Informe Técnico

SecureVision MX - IA y Software Inteligente para Seguridad Industrial

1. DATOS DEL PARTICIPANTE

- **Nombre Completo:** Diego López Facundo
- **Correo Electrónico:** diego_lopez233@outlook.com ó a302486@alumnos.uaslp.mx
- **Universidad y Carrera:** Universidad Autónoma de San Luis Potosí - Ingeniería en Sistemas Inteligentes.

- **Fecha de Entrega:** 12 de febrero del 2025
-

2. INTRODUCCIÓN

- Explica brevemente tu enfoque para resolver el reto.

El sistema propuesto aborda tres frentes críticos en seguridad industrial:

1. **Detección de anomalías en procesos productivos** mediante el modelo **MemAE** (Memory-augmented Deep Autoencoder), entrenado con el dataset **IPAD** (Industrial Process Anomaly Detection), que incluye 16 tipos de dispositivos industriales y anomalías en forma del producto, anomalias en el color, mal manejo del material, desviaciones periódicas, entre otras. Esto ayudará a la supervisión de las maquinas en producción.

2. **Vigilancia de intrusiones y comportamientos anómalos** usando el framework **Real-world-Anomaly-Detection-in-Surveillance-Videos-pytorch**, entrenado con datos de actividades como intrusiones no autorizadas, movimientos en zonas restringidas y accidentes laborales. Este sistema mantendrá una vigilancia perimetral y en el interior de forma automatizada identificando situaciones de peligro o fuera de lo común.

3. **Detección de anomalías en redes** con un modelo híbrido **Hybrid Isolation Forest** (basado en el trabajo de Pierre Marteau), es una implementación propia que se optimizó en tiempo y memoria para identificar patrones sospechosos en flujos de red. Ideal para mantener seguridad en redes locales en donde trabajen maquinas o donde se manejes importantes flujos de información y evitar posibles ciberataques que frenen la producción o filtren información privada de la empresa.

Además, para un mejor manejo de datos se implementó:

1. **Agente inteligente para gestión de alertas** integrado en **LangChain** con el modelo **Ollama** acorriendo localmente para generar reportes automatizados, responder consultas en lenguaje natural y priorizar incidentes.

2. **API unificada** que lo vuelve un sistema centralizado en **Flask** y conecta todos los módulos, gestiona los datos en **JSON** y deja espacio a integraciones futuras con dispositivos IoT (ej: cerraduras automáticas, drones).

- Justifica por qué elegiste la solución propuesta.

Me pareció conveniente hacer un sistema centralizado en un agente inteligente ya que en gestión de seguridad es muy importante el tiempo de reacción y actuar de forma lógica y racional, ante situaciones en las que a los humanos nos constaria mantener la compostura o un pensamiento crítico, cosa que puede hacer a un agente AI buen candidato para la gestión general del sistema.

Esto no solo facilita el analisis y desgloce de la información que le llega por medio de cámaras, sensores, drones y, modelos de machine learning y deep learning que implementa, si no que nos da la posibilidad de automatizar acciones y tener tiempos de reacción ante adversidades mucho mas bajos. Además, facilita la sintetización de la información para personas sin muchos conocimientos técnicos sobre ciberseguridad, analisis de resultados de modelos y nos porporciona una vision general de la seguridad de la planta. El sistema tiene una base sólida para evolucionar hacia un SOC (Security Operations Center) autónomo

3. DESARROLLO TÉCNICO

- ¿Cómo funciona tu sistema? Explica en términos generales.

El sistema esta compuesto por 3 modelos de deteccion de anomalias diferentes, **MemAE** que fue utilizado en la detección de anomalias en las cadenas de producción, **I3D (Inflated 3D ConvNet)** para la deteccion de situaciones de peligro o sospechosas en la vigilancia con cámaras, y el **HIF (Hybrid Isolation Forest)** como algoritmo de deteccion de anomalías híbrido que aprende tanto de forma no supervisada con el estado normal de la red e historial de anomalias y de forma supervisada con el dataset **CICIDS_2017** con ataques de red simulados.

Estos modelos estan corriendo en el backend y envian alertas a la interfaz cada que se detecta una anomalía, además de guardar su historial de anomalias y demas informacion relevante. El agente inteligente por el momento solo funciona para la generacion de reportes y consultas sobre los datos inmediatos que arrojan los modelos, pero se dejó espacio para la implemenatación de actuadores y herramientas para automatizar procesos de respuesta a incidentes.

Como llm para el agente inteligente se utilizó Llama3.1 de Meta corriendo el modelo localmente, en caso de poder subirlo a un servidor se podría dar servicio a la empresa con un llm con fine-tunning para que sea experto en los temas de seguridad y ademas conozca más información de la empresa.

- ¿Qué tecnologías y herramientas usaste?

Python, Pytorch, Scikit-learn, OpenCV, Flask, LangChain, Ollama, HTML, CSS y Javascript.

- ¿Cómo se integran los diferentes componentes del sistema?

Los componentes se comunican entre si mediante un API implementada con python usando Flask, aquí se corren las inferencias de los modelos y el agente inteligente, además de manejarse la informacion mediante JSON.

- Código estructurado en GitHub con explicación breve.

El codigo se estructuro en una carpeta “modelos” en donde se encuentran los archivos y datasets que se utilizaron para entrenar los modelos, una carpeta “memae_env” donde se guarda un entorno virtual configurado para este codigo en especifico. La carpeta “flask-dos-service” guarda la aplicacion flask completa donde se encuentra el código para correr los modelos, la API de flask y el código de la interfaz web, así como los archivos con los que trabaja.

- Explicación del modelo de IA utilizado.

MemAE (Memory-Augmented Autoencoder): Es un **autoencoder mejorado con memoria externa**, diseñado para detectar anomalías en entornos industriales. Su arquitectura incluye un *encoder* que comprime datos (ej: imágenes de productos o lecturas de sensores) en un espacio latente, una *memoria* que almacena patrones de operación normales, y un *decoder* que reconstruye la entrada usando dichos patrones. Las anomalías se identifican cuando la reconstrucción difiere significativamente de la entrada original (usando métricas como **MSE**). Es ideal para detectar defectos en cadenas de producción o fallas en maquinaria.

I3D (Inflated 3D ConvNet): Modelo de **red neuronal convolucional 3D** para análisis de video. Utiliza convoluciones que procesan simultáneamente dimensiones espaciales (píxeles) y temporales (secuencia de fotogramas). Preentrenado en datasets de acciones humanas (ej: **Kinetics**), detecta comportamientos anómalos en vigilancia (ej: caídas, intrusiones) al identificar patrones de movimiento inusuales. Su fortaleza radica en capturar contexto dinámico, como gestos o desplazamientos sospechosos.

HIF (Hybrid Isolation Forest):
Algoritmo híbrido que combina **aprendizaje no supervisado** (mediante *Isolation Forest*, que mide la facilidad para aislar datos anómalos) y **supervisado** (entrenado con el dataset **CICIDS_2017**, que incluye ataques simulados como *Botnet Attacks* o *SQL injection*). Detecta anomalías genéricas en redes (tráfico inusual) y amenazas específicas (ej: *DDoS*), optimizando velocidad y precisión al integrar ambos enfoques.

Automatización y Optimización

- ¿Cómo optimiza el consumo de recursos?

El sistema tiene que tener como requisito funcional la velocidad de respuesta por lo que utilizar modelos en librerias como pytorch y que fueron previamente optimizados fue una prioridad, se

utilizaron algunas tecnicas de feature engineering y parallelización en gpu y cpu para el entrenamiento de los modelos y se usan técnicas similares para las inferencias.

- ¿Cómo gestiona la seguridad y las alertas?

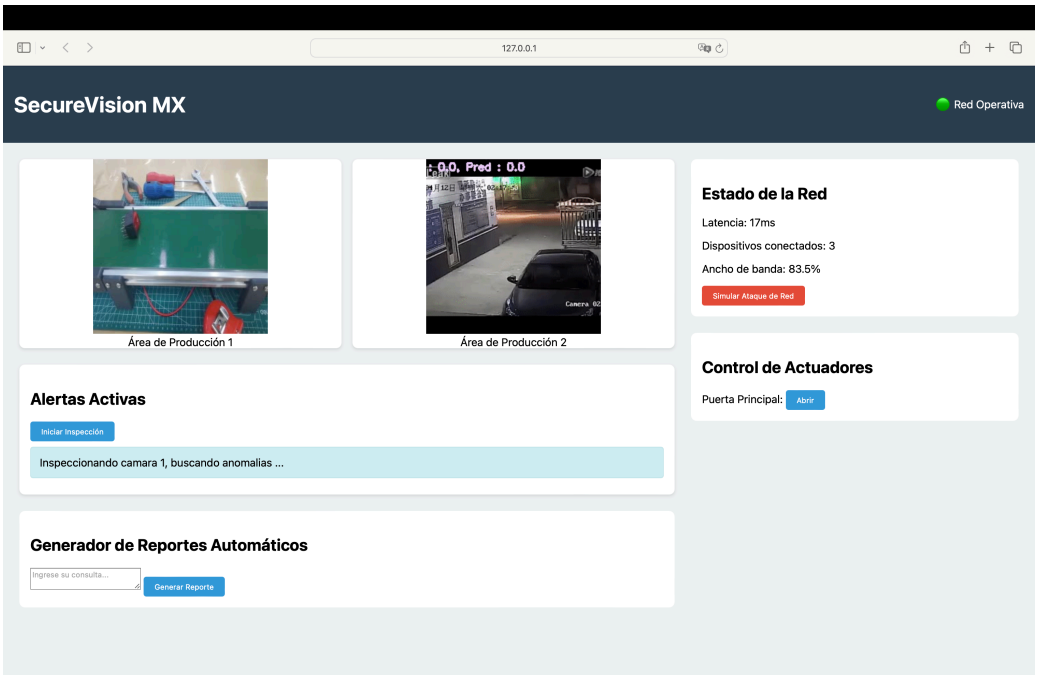
Cuando cualquiera de los modelos captura una alerta, esta se guarda en un archivo JSON compartido, esta informacion es utilizada por el agente inteligente para hacer los reportes y ultimadamente para automatizar acciones de respuesta. Además, en la interfaz se muestran las alertas en tiempo real.

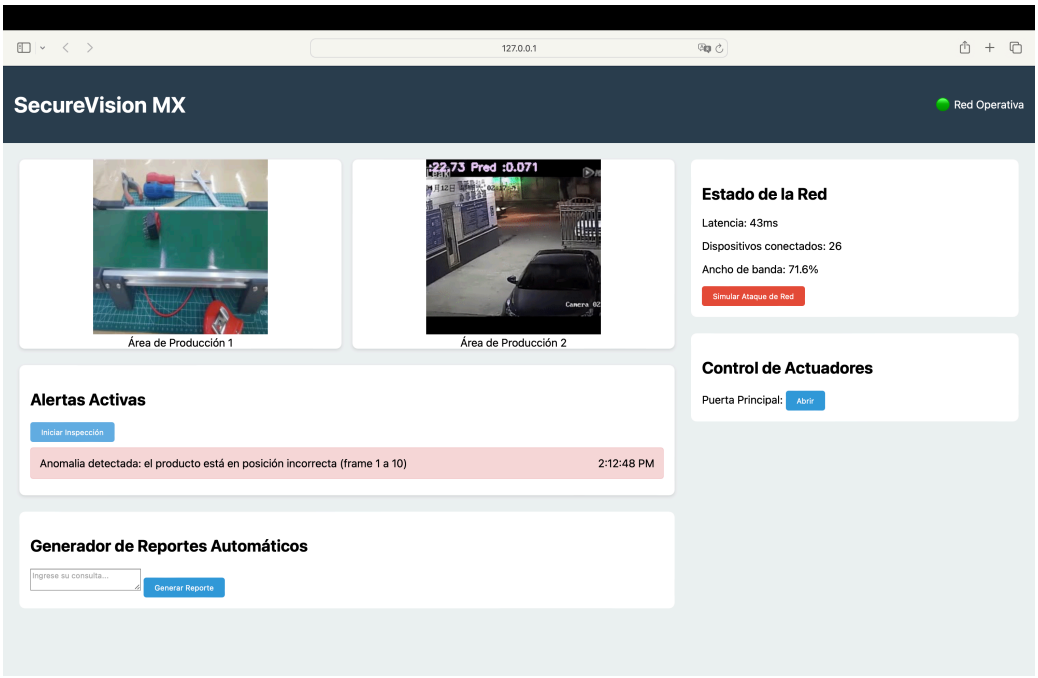
- ¿Qué mejoras de eficiencia propusiste?

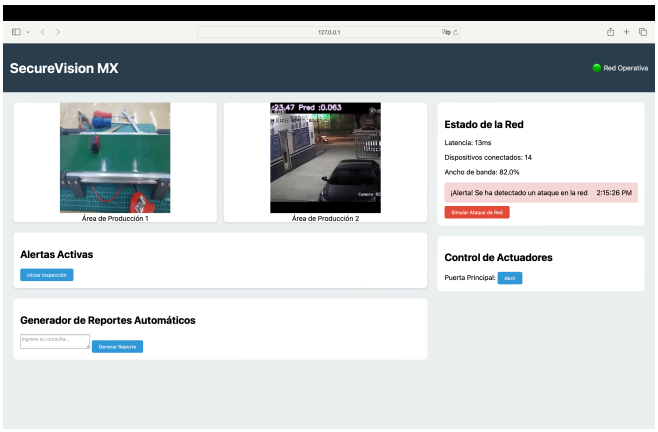
Automatizar la vigilancia tanto de las camaras de seguridad como del trafico de red agiliza los procesos de seguridad al dar respuesta a incidentes de forma inmediata, tener toda la información se seguridad centralizada y poder hacer consultas a un sistema experto con lenguaje natural y reduce el error humano en vigilancia.

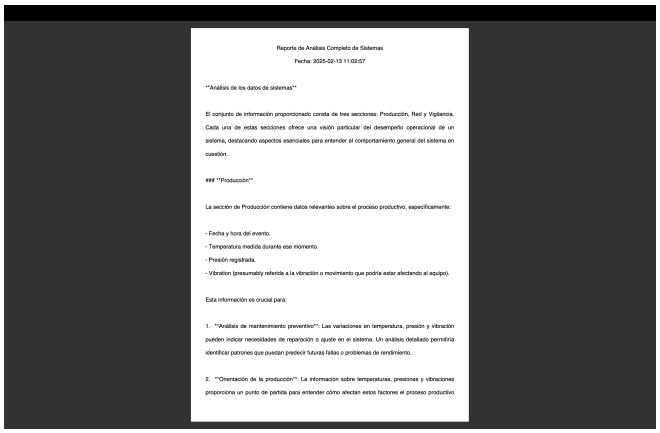
4. EVALUACIÓN DE RESULTADOS

- Pruebas realizadas (capturas de pantalla o videos).









- Principales retos y cómo los superaste.

El principal reto fue seleccionar la información y los modelos correctos para cada caso de uso y entrenarlos, por cuestión de tiempo y eficiciencia me apoyé de frameworks de machine learning y descargue modelos pre-entrenados así como datasets específicos para mi caso de uso.

- ¿Cómo podría escalarse el sistema a nivel industrial?

Este sistema puede ser escalado a nivel industrial al utilizar camaras y sistemas de vigilancia reales, además de poder trabajar con otros formatos como sonido o sensores de temperaturas, entre otros. Se podría integrar con cualquier dispositivo IoT para hacerlo

accionar cuando el agente inteligente lo mande a llamar y al tener un modelo de forma local, es posible entrenarlo con cuanta información haya disponible para que sea un experto en casos especificos de la empresa.

5. PREGUNTAS CLAVE

- Hábitos y Rutinas: ¿Cómo te organizaste para resolver el caso?

Siendo una semana ocupada con citas medicas, compromisos familiares y la renovacion de mi visa en Monterrey, estuve trabajando en mis tiempos libres y en tiempos de traslado. Primero hice un analisis general de los requerimientos y una busqueda de posibles soluciones a cada uno de los problemas planteado. Me tomó la mayor parte del tiempo la busqueda, filtración y pre-procesamiento de datos adecuados para este problema especifico y la implementacion de las API.

- Resiliencia: ¿Qué mejorarías si tuvieras 72 horas más?

Mejoraría la interfaz gráfica e implementaría las herramientas y lógica necesaria para que el agente IA pudiera tener acceso a dispositivos de seguridad IoT simulados, a la red local para hacer bloqueos a ips maliciosas y a una base de datos en donde pueda consultarse la informacion periodicamente.

- Visión a Futuro: ¿Cómo conectarías este caso con tu carrera profesional?

Mi objetivo es convertirme en un profesionista enfocado en la implementación de modelos de Machine Learning y DeepLearning, así como uso de APIs de Grandes Modelos de Lenguaje y la creación de Agentes Inteligentes, por lo que más que una tarea tediosa lo veo como un reto y una oportunidad para mejorar mis habilidades técnicas. Seguiré implementando proyectos, investigando y estudiando este campo de la informática. Me parece una oportunidad muy interesante ya que la inteligencia artificial esta evolucionando a gran velocidad y me gusta estar actualizado con las nuevas tecnologías.

- Trabajo en Equipo: ¿Cómo estructurarías este proyecto con más personas?

Cuando trabajo en equipo me gusta tener tareas especificas para cada miembro del equipo, tratando de aprovechar sus habilidades o experiencia en cada tipo de desarrollo y uso de herramientas, posiblemente trabajaría de forma modular y usando una metodología agil como SCRUM para la correcta captura de requerimientos, documentación y asignación de tareas específicas.

6. CONCLUSIÓN

- Resumen de lo aprendido.

Durante la investigación e implementación de este reto aprendí mucho sobre los algoritmos de visión por computadora y de como manejar datasets y archivos. Me queda como experiencia

para la resolución de otros problemas similares. Además, me obligó a investigar y comprender mejor como funciona una planta industrial y cuales serían sus necesidades reales de seguridad.

- Propuesta de siguientes pasos para mejorar la solución.

Para mejorar el sistema propondría implementar otro modelo llamado YOLOv8 para hacer detección de objetos o personas específicas en las escenas anomalas y poder obtener mas información valiosa de la situación y contexto. También le haría fine-tuning al modelo Ollama para que tuviera mejores conocimientos sobre ciberseguridad y seguridad perimetral en general y optimizar sus respuestas. Agregaría uno o varios accionadores de prueba para optimizar el uso de herramientas por medio del agente inteligente, probablemente me centraría en que además de que los modelos puedan dar predicciones correctas, tambien el agente inteligente sepa que hacer en cada caso y aprendiendo poco a poco de su entorno y contexto.

7. ENLACE DE ENTREGA

- Repositorio GitHub: https://github.com/Dicotomico23/SecureVisionMX