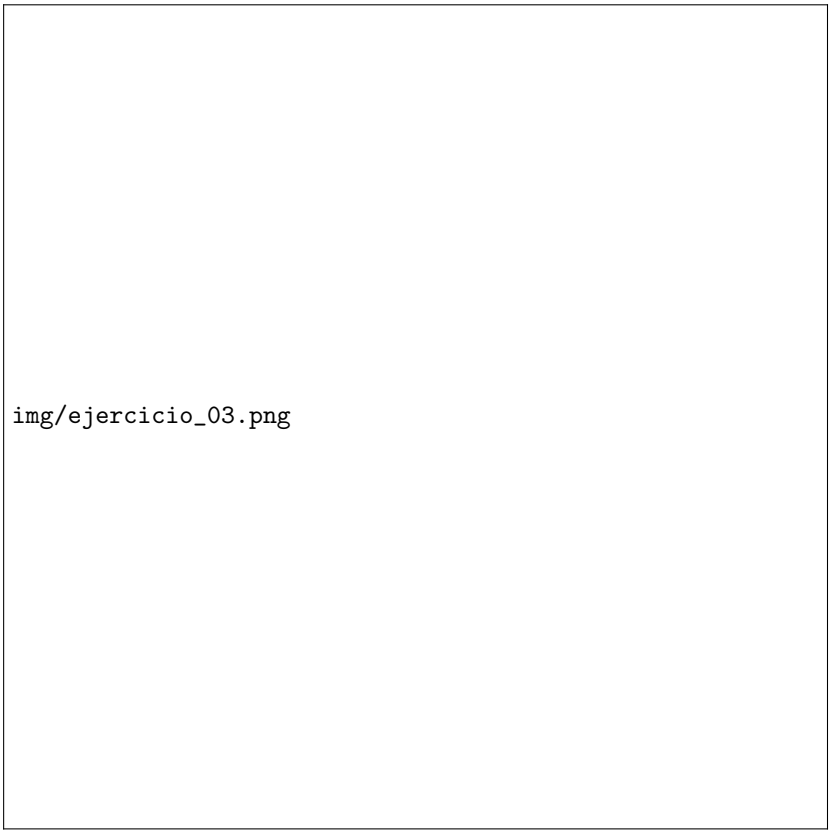

Figura 1: Evaluación de resultados en archivo *sketch.js*

5.2.2. Algoritmo kdtree

En la figura **??** se muestra la implementación de la función *closest_point_force_brute*.

Figura 2: Implementación de la función *closest_point_force_brute*

5.2.3. Algoritmo Descriptor "Bolsa de Palabras"

—

—

— —

[h!]
[width=0.7]img/ejercicio$_0$5$_c$losest$_p$oint$_b$rute$_f$orcepngEvaluaci\80\363ndedatosconlafunci\80\363nclosest\80\137point\
#I

5.2.4. Algoritmo Reductor de Dimensiones "MDS"

—

—

— —

[h!]
[width=0.7]img/ejercicio$_0$5$_c$losest$_p$oint$_b$rute$_f$orcepngEvaluaci\80\363ndedatosconlafunci\80\363nclosest\80\137point\
#I

6. Conclusiones

-
-
-

7. Referencias

1. Amalia Duch, Vladimir Estivill-Castro, Conrado Martínez, "Randomized K-Dimensional Binary Seach Trees", M-RR/LSI-98-48-R, Universitat Politècnica de Catalunya, Barcelona, 1998.
2. C. Martínez y S. Roura, "Randomized binary search trees", Journal of the ACM, Marzo 1998, Volumen 45, núm. 2, 288–323.
3. W. Cunto, G. Lau, y Ph. Flajolet, "Analysis of kd-trees: kd-trees improved by local reorganizations", In F. Dehne, J. R. Sack y N. Santoro editors, Work, Algorithms and Data Structures, 1989, Volumen 382, 24–38.
4. Andrew W. Moore, "An Introductory tutorial on kd-trees", Carnegie Mellon University, awm@cs.cmu.edu, 1992. http://www.autonlab.org.
5. Luc Devroye, Nicholas Broutin, Ketan Dalal y Erin McLeish, "The kdtreap", Personal Communication.
6. Redseguridad
7. ccn-cert
8. Geeksforgeeks