Convocatoria Colaboración Público-Privada CPP2021-009025

DICYME:

Dynamic Industrial CYberrisk Modelling based on Evidence

Iniciativa Conjunta de:

Universidad Rey Juan Carlos

ENTREGABLE 1.4:
Documentación de las técnicas y transformaciones de datos en información para su consumo en modelos de ciber riesgo OT

Coordinadores:

Romy R. Ravines (DeNexus Tech)
Isaac Martín de Diego (Universidad Rey Juan Carlos)
Alberto Fernández Isabel (Universidad Rey Juan Carlos)

Contenido

1 INTRODUCCIÓN Y OBJETIVOS

Este entregable detalla las técnicas y transformaciones utilizadas para convertir datos brutos en información útil que puede ser consumida por los modelos de ciber riesgo desarrollados en el marco del proyecto DICYME. La actividad se centra en describir los procesos que permiten estructurar, enriquecer y validar los datos provenientes de múltiples fuentes internas y externas, asegurando su compatibilidad con los modelos de probabilidad e impacto. El objetivo es garantizar que los datos transformados proporcionen una base sólida y confiable para el análisis dinámico del ciber riesgo en infraestructuras industriales.

En las siguientes secciones se explica:

1. Qué tipo de datos se procesan, los datos internos y externos utilizados incluyendo vulnerabilidades, registros de seguridad, características firmográficas y menciones en medios, y cómo se preparan para su análisis.
2. Cómo los datos se transforman en información útil para los modelos de ciber riesgo, el proceso de limpieza, normalización, y enriquecimiento mediante técnicas como la minería de reglas.
3. Qué técnicas y procesos de minería de reglas se han implementado. Se documenta cómo se extraen reglas asociativas de los datos usando métricas de soporte, confianza y *lift*, y cómo estas reglas se integran como elementos clave para modelos de probabilidad e impacto.
4. Qué datos y código se encuentran disponibles en este entregable (deriskGroup / DICyME Project · GitLab) y cómo pueden emplearse.

2 FUENTES DE DATOS Y CONTEXTO

El análisis y transformación de datos documentados en este entregable se basan en los descritos en los entregables previos (1.1, 1.2 y 1.3). Estas fuentes fueron seleccionadas por su relevancia para modelar el ciber riesgo en entornos OT, se integran en un pipeline automatizado que garantiza la recopilación continua y la generación de información útil para los modelos de probabilidad e impacto.

El objetivo principal es transformar estos datos en información estructurada que pueda alimentar los modelos de ciber riesgo. Para ello, se identifican patrones y relaciones entre atributos mediante técnicas de minería de reglas o algoritmos de aprendizaje automático.

3 TÉCNICAS DE TRATAMIENTO Y TRANSFORMACIÓN

El procesamiento de los datos recopilados de fuentes internas y externas requiere una serie de técnicas para garantizar su utilidad en los modelos de ciberriesgo OT. Estas técnicas abarcan desde la limpieza inicial y la normalización, hasta la transformación y enriquecimiento de los datos. El objetivo es convertir datos brutos en información estructurada y enriquecida que permita realizar análisis precisos y efectivos sobre probabilidad e impacto de incidentes para la ciberseguridad.

3.1 AT2ATK: Atractivo

3.1.1 Preprocesamiento de datos

Para llevar a cabo el preprocesamiento de los datos, se llevan a cabo las siguientes técnicas y pasos:

3.1.2 Limpieza de datos

- Eliminación de valores nulos, duplicados y registros incompletos. Cuando se ha optado por mantener registros con algún valor nulo, estos se han tratado como *NaN* o cadenas vacías "", en el caso de campos numéricos o de texto, o como listas vacías en el caso de campos de tipo lista.
- Corrección de formatos no estándar, como fechas o identificadores mal estructurados.

3.1.3 Estandarización

- Normalización de atributos clave, como información crítica encontrada en brechas de datos.
- Conversión de datos categóricos en formatos binarios o numéricos, según los requisitos de los algoritmos utilizados posteriormente.
    - Para el modelo de atractivo, los campos de facturación, ganancias y empleados se han discretizado empleando los cuartiles, resultando 4 grupos en cada variable.

3.1.4 Filtrado de datos

- Selección de características relevantes, como impacto financiero o activos expuestos, para reducir la complejidad del análisis.

3.1.5 Enriquecimiento

Este proceso aplica dentro de atractivo de dos maneras diferentes. Por un lado, se utilizan las fuentes de datos *RocketReach* y *CompaniesMarketCap* de manera complementaria. La primera de ellas, *RocketReach*, proporciona valores para el país y sector de la entidad, su facturación anual, cantidad de empleados y cotización en bolsa. Su base de datos contiene una cantidad de entidades muy elevada. Por otro lado, *CompaniesMarketCap* proporciona también información sobre país y sector, así como facturación anual, ganancias anuales, cantidad de empleados y cotización en bolsa. Sin embargo, su base de datos se reduce principalmente a empresas cotizadas. Como es evidente, algunos valores son proporcionados por ambas fuentes de datos. Cuando, para una entidad, se dispone de datos de ambas, prevalece el de *RocketReach*, por ser la fuente de datos principal para la totalidad de entidades buscadas. Sólo en el caso de existir valores faltantes en *RocketReach*, o para aquellos campos que únicamente proporciona *CompaniesMarketCap*, como las ganancias anuales, prevalece esta última.

No obstante, se han comprobado manualmente las posibles desviaciones entre campos comunes a ambas fuentes, y cuando se han encontrado desviaciones reseñables, se ha comprobado en fuentes abiertas el valor más exacto y corregido.

Por otro lado, dentro del procesado de datos del modelo de atractivo también se ha producido enriquecimiento para la obtención de los "no incidentes", o entidades altamente similares a las víctimas de ciber incidentes. Este proceso se realiza empleando *RocketReach* de nuevo, pues para una cantidad muy elevada de entidades devuelve otras similares a la analizada. De manera que la calidad de este nuevo dato sea la mayor posible, se comprueba que la nueva empresa tenga un nombre, que la categoría general de industria coincida con la de la entidad víctima del incidente y que tenga al menos algún valor para el campo facturación. Si alguna condición no se cumple o *RocketReach* no conoce empresas similares a la víctima, simplemente no se extraen datos de no incidente para esa víctima.

Por otra parte, la combinación de datos internos y externos contribuye a generar indicadores compuestos, como el índice de atractivo explicado en el entregable E2.2, basado en atributos como la popularidad, tamaño de la empresa y exposición.

Ilustración 1. Ejemplo del Indicator ATR2ATT

3.2 CVE2ATK: Vulnerabilidades

La información necesaria para este módulo se ha descrito en el entregable E1.3. En particular se utilizan las descripciones de los CVEs obtenidas de CAPEC y las descripciones de las técnicas de MITRE ATT&CK. Como preparación del dato para el modelo del mapeo, se realiza la clasificación de vulnerabilidades en 13 clases.

En el caso de la detección de vulnerabilidades (CVE), se realiza un mapeo sobre la matriz MITRE ATT&CK proporciona un contexto adicional sobre las posibles técnicas de ataque como se describe en el entregable E2.1 en detalle.

- Cadenas de texto construidas: CVE-NVD/CWE/CAPEC/TTs combinando varias fuentes de datos
- Limpieza y tokenización de descripciones de CVEs y técnicas de MITRE.

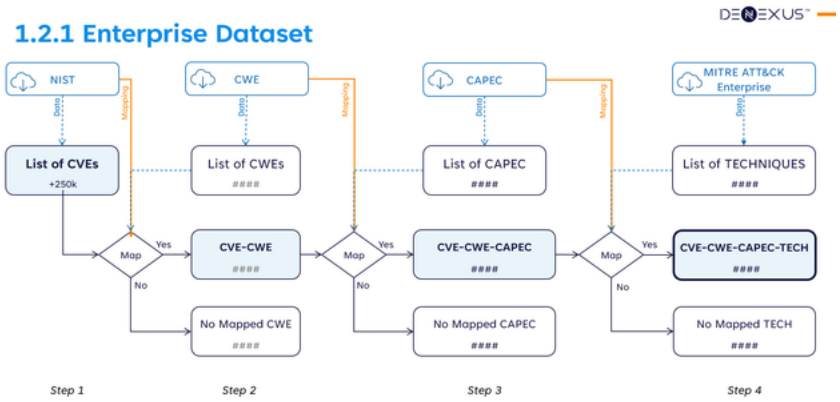


Ilustración 2. Construcción del conjunto de datos para CVE2ATK

El mapeo final se realiza con modelos de *machine learning* apropiados para el trabajo de textos como datos. Específicamente se usa modelo de red neuronal de dos capas para mapear automáticamente los CVE a las técnicas ATT&CK, de dos de sus matrices (Enterprise y ICS). Abordamos el problema de la falta de etiquetas para esta tarea, aprovechando la información disponible para la matriz Enterprise, en combinación con el conocimiento de expertos en la materia para la matriz ICS. Evaluamos el enfoque con el conjunto de datos que contiene la lista completa de CVE. Utilizando el modelo propuesto, mapeamos todos los registros de CVE a todas las técnicas ATT&CK de ambas matrices Enterprise e ICS.

El resultado principal de este trabajo es un sistema que aprovecha algoritmos de aprendizaje profundo que utilizan descripciones de vulnerabilidades, tipos de vulnerabilidades y descripciones de técnicas MITRE para vincular los CVE a las técnicas MITRE en ambas matrices, Enterprise e ICS.

El sistema recupera información de bases de datos públicas, periódicamente y vuelve a entrenar los algoritmos para actualizar la herramienta de mapeo.

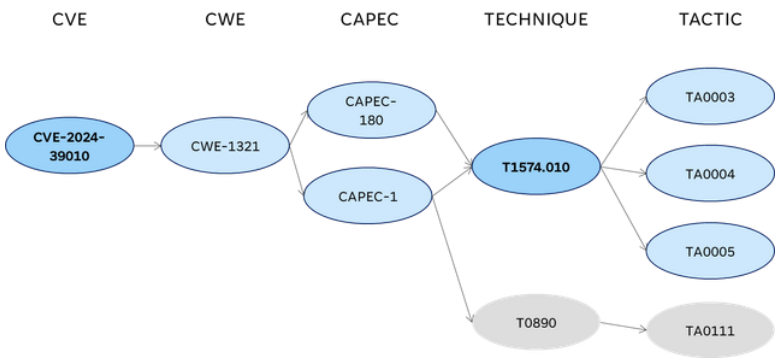


Ilustración 3.Ejemplo de mapeo para CVE-2024-39010. La última columna corresponde a Tácticas, donde las técnicas aparecen en MITRE ATT&CK. En gris, se muestra una técnica de ICS para completarla información

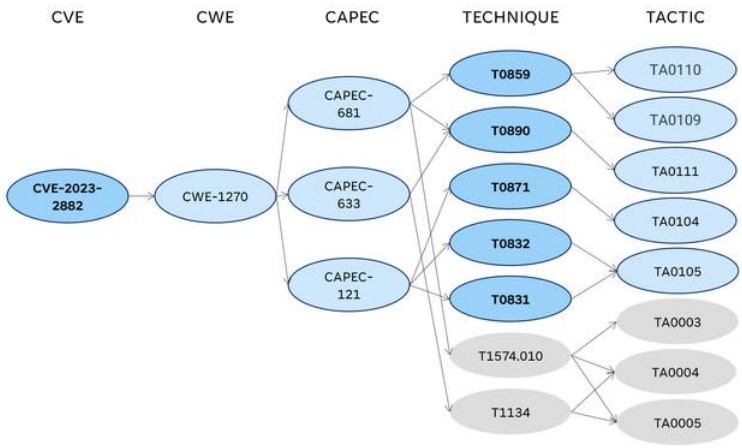


Ilustración 4. Ejemplo de mapeo para CVE-2023-2882. La última columna corresponde a Tácticas, donde aparecen las técnicas en MITRE ATT&CK. En gris, se muestran las técnicas de Enterprise para completarla información

3.3 THRACT: Threat Actor Score

La puntuación del actor se compone de tres puntuaciones:

- *Activity Score*: Puntuación de actividad
- *Capacity Score*: Puntuación de capacidad
- *Target Score*: Puntuación de objetivo

La fórmula para la puntuación del actor *(Actor Score*) es:

$$\text{Actor Score} = \text{Activity Score} * mean(\text{Capacity Score}, \text{Target Score})$$

Los indicadores parciales son:

$$\text{activity\_score} = \begin{cases} 1.0, & \text{if last\_activity\_date} \leq 30 \text{ days} \\ 0.8, & \text{if } 30 < \text{last\_activity\_date} \leq 90 \text{ days} \\ 0.6, & \text{if } 90 < \text{last\_activity\_date} \leq 365 \text{ days} \\ 0.3, & \text{if last\_activity\_date} > 365 \text{ days} \end{cases}$$

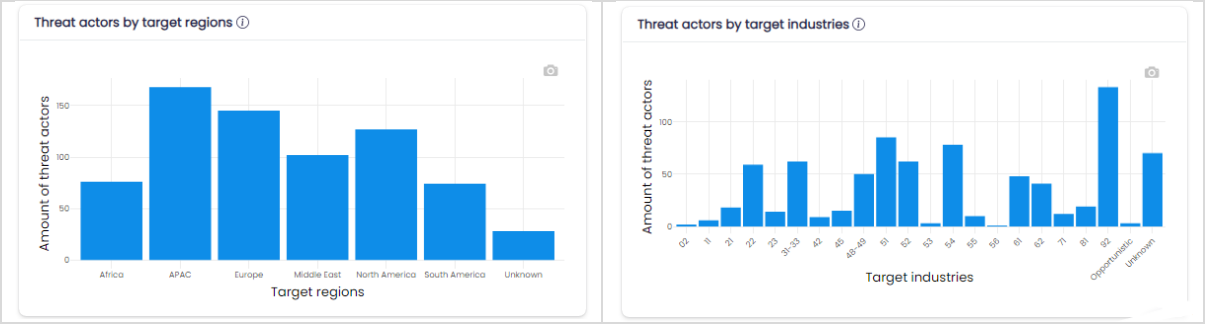
Donde *last_activity_date* es la fecha (dd/mm/yyyy) en la que se ha reconocido actividad de dicho actor.

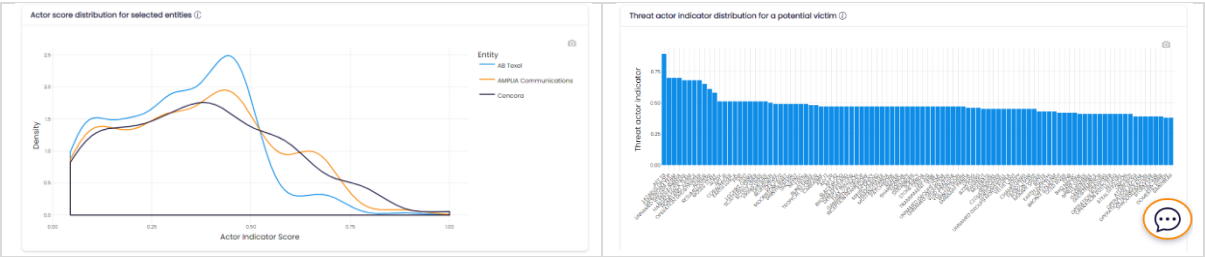
El target score es:

$$\text{Target Score} = \frac{w_{\text{country\_ts}} * \text{country\_ts} + w_{\text{industry\_ts}} * \text{industry\_ts} + w_{\text{motivations\_s}} * \text{motivations\_s}}{w_{\text{country\_ts}} + w_{\text{industry\_ts}} + w_{\text{motivations\_s}}}$$

donde:

- $$\text{country\_target\_score} = \begin{cases} 1.0, & \text{if target county is in target\_countries\_values} \\ 0.6, & \text{if some adjanced county is in target\_countries\_values} \\ 0.4, & \text{otherwise'} \end{cases}$$
- $$\text{industry\_target\_score} = \begin{cases} 1.0, & \text{if target industry is in target\_industries\_values} \\ 0.4, & \text{otherwise} \end{cases}$$
- $$\text{motivations\_score} = \begin{cases} 1.0, & \text{if motivations\_values is equal to 'Criminal'} \\ 0.8, & \text{if motivations\_values is equal to 'State-Sponsored'} \\ 0.5, & \text{if motivations\_values is equal to 'Hacktivism'} \\ 0.1, & \text{otherwise'} \end{cases}$$

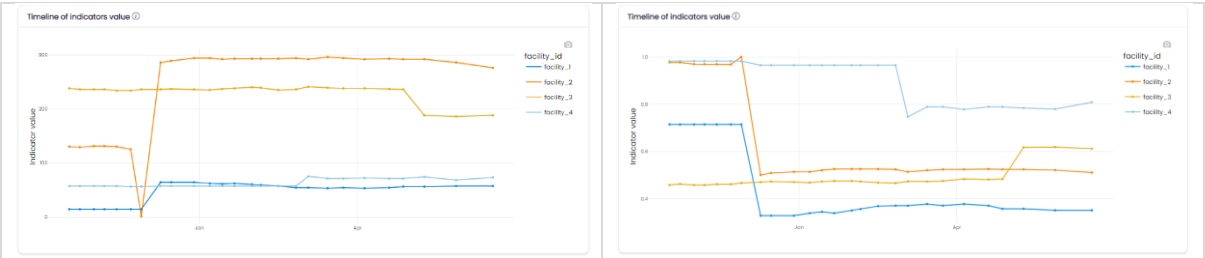




3.4 Indicadores de evidencias internas

En el entregable *E1.1* del proyecto se definieron diversos indicadores para la cuantificación del ciber riesgo industrial. Estos indicadores se generan a partir de evidencias internas extraídas de diferentes fuentes, como sistemas EDR (*Endpoint Detection and Response*), IDPS (*Intrusion Detection and Prevention System*) y SIEM (*Security Information and Event Management*). Las técnicas y transformaciones aplicadas a los datos no están separadas de la definición de estos indicadores, sino que forman parte intrínseca del proceso, ya que involucran agrupaciones, conteos y otras operaciones que permiten convertir datos brutos en información estructurada y relevante para la gestión del ciber riesgo.

El primer prototipo de este módulo ha trabajado principalmente con datos provenientes de sistemas IDPS en entornos industriales, seleccionando aquellos que proporcionan una visión de nivel medio sobre los activos, sus vulnerabilidades y su estado de seguridad. Estos datos han sido procesados y transformados en métricas concretas, como el recuento de activos evaluados en busca de vulnerabilidades, la cantidad de vulnerabilidades abiertas y el tiempo medio de resolución de vulnerabilidades. Estas métricas, al estar basadas en técnicas de agregación y clasificación, permiten alimentar los modelos de cuantificación del riesgo que se desarrollarán en otras actividades del proyecto.



4 DATOS Y CÓDIGO DISPONIBLES

El repositorio del proyecto es *deriskGroup / DICyME Project · GitLab* La carpeta correspondiente a este entregables es *E1.4 Documentation of Techniques and Data Transformations into Information for Use in OT Cyberrisk Models*.

Cabe aclarar algunos aspectos sobre el entorno de desarrollo empleado para generar el prototipo. Se han utilizado:

- *Python 3.9.5*
- *Databricks Runtime 12.2 LTS*
- *Apache Spark 3.3.2*
- *Pandas 1.4.2*
- *Numpy 1.21.5*
- *Beautiful Soup 4.12.3*
- *Matplotlib 3.5.1*

Para poder ejecutar este código sin errores se recomienda usar el mismo entorno que el de desarrollo, arriba señalado1.

5 CONCLUSIONES Y SIGUIENTES PASOS

En conclusión, este trabajo ha desarrollado la transformación de datos en información útil para los modelos de ciber riesgo OT. A través del uso de técnicas como la minería de reglas y procesos de tratamiento de datos estructurados, se ha logrado convertir grandes volúmenes de datos internos y externos en indicadores clave que fortalecen la capacidad predictiva y analítica de los modelos desarrollados.

El análisis exploratorio y los resultados generados demuestran la utilidad de estas transformaciones al identificar patrones significativos entre vulnerabilidades, vectores de ataque y posibles impactos. La integración de datos mapeados a estándares como MITRE ATT&CK añade un contexto táctico y estratégico que facilita la toma de decisiones en tiempo real y la priorización de medidas de mitigación.

Además, los gráficos obtenidos del análisis exploratorio destacan diferencias significativas entre conjuntos de datos, resaltando el potencial para ajustar los parámetros de la minería de reglas para optimizar aún más los resultados. Las métricas clave como soporte, confianza y *lift* han proporcionado oportunidades de mejora, permitiendo identificar relaciones críticas entre atributos y mejorar la relevancia de las reglas generadas.

[1] Es posible que a lo largo del desarrollo del Proyecto algunas versiones sean actualizadas.