Detección temprana de correos maliciosos usando IA

Carlos Cortés Danta
Director adjunto de Ciberseguridad y Servicios Gestionados de Ingenia (a Babel Company)

Adrián Rodríguez García
Analista de Malware de Ingenia (a Babel Company)

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 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.

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

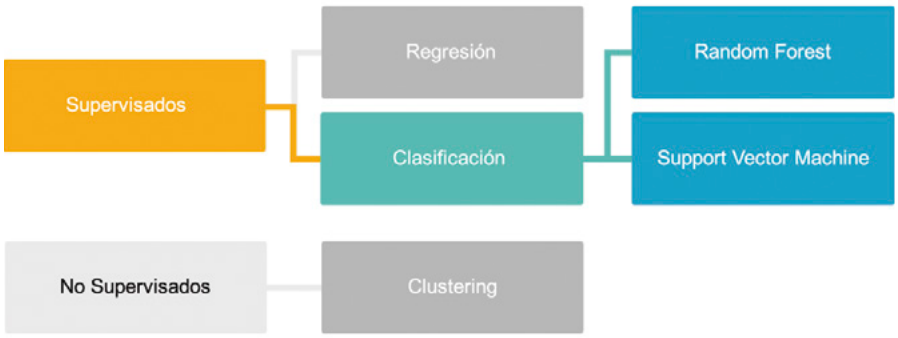
Otro documento de ENISA sobre ataques de *phishing* aporta un dato muy revelador: la ciberdelincuencia basada en la ingeniería social o *hacking* humano se ha amplificado un 600 por ciento durante la pandemia. Hay claramente, por tanto, un crecimiento significativo y alarmante.

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.

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.



Cuadro 1. Modelos del Machine Learning.

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.

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.

Cuadro 2. Modelos del Deep Learning.

Modelos

Entrando en materia con los modelos, *Machine Learning* se sirve de dos familias principales: los modelos supervisados y los no supervisados (ver cuadro 1). Los supervisados son aquellos en los que entrenamos al modelo con un número elevado de muestras que han sido previamente clasificadas. En base a ese aprendizaje, es capaz de predecir luego situaciones futuras.

Por su parte, los modelos no supervisados no cuentan con esas muestras ya clasificadas de antemano, y es el propio modelo el que se las ingenia para su clasificación en base a una serie de características comunes como la densidad, la cercanía, etc.

No obstante, los modelos de *Deep Learning* funcionan de manera distinta. Sus redes neuronales se clasifican en tres familias: redes supervisadas, no supervisadas y convolucionales (ver cuadro 2). Estas últimas se caracterizan por su efectividad en tareas de visión artificial, siendo muy usadas en la clasificación y segmentación de imágenes previo conocimiento.

En definitiva, hablar de IA es hablar de algoritmos complejos cuyos modelos deben recalibrarse continuamente para mejorar su eficiencia. Por tanto, es vital entrenarlos exhaustivamente y disponer de grandes cantidades de datos (correos electrónicos en este caso de uso) para afi-

La lucha contra las amenazas a las que se expone el correo electrónico necesita la aplicación complementaria de métodos de IA

nar lo mejor posible en las predicciones. No hay modelos mejores que otros, por lo que es primordial estudiar y comprender el problema bien para implementar los que más se adecuen a la situación.

Detección de ataques

El Área de Innovación en Ciberseguridad de Ingenia ha desarrollado sus propias herramientas de uso interno, en el ámbito del SOC, basadas en IA para la detección temprana de ataques que utilizan el correo electrónico para materializar su objetivo. Tras múltiples iteraciones con ensayos e implementaciones y una recalibración constante de los modelos para su perfeccionamiento, son los modelos supervisados de *Machine Learning* basados en la clasificación como *Random Forest* y *Support Vector Machine* los que han aportado mejores resultados en las predicciones.

En el ámbito del *Deep Learning*, el modelo de perceptrón multicapa ha resultado ser el de mayor rendimiento.

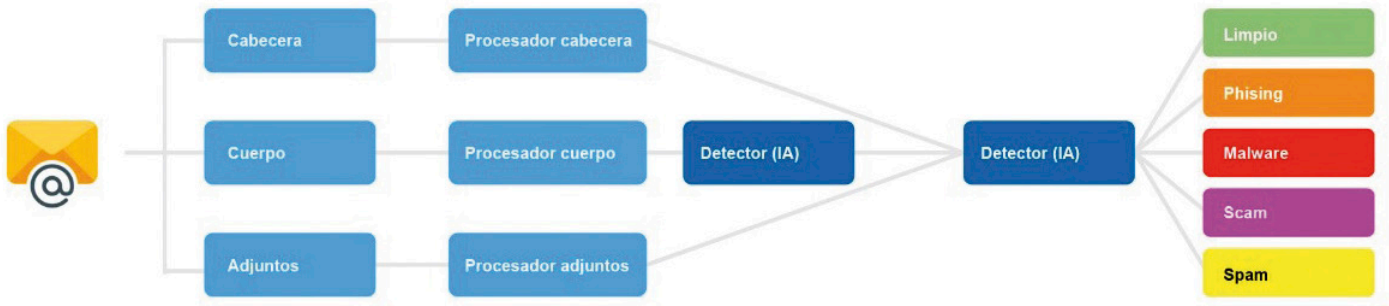
A grandes rasgos, la funcionalidad implementada consiste en extraer los datos y metadatos de un correo electrónico y clasificarlos en tres grandes bloques para su análisis (ver cuadro 3):

- **Cabecera**: Con los datos y metadatos procedentes de la cabecera intentamos obtener el origen real del correo o, en su defecto, una serie de indicios que aporten luz a la veracidad del remitente.
- **Cuerpo**: Se realiza un análisis riguroso de comportamiento sobre el texto donde formulamos cuestiones como ¿tiene el correo un hilo conductor general que tenga sentido?, ¿hay concordancia de género y número?, ¿tiene faltas ortográficas?, ¿parece escrito por una máquina? Las respuestas a estas preguntas nos pueden dar una idea del tipo de amenaza.
- **Adjuntos**: También son analizados para extraer información sobre reputación del archivo y posible contenido de *software* malicioso.

A partir del análisis de toda esta información se llevan a cabo una serie de operaciones matemáticas para obtener como resultado un vector con el que se alimenta el modelo de IA. Finalmente, la herramienta es capaz de devolver cinco resultados posibles: correo limpio (no malicioso), *phishing* (o ingeniería social), *malware* (*software* malicioso), *scam* (técnicas de extorsión al usuario) y *spam* (publicidad molesta).

Debido al gran volumen de ataques que se siguen observando a través del correo electrónico, el impacto significativo que pueden tener y la dificultad para detectarlos a tiempo, estas amenazas no pueden bloquearse solamente mediante enfoques clásicos. Por consiguiente, se hace indispensable la aplicación complementaria de métodos de IA con los que aumentar las capacidades de detección.

En el caso de uso expuesto, la aplicación está enfocada en la mejora de las capacidades de un SOC para prevenir y dar respuesta temprana a una de las principales amenazas para la continuidad del negocio en nuestras organizaciones.



Cuadro 3. Bloques de análisis de un correo electrónico.