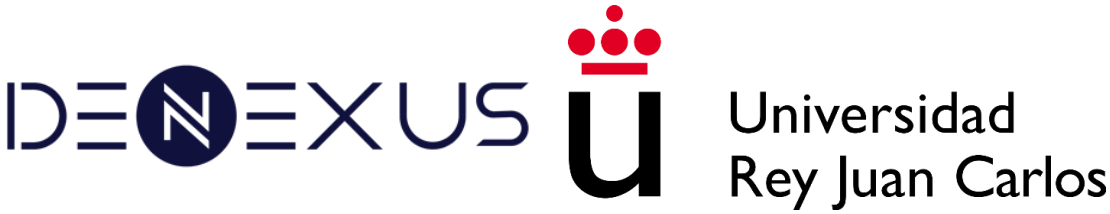
Convocatoria Colaboración Público-Privada CPP2021-009025

DICYME:

Dynamic Industrial CYberrisk Modelling based on Evidence

Iniciativa Conjunta de:



ENTREGABLE 2.3:

Prototipo final del módulo de medida/estimación de probabilidad e impacto

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 documento resume los aspectos más importantes de diseño e implementación del prototipo final del módulo de medida/estimación de probabilidad e impacto. Este entregable es fruto del trabajo realizado en las tareas 2.1, 2.2, 2.3 y 2.4 del Proyecto, dentro de la Actividad 2 “Módulos de medida/estimación de probabilidad e impacto”.

A lo largo de este documento se presentan diversos aspectos relacionados tanto con el prototipo final de estimación de probabilidad e impacto, como con la implementación de un nuevo modelo. En primer lugar, se detallan las mejoras introducidas en el modelo del prototipo existente, orientadas a resolver limitaciones previamente identificadas. Posteriormente, se introduce un nuevo modelo que permite combinar datos relativos a entidades víctimas de diversas fuentes con el objetivo de enriquecer, mejorar la calidad y completitud de los datos existentes en la aplicación.

En las próximas secciones del documento se explicarán los siguientes aspectos:

- Mejora de los modelos para el cálculo del atractivo.
- Resultados obtenidos de la evaluación de los nuevos modelos de atractivo.
- Descripción del nuevo modelo de combinación de víctimas VISCA.
- Visualización y descripción del flujo del nuevo modelo.
- Método para combinar los datos finales del nuevo modelo.
- Resultados obtenidos de la evaluación del nuevo modelo de combinación de víctimas VISCA.

2 ATR2ATK: ATRACTIVO DE UNA ORGANIZACIÓN A CIBERATAQUES

En los entregables E2.1 y E2.2 se introdujo el nuevo modelo de cuantificación del atractivo de una entidad a un atacante o actor de amenazas, dados diversos atributos que definen a la empresa por su ubicación, tamaño, etcétera. En el último prototipo, se han analizado los sesgos introducidos por la limitación del tamaño del conjunto de datos de Perfil de víctimas por diferentes motivos. Para solventarlo, se han tomado dos medidas diferentes, la primera relativa a los modelos de *Machine Learning* empleados para el cálculo de la métrica, y la segunda, posterior en el tiempo, para mejorar el conjunto de datos y poder mejorar la precisión del sistema.

2.1 Uso de *Random Forest*

Durante los análisis de los modelos aplicados, aún con el conjunto de datos inicial (incluye datos de víctimas de incidentes Diciembre 2023 a Diciembre 2024, así como de empresas similares a las afectadas, denominadas “no-incidentes”), se detectó una alta complejidad en la combinación de modelos de una manera que implicaba realizar

diversas fragmentaciones del conjunto de datos, cuyas dimensiones eran limitadas, sin implicar una mejoría en las métricas de rendimiento respecto a otras pruebas o alternativas. La explicabilidad tampoco era un punto a favor, puesto que a pesar de que la minería de reglas sí que aportaba combinaciones más sencillas de entender por los usuarios del sistema, la combinación de métricas posterior (cantidad de reglas cumplidas, suma de soporte, *lift* y confianza) que se usaba para definir el árbol de decisión complicada la interpretabilidad del modelo. Dadas todas las circunstancias mencionadas, se optó por simplificar la propuesta y pasar a usar un modelo basado en *Random Forest* que tiene en cuenta todas las variables definidas (Atractivo basal, Reputación en línea y Victimización). Esta decisión se fundamenta, además de lo anterior, en la mejoría de las métricas de evaluación respecto a la opción descrita en el segundo prototipo (minería de reglas y árbol de decisión).

Entre las pruebas realizadas para definir el modelo final, se incluyeron experimentos para mantener las tres componentes de Atractivo definidas desde el primer prototipo (Atractivo basal, Reputación en línea y Victimización) y combinarlas con una regresión logística como se venía haciendo previamente, así como introducir todas las variables directamente en el modelo de manera conjunta. Finalmente se optó por esta última opción, dado que reduce la complejidad del modelo y se obtienen unos parámetros de rendimiento similares (ligeramente mejores incluso) que en la opción de la combinación.

En el modelo final se ha realizado validación cruzada *K-Fold*, maximizando el *F1-score* para obtener el mejor balance entre precisión (*accuracy*) y sensibilidad (*recall*). Esta estrategia busca que el modelo sea capaz de identificar correctamente los incidentes reales (alto *recall*), al tiempo que minimiza las falsas alarmas (alto *accuracy*) (véase la *Tabla 1. Métricas de rendimiento del nuevo modelo ATR2ATK*). Esto se traduce en una mayor capacidad para detectar amenazas reales sin activar respuestas innecesarias, lo cual es crucial para una gestión eficiente de los recursos operativos. No obstante, se penaliza la confianza en los casos cuya predicción es negativa, que requerirán análisis adicionales por otras vías. La validación cruzada proporcionó la siguiente configuración del modelo:

- num.trees = 10
- mtry = 2
- splitrule = "gini"
- min.node.size = 1
- sample.fraction = 0.5

Tabla 1. Métricas de rendimiento del nuevo modelo ATR2ATK

Métrica	Entrenamiento	Test
Accuracy	86%	64%
Precision	87%	73%
Recall/Sensitivity	95%	80%
F1-score	91%	76%

No obstante, este cambio sigue reflejando demasiada afinidad al conjunto de datos, y supone problemas al generalizar el indicador a muestras no incluidas en los datos para poder integrar el indicador de forma satisfactoria en DeRISK.

En pruebas iniciales se intentó generar no-incidentes de forma sintética, haciendo uso de técnicas como SMOTE, para tratar de balancear el dataset, aunque los resultados no mostraron mejoría significativa ya que las categorías de los datos no diferían mucho de las reales. Por ello, se optó por desarrollar una propuesta más completa a la par que compleja, descrita en la ***Error! Reference source not found.***.

3 VISCA: VICTIM INSIGHT SYSTEM FOR CYBER ATTACKS

Este nuevo sistema, desarrollado en los últimos meses del proyecto e introducido en este último prototipo, busca mejorar aún más la recolección de datos de víctimas de ciber incidentes. La ejecución del modelo se inicia a partir de un conjunto de incidentes cibernéticos recopilados de las bases de datos ICSStrive y EuRepoC, lo que permite disponer de un dataset combinado de ciber incidentes. Este conjunto constituye un resultado valioso en sí mismo, ya que ofrece una base estructurada para la identificación de víctimas y posterior enriquecimiento con datos firmográficos, aunque en esta fase no se ha abordado aún la deduplicación de incidentes sino el proceso de identificación de las víctimas y sus características.

Una vez generado el conjunto de datos empleando las automatizaciones y prototipos de recogida de evidencias para incidentes desarrollados en la Actividad 1, se ha aplicado un filtrado para mantener únicamente aquellos que afectan a infraestructuras críticas o arquitecturas OT, que son las más relevantes para DICYME y DeNexus dado el propósito del proyecto y del propio DeRISK™. De esta manera, tras reconocer la víctima se obtienen datos firmográficos de ella, aprovechando de nuevo los prototipos de la Actividad 1, en esta ocasión los de extracción de información firmográfica. En este caso se consigue una mayor automatización en el reconocimiento de víctimas y una estandarización en la combinación de fuentes de datos, gracias a los modelos de lenguaje de gran tamaño (*Large Language Models*, LLMs) y al uso de nuevas fuentes de datos que posibilitan los LLMs, como motores de búsqueda.

Toda esta información puede ser empleada posteriormente en el entrenamiento y validación de ATR2ATK, pero también en el propio flujo de cuantificación del ciberriesgo. Si a partir de la descripción de un incidente se puede obtener la información de la víctima que usa el modelo de cuantificación (*Cyber Risk Quantification*, CRQ) detallado en el prototipo E2.2, el usuario del sistema sólo debe introducir la información privada adicional que no se puede obtener directamente (como los CVEs de la entidad) para poder ejecutar satisfactoriamente todo el flujo de cuantificación.

3.1 Flujo del modelo

Gracias a las capacidades de los LLM es posible procesar grandes volúmenes de datos, diseñar arquitecturas complejas, dotar al modelo de herramientas para la ejecución de tareas específicas, dar formato a la información y llevar a cabo múltiples operaciones adicionales que resultan fundamentales para los objetivos de este sistema. A

continuación, se describen las funciones y los procesos que lleva a cabo el sistema a partir del conjunto de datos de incidentes de ICSStrive y EuRepoC:

- **Obtención de información básica de la víctima:** a partir de un ciber incidente como fuente de entrada, el modelo es capaz de realizar su primera tarea, que es la extracción de la información básica que aparece en la descripción del incidente como puede ser el nombre de la entidad víctima, el país o la categoría de la industria. Adicionalmente, el modelo consultará la fuente de datos externa de Google con el que se realizarán consultas a partir del nombre de la víctima extraído (si existe) con el objetivo de obtener la URL de la página web de la entidad víctima. Estos datos básicos que se encuentran mencionados en la descripción del incidente se utilizan para posteriores procesos con los que obtener la información firmográfica.
- **Consulta de los datos firmográficos (*firmographics*):** con la información extraída previamente, el modelo consultará diversas fuentes de datos externas con los que se tratará de obtener 8 campos firmográficos de la entidad: nombre de la víctima, categoría de industria, año de fundación, número de empleados, país, ganancias, si cotiza en bolsa y códigos NAICS de industria.

  El modelo solicitará información sobre la entidad víctima a partir de cuatro fuentes principales: Google, DBpedia, BigPicture y RocketReach. DBpedia y BigPicture son fuentes abiertas, accesibles por cualquier usuario que quiera consulta información sobre empresas. Por su parte, RocketReach es una plataforma de pago que proporciona datos con un mayor nivel de confiabilidad que las demás al tratarse de un producto más privado. Google, en cambio, accede a información proveniente de diversas fuentes externas.
- **Combinación de los datos extraídos:** una vez extraída la información proveniente de diferentes fuentes de datos, el último paso que realiza el modelo es la combinación de los datos para generar un conjunto final que reúne la mayor cantidad de datos posibles. Para ello, a cada conjunto proveniente de las fuentes se le asigna un valor de confianza, basado en la cantidad de campos encontrados. Se selecciona la fuente con mayor confianza y, en el caso de que algún campo de *firmographics* esté ausente, se realiza una combinación con las otras fuentes siempre y cuando el valor de confianza supere cierto umbral.

Este proceso lo lleva a cabo una arquitectura multi-agente donde cada una de las tareas la realiza un LLM en concreto, siguiendo la arquitectura mostrada en la *Ilustración 1. Arquitectura del modelo de combinación de datos de víctimas*.

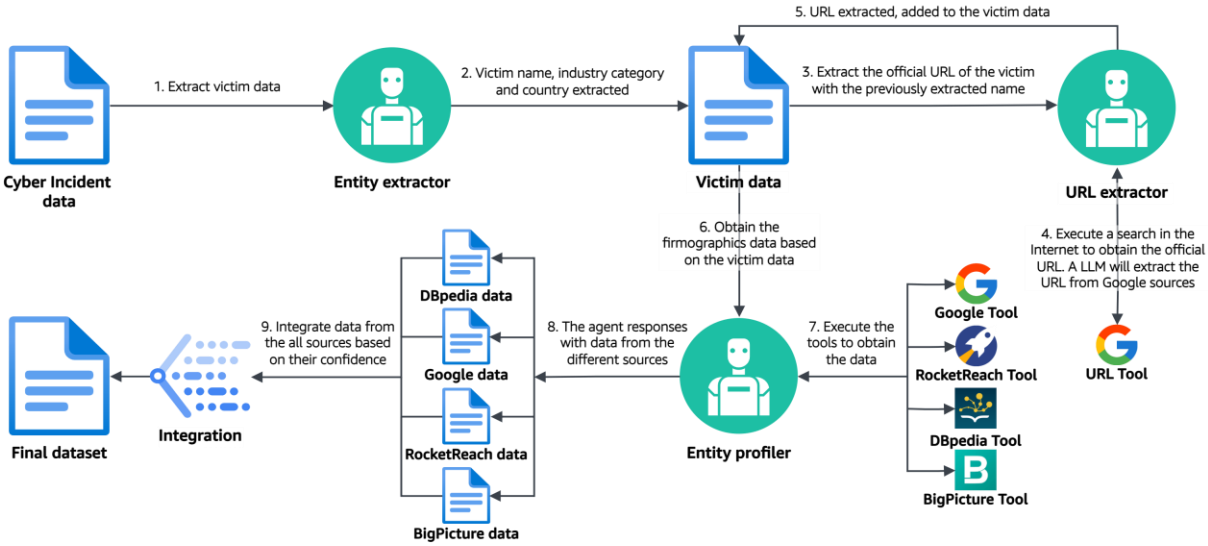


Ilustración 1. Arquitectura del modelo de combinación de datos de víctimas

La arquitectura sigue un flujo en el que la información que extraiga un agente será utilizada por el siguiente de ellos. Hay un total de 3 agentes, cada uno encargado de realizar una tarea concreta. El primer LLM llamado *Entity extractor* se encarga de realizar la obtención de la información básica de la víctima, donde a partir de un ciber incidente el modelo LLM será capaz de extraer dicha información si apareciese. El modelo *URL extractor* se encarga de complementar esta información mediante la consulta de la URL de la página web de la víctima. Finalmente, el agente *Entity profiler* recogerá la información de *firmographics* de varias fuentes externas y la función *Integration* realizará la combinación de los datos en un conjunto final.

La ejecución de este flujo completo permite extraer información valiosa de entidades que han sido víctimas de algún ciberataque, permitiendo así la creación de un conjunto de datos en el cual se encuentra toda esta información de diferentes víctimas. Este conjunto puede ser usado para DICYME para poder enriquecer la información de la aplicación, pudiéndose ser usado para, además, representar información o trabajar con dichos datos para diferentes objetivos.

3.2 Integración y combinación de los datos

En caso de que el nombre de la víctima esté presente en el ciber incidente, la arquitectura ejecutará el flujo completo. Al finalizar, un agente devolverá cuatro conjuntos de datos provenientes de cuatro fuentes distintas, con los campos firmográficos completos o incompletos. Para obtener un conjunto final que reúna la mayor cantidad de información posible, se ejecuta la función de integración del flujo. La función asigna un valor numérico de confianza a cada una de las fuentes, con un rango de entre 0 a 1. Este valor permite representar que tan confiable es una fuente en función del número de campos encontrados con valor, donde no mide la calidad de los datos.

La función seleccionará la fuente con mayor confianza y realizará combinación de resultados si se considera oportuno. La combinación se realiza si falta alguno de los campos, donde se comprueban si la confianza de las demás fuentes supera cierto umbral, donde en caso afirmativo se completan los campos faltantes si dichos campos contienen valor.

3.3 Resultados

Para comprobar la precisión y el comportamiento de la arquitectura a la hora de extraer datos tanto de un incidente como de otras fuentes, así como su capacidad para combinar los datos, la evaluación se ha dividido en dos partes: la primera analiza que tan bien extrae la víctima de un incidente y la segunda evalúa la calidad de los datos firmográficos obtenidos.

Para verificar el nombre de la víctima extraído por el LLM, se utilizan dos métodos de evaluación, uno basado en la distancia de *Levenshtein* y otro en el que un LLM actúa como juez. Las métricas utilizadas están basadas en el número de verdaderos y falsos positivos y negativos. En nuestros experimentos, se consideran las siguientes definiciones:

- TP (*true positives*): La víctima predicha coincide con la víctima real.
- TN (*true negative*): El modelo no predice ninguna víctima, pero en realidad hay una víctima.
- FP (*false positive*): El modelo predice una víctima, pero no existe ninguna.
- FN (*false negative*): La víctima predicha no coincide con la víctima real.

En el primer método basado en la distancia de *Levenshtein*, para considerar que una víctima predicha coincide con la real se calculan las distancia entre la víctima real y las víctimas predichas, ya que el modelo devuelve una lista de víctimas que aparecen en el incidente y pueden ser víctimas principales o secundarias. Con esta distancia calculada se establece un umbral, donde si la distancia pasa el umbral se considera FP y si la distancia no lo supera se considera TP.

El segundo método de evaluación se basa en la ejecución de un LLM que actúa como un juez, donde a partir de la introducción de un *prompt* preciso y extenso, el LLM es capaz de devolver una de las 4 métricas anteriores según considere. El LLM considera como un TP si alguna de la lista de las víctimas es muy similar a la víctima real, puesto que el nombre de la víctima puede ir acompañada de palabras referenciando subsidiarias o abreviaciones.

Los resultados de los dos métodos de evaluación se muestran en la *Tabla 2. Rendimiento de la extracción de los nombres de las víctimas*, donde para cada uno de ellos se han calculado el *Accuracy*, *Precision*, *Recall* y *F1 score*.

En los resultados previos se observa que el rendimiento medido por el evaluador basado en Levenshtein presenta una tasa de error más elevada que el evaluador basado en LLM. Esto se debe principalmente a que el nombre de la víctima suele aparecer con alguna abreviación, sucursal o subsidiaria, y aunque el nombre varíe ligeramente, sigue refiriéndose a la misma entidad real. En general, el agente es capaz de predecir con bastante acierto el nombre de la víctima de un incidente.

Tabla 2. Rendimiento de la extracción de los nombres de las víctimas

Dataset	Method	TP	TN	FP	FN	Accuracy	Precision	Recall	F1
ICSStrive	Levenshtein	154	0	46	0	0.77	0.77	1	0.87
ICSStrive	LLM-as-a-judge	193	0	7	0	0.96	0.96	1	0.98
EuRepoC	Levenshtein	130	0	70	0	0.65	0.65	1	0.79
EuRepoC	LLM-as-a-judge	189	0	9	2	0.95	0.95	0.99	0.97

Como se ha mencionado previamente, la segunda y última parte de estos experimentos consiste en la validación de los datos firmográficos extraídos de varias fuentes de datos. Los resultados han sido expresados mediante la proporción de los campos obtenidos con los campos totales de cada uno de los campos de *firmographics*, para cada una de las 4 fuentes. Los campos sobre los cuales se ha evaluado esta proporción han sido el nombre de la víctima, categoría, año de fundación, número de empleados, país, ganancias y código de NAICS (véase la ***Error! Reference source not found.***). Se han evaluado primeramente para la base de datos de ciber incidentes de ICSStrive y posteriormente para la base de datos de EuRepoC.

En los resultados obtenidos se puede apreciar que las fuentes de datos más fiables, es decir los que tienen de media la mayor confianza y retornan la mayoría de los campos, son las fuentes de Google y RocketReach. En cambio, DBpedia y BigPicture retornan de media pocos campos debido a que se realizan consultas directas a sus respectivas bases de datos, además de que estas fuentes no contienen todos los campos como pueden ser los códigos NAICS o el valor booleano que indica su cotiza en bolsa la víctima. Nótese que pese a que las proporciones de RocketReach son bastante altas su valor de confianza es más menor que si lo comparamos con Google. Esto se debe principalmente a que esta fuente, por cada consulta realizada, devuelve generalmente más de un resultado, por lo que al aplicar la fórmula de la confianza este valor quedará penalizado considerablemente.

Tabla 3. Rendimiento de la recolección de los datos de firmographics de todas las fuentes

Dataset	**Source**	**vic**	**cat**	**fou**	**siz**	**cou**	**rev**	**pub**	**nai**	**$\overline{conf}_s$**
ICSStrive	Google	0.94	0.86	0.69	0.87	0.84	0.74	0.91	0.64	0.81
	RocketReach	0.91	0.86	0.78	0.91	0.91	0.83	0.91	0.91	0.61
	DBpedia	0.37	0.27	0.27	0.27	0.06	0.15	0	0	0.20
	BigPicture	0.57	0.55	0.58	0.54	0.48	0	0	0	0.27
EuRepoC	Google	0.80	0.74	0.59	0.78	0.69	0.60	0.80	0.47	0.68
	RocketReach	0.93	0.90	0.75	0.93	0.89	0.75	0.93	0.93	0.64
	DBpedia	0.31	0.14	0.15	0.16	0.03	0.11	0	0	0.14
	BigPicture	0.54	0.53	0.57	0.51	0.47	0	0	0	0.22

Para finalizar, se evalúa el rendimiento final del sistema tras la integración de datos. Anteriormente, un mismo ciber incidente contenía un total de cuatro conjuntos de datos

distintos provenientes de diversas fuentes. Ahora, gracias al proceso de integración, se dispone de un único conjunto de datos unificado por incidente. Sobre este conjunto final se llevará a cabo la evaluación, cuyos resultados se presentan a continuación (véase la *Tabla 4. Rendimiento final de la recolección de los datos de firmographics*).

Tabla 4. Rendimiento final de la recolección de los datos de firmographics

Dataset	Source (n)	vic	cat	fou	siz	cou	rev	pub	nai	conf_total
ICSStrive	Google (59)	1.00	0.97	0.86	0.95	0.95	0.88	0.98	0.80	0.91
	RocketReach (34)	0.94	0.94	0.94	0.94	0.88	0.94	0.94	0.92	0.79
	DBpedia (5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95
	BigPicture (2)	1.00	0.50	1.00	1.00	1.00	0.00	0.50	0.00	0.59
	All (100)	0.98	0.95	0.90	0.95	0.95	0.87	0.96	0.84	0.87
EuRepoC	Google (44)	1.00	0.93	0.82	0.98	0.91	0.82	1.00	0.73	0.89
	RocketReach (44)	0.91	0.86	0.80	0.91	0.89	0.77	0.91	0.91	0.73
	DBpedia (5)	1.00	0.80	0.60	0.80	0.80	0.60	0.60	0.40	0.63
	BigPicture (7)	0.71	0.71	0.71	0.71	0.57	0.00	0.43	0.00	0.44
	All (100)	0.94	0.88	0.79	0.92	0.87	0.73	0.90	0.74	0.77

En los resultados presentados anteriormente, se ha evaluado un total de 100 incidentes correspondientes a las bases de datos ICSStrive y EuRepoC. A estos incidentes se les ha aplicado el proceso de integración descrito previamente, cuyos valores iniciales se muestran en la ***Error! Reference source not found.***, mientras que los resultados tras la integración se presentan en la *Tabla 4. Rendimiento final de la recolección de los datos de firmographics*. Tal como se indicó anteriormente, para cada incidente se selecciona la fuente con el mayor nivel de confianza entre las cuatro disponibles. En caso de que dicha fuente no contenga información para alguno de los campos firmográficos, se recurre a las demás fuentes, siempre y cuando superen un umbral mínimo de confianza previamente definido. En los resultados se aprecia que tanto en ICSStrive como EuRepoC las fuentes más seleccionadas son RocketReach y Google. Concretamente en ICSStrive de los 100 incidentes en 54 de ellos se ha seleccionado la fuente de datos de Google y 34 de RocketReach. En EuRepoC se han obtenido resultados similares con 44 selecciones de Google y 44 de RocketReach. Raramente se seleccionan las fuentes de datos de DBpedia o BigPicture al contar con una confianza más baja en la mayoría de los casos. Pese a ello, estas fuentes son útiles para realizar la combinación de datos.

4 OTROS MODELOS Y COMBINACIÓN

El modelo VISCA es empleado para la recolección de datos a mayor escala de la posible anteriormente, empleando no solo algoritmos de obtención de información, sino modelos LLMs y un modelo de combinación de datos propio basado en las fórmulas de confianza que permite mejorar el entrenamiento de ATR2ATK con datos actualizados y a mayor escala. El resto de los modelos expuestos en los prototipos previos E2.1 y E2.2 (CVE2ATK o CVE2TTPs, y THRACT), así como los conjuntos de datos generados (perfil de víctimas, indicadores de IDS y ciber incidentes) y las entradas del usuario (constantes

de ciber pólizas, CVEs presentes en la infraestructura y parámetros de las simulaciones) se encargan de nutrir al sistema de cuantificación del ciber riesgo (Cyber Risk Quantification, CRQ) detallado también en el prototipo E2.2.

Este modelo se encarga de combinar todas las entradas generadas para otorgar una puntuación de ciber riesgo personalizada para la entidad teniendo en cuenta el atractivo, los actores potenciales para la entidad en concreto, las tácticas y técnicas de MITRE ATT&CK dados los CVEs que sufre en sus sistemas, etc. Para ello, el sistema emplea distribuciones estadísticas y simulaciones de Monte Carlo, que permiten estimar de forma robusta la distribución de posibles escenarios de riesgo y cuantificar su impacto esperado.

Además, tal y como está diseñada la aplicación web de DICYME de visualización y soporte a la toma de decisiones, se permite al usuario la ejecución del modelo CRQ tanto para las empresas recopiladas en el conjunto de datos de perfil de víctimas como para una nueva que él defina a través de los parámetros necesarios, como el país, la industria o la cantidad de empleados, pudiendo también introducir el nombre de la misma para reflejarlo en el informe descargable que se ofrece con todos los detalles del proceso.

5 DATOS Y CÓDIGO

El código de este módulo se encuentra en el repositorio del proyecto: *deriskGroup / DICYME Project · GitLab*, dentro del directorio E.*2.3 Final prototype of measurement and estimation of probability and impact of a cyberincident*. Este contiene, a su vez, dos subdirectorios diferentes:

- Attractiveness: contiene el código del entrenamiento, K-Fold y un *script* para la predicción mediante el modelo basado en *Random Forest*. Asimismo, contiene el conjunto de datos empleado en su entrenamiento y validación, y un conjunto adicional y anonimizado para hacer funcionar el código de predicción. También está disponible el modelo entrenado tanto en formato RDA como RDS, y se emplea renv para la gestión de dependencias, asegurando que cualquier usuario pueda replicar los experimentos utilizando los mismos paquetes y versiones que los autores.
- Visca: en este directorio se incluye todo el código relativo al modelo VISCA, los *prompts* empleados por los agentes y modelos LLM, así como la documentación de todo el código tanto en formato PDF como en HTML.

Ambos subdirectorios contienen un fichero README.md con detalles sobre la ejecución y resultados de ambos modelos para facilitar la comprensión y replicación de los experimentos.

6 CONCLUSIONES Y SIGUIENTES PASOS

En esta actividad se ha desarrollado el prototipo final para la estimación de probabilidad e impacto, así como un nuevo modelo complementario cuyo objetivo principal es enriquecer los datos empleados en la aplicación DICYME.

A lo largo del documento se ha presentado una nueva versión del modelo principal, orientada a solventar las limitaciones identificadas en entregables previos. La sustitución de los modelos de *Machine Learning* anteriormente utilizados por un enfoque basado en *Random Forest*, junto con la incorporación de distintas técnicas previamente descritas, ha permitido mejorar el rendimiento general del sistema. No obstante, el modelo aún presenta margen de mejora, especialmente en su capacidad para generalizar frente a muestras provenientes de otros conjuntos de datos.

Adicionalmente, se ha introducido un modelo complementario diseñado para la extracción de datos relevantes sobre entidades víctimas de ciber incidentes. Esta información puede emplearse en tareas posteriores de análisis o visualización, tanto la mejora de ATR2ATK como en el modelo de CRQ, recuperando automáticamente muchas de las entradas de este. Aunque el modelo logra recuperar con éxito la información clave en la mayoría de los casos, el flujo de ejecución presenta aún áreas susceptibles de optimización. Una mejora futura podría centrarse en la evaluación de la calidad de los datos extraídos mediante métricas más específicas. Asimismo, uno de los principales desafíos identificados es el elevado tiempo de ejecución del proceso completo, el cual podría reducirse mediante la adopción de modelos alternativos o una reconfiguración de la arquitectura implementada, empleando mayores y mejores recursos de cómputo de los que disponen los equipos de la URJC y DeNexus actualmente.