IA EN CIBERSEGURIDAD

PRÁCTICA 2, MÓDULO 8

Descripción breve

Informe del estudio realizado sobre la aplicación de la Inteligencia Artificial en el mundo de la ciberseguridad.

JOSE LUIS MEZQUITA JIMENEZ

Máster en Inteligencia Artificial, Universidad Europea

Índice:

- Capítulo 1: Introducción.
- Capítulo 2: Aplicación de la IA a la ciberseguridad.
- Capítulo 3: Aplicación propuesta.
- Capítulo 4: Modelos empleados.
- Capítulo 5: Conclusiones.

Capítulo 1: Introducción

La Inteligencia Artificial (IA) es una de las tecnologías más revolucionarias y transformadoras de nuestro tiempo, con un impacto que se extiende a prácticamente todos los sectores de la sociedad. Su capacidad para imitar y superar funciones cognitivas humanas, como el reconocimiento de patrones, el razonamiento y la toma de decisiones, ha abierto un abanico de posibilidades en diferentes ámbitos.

Uno de los sectores donde la IA ha tenido un impacto muy significativo es la salud. A través de algoritmos avanzados, la IA ayuda a diagnosticar enfermedades con mayor precisión y rapidez, lo que mejora considerablemente los tratamientos personalizados para los pacientes. Además, sistemas de IA son capaces de analizar grandes cantidades de datos médicos, como imágenes o historiales clínicos, permitiendo detectar patologías de forma temprana.

En la industria automotriz, la IA impulsa el desarrollo de vehículos autónomos, que pueden operar sin intervención humana directa, aumentando la seguridad vial y transformando la movilidad urbana. Estos sistemas integran sensores y modelos predictivos para tomar decisiones en tiempo real.

La educación también ha visto cómo la IA facilita la personalización del aprendizaje. Plataformas educativas inteligentes adaptan los contenidos y ejercicios a las necesidades específicas de cada estudiante, mejorando su experiencia y resultados académicos.

En el ámbito del comercio, la IA permite optimizar la gestión de inventarios, predecir tendencias de consumo y mejorar la experiencia del cliente mediante recomendaciones personalizadas y asistentes virtuales que responden consultas en tiempo real.

Las finanzas se benefician de la IA en la detección de fraudes, la evaluación de riesgos crediticios y la automatización de procesos complejos, ayudando a tomar decisiones más seguras y eficientes.

El entretenimiento también ha sido transformado por la IA, que permite desde la creación de contenido personalizado hasta el desarrollo de videojuegos con comportamientos inteligentes y adaptativos.

En agricultura, la IA ayuda a optimizar cultivos mediante el análisis de datos climáticos y del suelo, facilitando prácticas más sostenibles y productivas.

Estas aplicaciones tienen en común el uso de técnicas de aprendizaje automático o Machine Learning, que consisten en entrenar modelos matemáticos a partir de datos para que aprendan patrones y puedan realizar predicciones o clasificaciones sin ser explícitamente programados para cada tarea específica.

Gracias al Machine Learning, los sistemas pueden mejorar su rendimiento con el tiempo, adaptándose a nuevos escenarios y automatizando procesos que antes requerían intervención humana directa. Así, la IA se posiciona como una herramienta esencial para la innovación y la mejora de la calidad de vida, transformando industrias y generando nuevos modelos de negocio en una sociedad cada vez más digitalizada.

Capítulo 2: Aplicación de la IA a la ciberseguridad.

La Inteligencia Artificial (IA) se ha convertido en un pilar fundamental en el área de la ciberseguridad, especialmente debido a la creciente complejidad y sofisticación de las amenazas digitales que enfrentan las organizaciones y usuarios hoy en día. En un entorno en constante evolución, donde los ataques cibernéticos se vuelven más frecuentes, avanzados y difíciles de detectar mediante métodos tradicionales, la IA ofrece soluciones innovadoras que permiten mejorar significativamente la defensa y protección de sistemas, redes y datos.

Una de las grandes ventajas de la IA en ciberseguridad es su capacidad para procesar y analizar grandes volúmenes de datos en tiempo real. Estos datos pueden provenir de diversas fuentes, como registros de acceso, tráfico de red, eventos de seguridad o información generada por dispositivos conectados. Al utilizar técnicas avanzadas de análisis, la IA es capaz de identificar patrones y comportamientos anómalos que podrían pasar desapercibidos para los sistemas convencionales o incluso para los analistas humanos. Esta detección temprana es clave para prevenir incidentes de seguridad antes de que se conviertan en ataques graves.

Entre las técnicas más empleadas en IA para ciberseguridad destacan el aprendizaje supervisado y no supervisado. En el aprendizaje supervisado, los modelos se entrenan con datos etiquetados que contienen ejemplos de actividades maliciosas y benignas, lo que permite clasificar y predecir nuevas amenazas basándose en ese conocimiento previo. En cambio, el aprendizaje no supervisado busca identificar anomalías o patrones desconocidos sin necesidad de etiquetas, lo que resulta especialmente útil para detectar amenazas nuevas o poco conocidas, como ataques de día cero.

Los sistemas inteligentes basados en IA pueden identificar diferentes tipos de amenazas cibernéticas, entre las que se incluyen el malware, que abarca virus, troyanos y ransomware; ataques de phishing, donde se intenta engañar a los usuarios para obtener información sensible; intrusiones en redes y sistemas, que buscan explotar vulnerabilidades para obtener acceso no autorizado; y otras actividades sospechosas que comprometen la integridad, confidencialidad y disponibilidad de la información.

Además, la IA no solo contribuye a la detección, sino que también mejora la capacidad de respuesta automática ante incidentes de seguridad. Los sistemas pueden tomar decisiones rápidas y precisas para contener ataques en curso, aislar componentes comprometidos, bloquear conexiones maliciosas o generar alertas para que los equipos de seguridad intervengan con mayor eficacia. Esto reduce significativamente el tiempo de respuesta y minimiza el impacto de los ataques.

La protección de infraestructuras críticas, como servicios financieros, sistemas de salud, redes eléctricas o gobiernos, es otro ámbito en el que la IA desempeña un papel fundamental. Estas infraestructuras suelen ser objetivos prioritarios para los ciberataques debido a su importancia estratégica y al alto costo que supondría una interrupción. Por ello, la implementación de sistemas de seguridad inteligentes resulta esencial para garantizar su resiliencia y continuidad operativa.

Además, con la proliferación de dispositivos conectados en el Internet de las Cosas (IoT), la superficie de ataque se ha ampliado considerablemente, haciendo indispensable el uso de IA para monitorizar y proteger estos entornos heterogéneos y distribuidos, donde las capacidades tradicionales de seguridad son insuficientes.

La IA también contribuye a la gestión y mitigación de riesgos en entornos digitales mediante la evaluación continua de vulnerabilidades y la priorización de medidas de seguridad. Al identificar qué activos son más críticos y cuáles presentan mayor riesgo, se pueden asignar recursos de forma más eficiente y establecer estrategias de defensa proactivas.

Por último, cabe destacar que la integración de la IA en ciberseguridad está en constante evolución y enfrenta desafíos importantes, como la necesidad de evitar falsos positivos, proteger la privacidad de los datos y asegurar que los modelos sean robustos frente a ataques adversarios que intentan engañarlos. Sin embargo, la capacidad de la IA para aprender, adaptarse y mejorar con el tiempo la convierte en una herramienta indispensable para proteger el entorno digital ante las amenazas actuales y futuras.

Capítulo 3: Aplicación propuesta.

Para realizar el análisis y la aplicación de técnicas de Machine Learning en este informe, se ha seleccionado el dataset conocido como NATICUSdroid, que se centra en el estudio de permisos otorgados a aplicaciones que operan en dispositivos con sistema operativo Android. Este dataset es especialmente valioso debido a la creciente importancia que tienen los permisos dentro del ecosistema móvil, ya que muchas aplicaciones pueden solicitar permisos que, si se usan de forma indebida, pueden comprometer la privacidad y seguridad del usuario.

El conjunto de datos NATICUSdroid contiene registros detallados de las aplicaciones y los permisos que estas solicitan para funcionar correctamente en el dispositivo. Estos permisos pueden incluir acceso a la cámara, micrófono, ubicación, contactos, almacenamiento, entre otros. Analizar estos permisos resulta fundamental porque muchas amenazas móviles, como el malware o aplicaciones maliciosas, tienden a abusar de ciertos permisos para realizar acciones no autorizadas o perjudiciales para el usuario.

El objetivo principal de utilizar este dataset es identificar patrones en la solicitud y uso de permisos que puedan diferenciar a las aplicaciones benignas (aquellas que no representan una amenaza) de las maliciosas, que intentan explotar estos permisos para fines dañinos. Por ejemplo, una aplicación aparentemente inofensiva puede solicitar permisos excesivos o innecesarios que no concuerdan con su funcionalidad principal, lo que podría indicar un comportamiento sospechoso.

Esta característica convierte al dataset NATICUSdroid en una herramienta útil para entrenar modelos de aprendizaje automático capaces de clasificar aplicaciones según su nivel de riesgo o potencial malicioso. Los modelos aprenden a reconocer combinaciones de permisos y características que suelen estar asociadas a aplicaciones maliciosas, ayudando así a mejorar la detección automática y la seguridad en dispositivos Android.

Además, el análisis basado en este dataset contribuye a entender mejor cómo se utilizan y solicitan los permisos en el ecosistema móvil, lo que puede ayudar a

diseñar mejores políticas de seguridad y recomendaciones para desarrolladores y usuarios finales.

El dataset representa un reflejo realista del comportamiento actual de las aplicaciones en el mercado, lo que permite que los resultados obtenidos tengan aplicabilidad práctica y sean relevantes para enfrentar amenazas reales en el ámbito de la ciberseguridad móvil.

Por tanto, NATICUSdroid no solo sirve como base para la clasificación, sino también como fuente de información para comprender tendencias y riesgos asociados al uso de permisos en aplicaciones Android, facilitando la implementación de soluciones más efectivas y adaptadas a los desafíos de seguridad modernos.

El dataset contiene información detallada sobre la presencia o ausencia de una gran variedad de permisos solicitados por aplicaciones Android. Cada permiso aparece con un conteo que indica cuántas aplicaciones no lo solicitan (valor 0) y cuántas sí lo solicitan (valor 1).

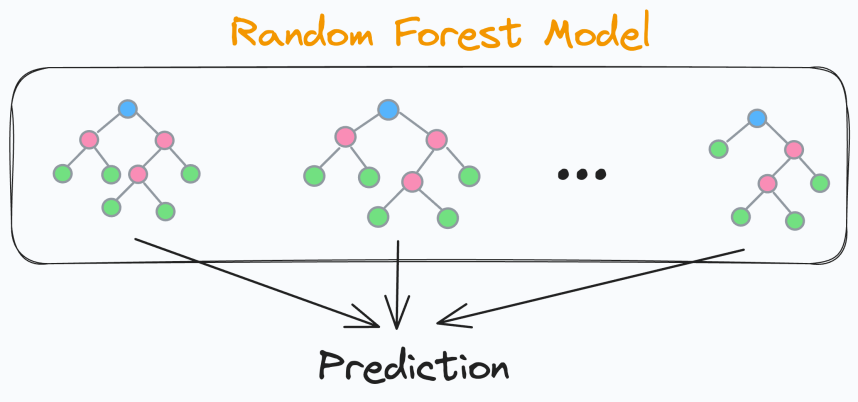
Por ejemplo, permisos comunes como **android.permission.INTERNET** son solicitados por la mayoría de las apps, mientras que otros más específicos, como **android.permission.MANAGE_ACCOUNTS**, se usan menos frecuentemente.

Este formato binario permite identificar patrones en el uso de permisos que pueden estar asociados con aplicaciones maliciosas o benignas. Además, el dataset está equilibrado en cuanto a la cantidad de aplicaciones benignas y maliciosas, lo que facilita entrenar modelos predictivos con un buen rendimiento para clasificar correctamente ambas clases. Así, el análisis de estos permisos y sus distribuciones es fundamental para detectar comportamientos anómalos y posibles amenazas en el entorno Android.

Capítulo 4: Modelos empleados.

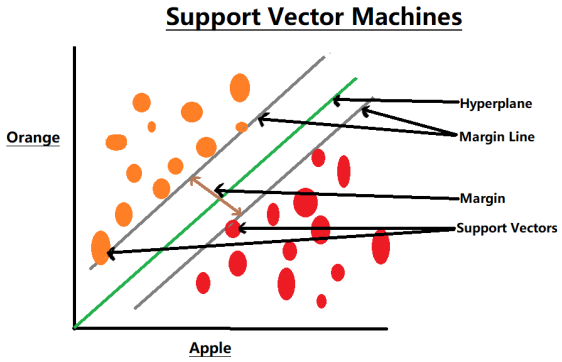
Para la clasificación de aplicaciones en el dataset NATICUSdroid, se han seleccionado dos algoritmos de Machine Learning ampliamente reconocidos por su eficacia y capacidad para manejar datos complejos: Random Forest y Support Vector Machine (SVM). Ambos métodos ofrecen enfoques complementarios para resolver problemas de clasificación, lo que permite comparar resultados y obtener mejores insights sobre el comportamiento del modelo.

Random Forest es un método de ensamblaje que construye múltiples árboles de decisión durante el proceso de entrenamiento y combina sus predicciones para mejorar la exactitud y robustez del modelo final. Cada árbol se construye usando un subconjunto aleatorio de datos y características, lo que reduce el riesgo de sobreajuste y mejora la generalización. Esta característica es especialmente valiosa para el dataset NATICUSdroid, donde las aplicaciones pueden solicitar diferentes combinaciones de permisos, y las relaciones entre estos pueden ser complejas y no lineales. Además, Random Forest permite calcular la importancia de cada permiso en la clasificación, lo cual ayuda a entender cuáles permisos son más indicativos de comportamientos maliciosos o benignos. Esta interpretabilidad resulta fundamental para la confianza en el modelo, ya que facilita el análisis de qué factores influyen más en la decisión.



Por otro lado, el **Support Vector Machine (SVM)** es un algoritmo que busca un hiperplano óptimo para separar las dos clases (aplicaciones benignas y maliciosas) maximizando la distancia o margen entre los puntos de datos más cercanos de cada clase. Esta maximización del margen ayuda a crear un modelo con buena capacidad de generalización. SVM es especialmente efectivo en conjuntos de datos con alta dimensionalidad, como es el caso de NATICUSdroid, donde cada permiso representa una característica que contribuye al espacio de clasificación. Además, SVM puede utilizar distintas funciones kernel (lineales, polinomiales, RBF, entre otras) para mapear los datos a espacios de mayor dimensión, donde una separación lineal es posible, incluso cuando las relaciones entre permisos y la

etiqueta no son lineales. Esto aporta flexibilidad y potencia al modelo para capturar patrones complejos.



Ambos algoritmos tienen ventajas y limitaciones. Random Forest suele ser más robusto frente al ruido y menos sensible a la selección de hiperparámetros, mientras que SVM puede requerir una mayor calibración para seleccionar el kernel y parámetros adecuados, pero puede ofrecer una separación más precisa cuando el margen es claramente definible. Utilizar ambos permite evaluar cuál se adapta mejor a las características del dataset y al problema de clasificación específico.

Capítulo 5: Conclusiones

Los resultados obtenidos al aplicar los algoritmos Random Forest y Support Vector Machine (SVM) para la clasificación de aplicaciones en el dataset NATICUSdroid demuestran un desempeño muy alto y comparable entre ambos métodos, aunque con una ligera ventaja para Random Forest.

El modelo Random Forest alcanzó una precisión global (accuracy) del 97.03%, con valores muy equilibrados en las métricas de precisión, recall y F1 Score, todos cercanos a 0.97. Esto indica que el modelo no solo es capaz de clasificar correctamente la mayoría de las aplicaciones, sino que también mantiene un buen equilibrio entre la detección de aplicaciones maliciosas y benignas. El análisis de la matriz de confusión confirma esta alta tasa de acierto, mostrando un bajo número de falsos positivos (71) y falsos negativos (103), lo que refuerza la robustez y confiabilidad del modelo.

Por su parte, el algoritmo SVM también mostró un rendimiento excelente con una precisión general del 96.13%, ligeramente inferior a la de Random Forest. Las métricas de precisión, recall y F1 Score permanecen muy balanceadas y cercanas al 0.96, reflejando un modelo efectivo y consistente. Sin embargo, la matriz de confusión revela un mayor número de errores (114 falsos positivos y 113 falsos negativos) en comparación con Random Forest, lo que puede influir en la capacidad de SVM para minimizar riesgos en aplicaciones críticas.

En términos prácticos, ambos modelos son adecuados para la clasificación de aplicaciones basándose en sus permisos, pero Random Forest presenta una mejor capacidad para generalizar y manejar las características del dataset, con un desempeño ligeramente superior en todas las métricas evaluadas. Además, su facilidad para interpretar la importancia de las variables puede ofrecer ventajas adicionales en la comprensión del comportamiento del modelo y en la identificación de permisos clave relacionados con actividades maliciosas.

En conclusión, Random Forest se posiciona como la opción más sólida para esta tarea específica, aunque SVM también constituye un método viable con resultados muy competitivos. La elección final puede depender del contexto de aplicación, la necesidad de interpretabilidad y la tolerancia al error en la detección de aplicaciones maliciosas.