

DICYME:

Dynamic Industrial CYberrisk Modelling based on Evidence

Iniciativa Conjunta de:



Universidad
Rey Juan Carlos

ENTREGABLE 4.3:

Segundo Prototipo del Sistema Completo e Informe de
Resultados de Pruebas

Coordinadores:

Romy R. Ravines (DeNexus Tech)

Isaac Martín de Diego (Universidad Rey Juan Carlos)

Alberto Fernández Isabel (Universidad Rey Juan Carlos)



Contenido

1	INTRODUCCIÓN.....	4
2	EL SISTEMA	4
2.1	Repositorio de datos.....	5
2.2	Indicadores DICYME	5
2.3	Simulador de ciber riesgo	6
2.4	Recomendador DICYME	7
2.5	Aplicación web	7
2.6	Informe de resultados.....	8
2.7	Agente IA	8
3	FUNCIONAMIENTO DEL SISTEMA	9
3.1	Data. Sistema ETL.....	9
3.2	Indicadores. Modelos.....	10
3.3	Calculadora de ciber riesgo	11
3.3.1	Frecuencia (Loss Event Frequency).....	11
3.3.2	Severidad (Loss Magnitude).....	12
3.4	Agente IA	13
4	INFORME DE RESULTADOS DE PRUEBA	15
5	COMENTARIOS FINALES.....	16

1 INTRODUCCIÓN

En este documento se describe como será el sistema completo propuesto por DICYME. Los indicadores y modelos diseñados en el proyecto serán utilizados para calcular el potencial impacto financiero que un evento de ciberseguridad puede causar en una compañía¹.

De esta manera se mostrará cómo funcionan las sugerencias derivadas de DICYME en el cálculo del ciber riesgo y el sistema de recomendación en una web app de acceso público.

2 EL SISTEMA

El Sistema DICYME se presenta en Ilustración 1 y sus principales componentes son:

1. Repositorio de datos
2. Indicadores DICYME
3. Simulador de ciber riesgo
4. Recomendador DICYME
5. Aplicación web
6. Generador de informes de resultados
7. Agente IA para describir datos e indicadores en la aplicación web.

Es importante resaltar que el sistema DICYME se encuentra aún en desarrollo y en este entregable se describe el estado actual del mismo. El contenido descrito a continuación está sujeto a modificaciones según que vaya concretizando el desarrollo de la aplicación DICYME. La aplicación web correspondiente, que integra estos datos y módulos de otras actividades del proyecto, está disponible a través de la siguiente dirección web <https://gondor.etsii.urjc.es:3866/>.

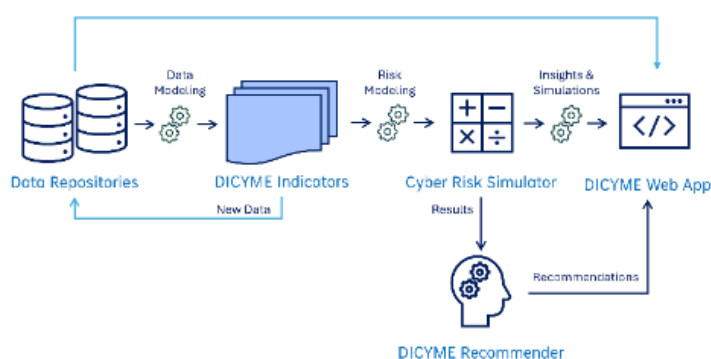


Ilustración 1. Sistema DICYME

¹ En la memoria técnica se indica que los indicadores se integrarán con DeRISK™; sin embargo, como se explica en el entregable E4.1, con la finalidad de disponer de un producto end-to-end (E2E) que sea publicado como web app de DICYME, y proporcione una solución completa de cálculo del impacto financiero del riesgo cibernético, se usará un sistema alternativo a DeRISK.

2.1 Repositorio de datos

Una de las principales contribuciones de DICYME es la **recolección automática de datos** con información sobre factores o *drivers* del ciber riesgo. Para ello, se han implementado procesos específicos de recolección adaptados a cada fuente, con los datos almacenados en una base de datos MongoDB en los servidores de la URJC.

Como descrito en anteriores documentos, durante el desarrollo del proyecto, la extracción y actualización el repositorio de datos se realizará de manera continua con los mismos tipos de datos. así como la generación de las relaciones necesarias para el desarrollo de otros sistemas como la aplicación web del proyecto, para la cual se han relacionado, por ejemplo, las tácticas y técnicas de MITRE entre sí siguiendo la matriz MITRE ATT&CK.

El código y la infraestructura relacionados con la recolección de datos están alojados en repositorios privados del proyecto en *GitHub* y *GitLab*, mientras que el almacenamiento y procesamiento se realiza en servidores de la URJC.

2.2 Indicadores DICYME

Otra contribución de DICYME es el diseño e implementación de indicadores para evaluar el ciber riesgo industrial. El módulo de indicadores integra subsistemas independientes, cada uno con objetivos específicos, metodologías propias y repositorios de datos, transformando información recolectada en métricas dinámicas que reflejan aspectos clave del perfil de riesgo de las organizaciones.

Los indicadores desarrollados son los ya mencionados en el E4.1: **ATR2ATK**, que mide el atractivo de una organización para ciberataques; **CVE2TTs**, que relaciona vulnerabilidades con técnicas explotables según MITRE ATT&CK; y **THRACT**, que evalúa la evolución de actores de amenazas frente a objetivos específicos. Sin embargo, se han actualizado y mejorado los modelos, incorporando nueva información obtenida del módulo de recolección automática descrito previamente, con el fin de incrementar la precisión y relevancia de los resultados.

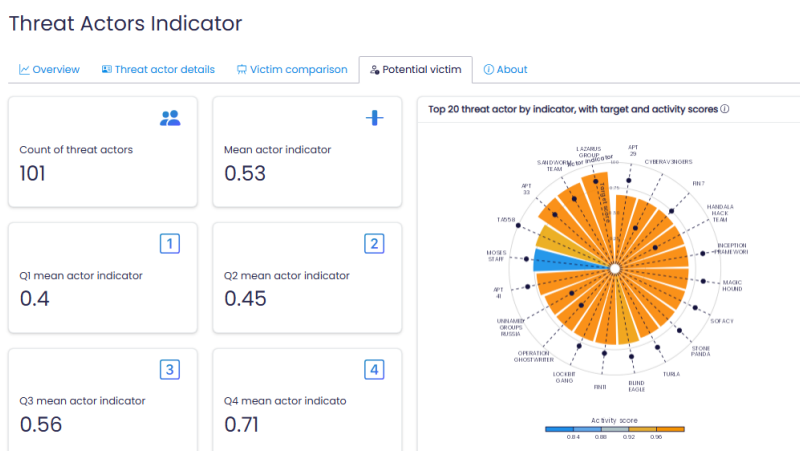


Ilustración 2. Indicador THRACT

Attractiveness indicator

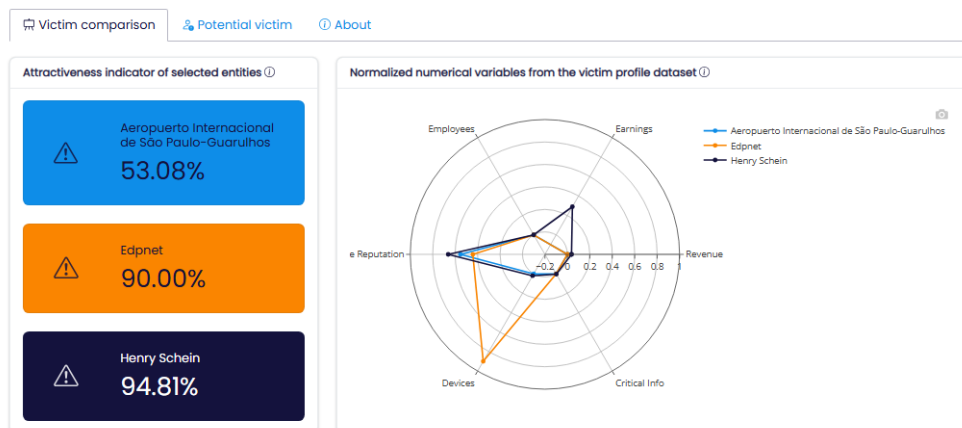


Ilustración 3. Indicador AT2ATK

2.3 Simulador de ciber riesgo

Este módulo constituye el motor de cálculo de ciber riesgo en DICYME. Basado en una adaptación de la metodología FAIR, simula la frecuencia e impacto de ciber incidentes utilizando distribuciones de probabilidad que, en esta versión, ya incorporan los resultados de algunos de los indicadores desarrollados en el proyecto. El funcionamiento detallado de este simulador con la primera integración de algunos indicadores se describe en la Sección 0, donde también se exploran posibles escenarios y recomendaciones para la mitigación del riesgo.

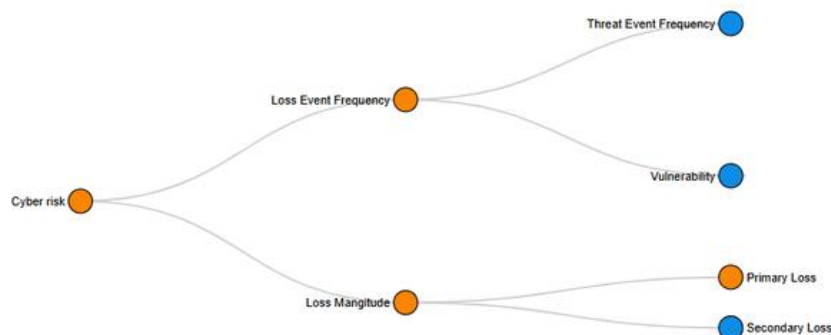


Ilustración 4. Taxonomía de ciber riesgo utilizada en DICYME

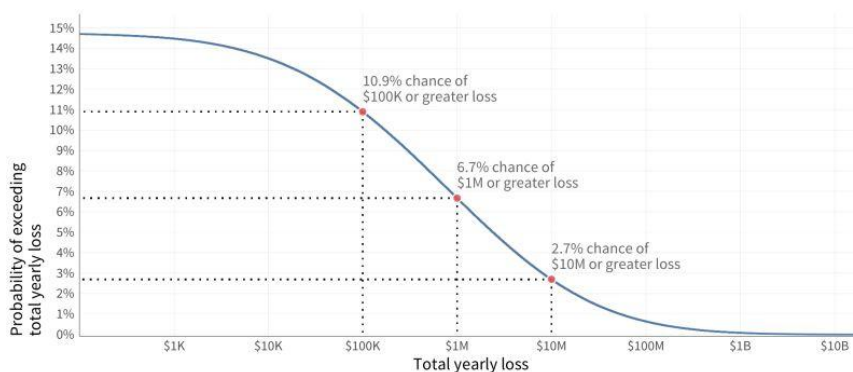


Ilustración 5. Curva de pérdidas (Loss Exceedance Curve)

2.4 Recomendador DICYME

El sistema recomendador propuesto se centra en la evaluación y gestión del ciberriesgo mediante un enfoque interactivo y visual. Actualmente, como se detalla en la sección 3.3 de este documento, el equipo de desarrollo está trabajando en el módulo CRQ, que calcula y agrega distintos parámetros que influyen en la estimación del ciberriesgo para obtener un resultado final del mismo. El recomendador inicial adoptará una metodología del tipo *what-if*, permitiendo al usuario ajustar diversos inputs a través de controles visuales, como deslizadores (*sliders*) u otros elementos interactivos. De este modo, se ofrecerá una experiencia dinámica donde el usuario podrá observar cómo la modificación de ciertos parámetros influye directamente en las simulaciones del riesgo cibernético, facilitando así el análisis de escenarios hipotéticos a la vez que analiza los valores calculados por el sistema.

A largo plazo, se planea enriquecer este sistema recomendador mediante la incorporación de escenarios prefijados en los que los parámetros varíen automáticamente en un rango fijo (por ejemplo, del 5% o 10% tanto al alza como a la baja). Esto permitiría evaluar la sensibilidad de las simulaciones ante pequeñas variaciones de los inputs, aportando conclusiones más precisas sobre el impacto de cada factor y facilitando la toma de decisiones fundamentadas para mitigar el ciberriesgo. Además, se valora también enriquecer el sistema haciendo uso de grandes modelos de lenguaje (LLMs), ya integrados en las pestañas de datos y modelos, para generar interpretaciones automáticas de los resultados, mejorando así la comprensión de los factores críticos que determinan el riesgo cibernético.

Si bien un algoritmo de optimización sería la solución más robusta para identificar configuraciones óptimas que minimicen el ciberriesgo, la complejidad de su desarrollo parece que va a exceder los plazos establecidos para el proyecto DICYME. No obstante, se mantiene como una opción de mejora a considerar en futuras fases de desarrollo, siempre que el alcance y los recursos disponibles lo permitan.

2.5 Aplicación web

La aplicación web es el nombre que usamos para el sistema de visualización y herramienta de recomendaciones que se cita en la memoria técnica de DICYME. Esta aplicación se está desarrollando y está alojada en los servidores de la URJC, accesible públicamente en <https://gondor.etsii.urjc.es:3866/>.

Además, esta aplicación incluye el trabajo desarrollado en esta actividad 4, dadas las limitaciones de integración directa con DeRISK ya descritos. La novedad de esta línea de trabajo es que dicha herramienta será independiente de DeRISK, incluyendo un motor de cálculo sencillo pero eficaz, basado en la metodología FAIR.

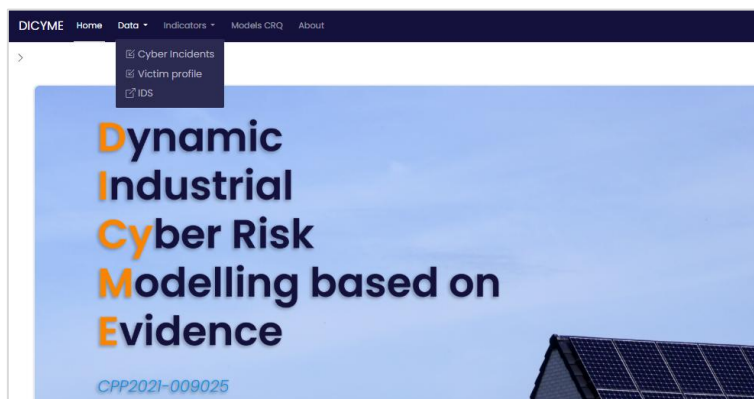


Ilustración 6. Página principal

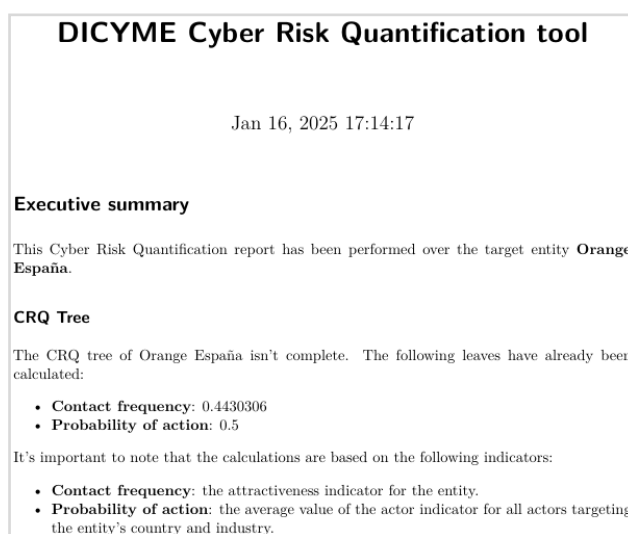


Ilustración 7. Informe de resultados

2.6 Informe de resultados

Con el fin de soportar la toma de decisiones, alineado tanto con la actividad 3 como la 4 del proyecto, se ha incluido la generación de un informe dinámico que incluye los resultados de la estimación del ciber riesgo. Por el momento, se genera un informe en formato PDF como el presentado en la Ilustración 7. De esta manera, se puede compartir información de la aplicación en un formato estandarizado y altamente compatible para reflejar un resultado concreto, de forma que los diferentes actores involucrados en la toma de decisiones e implementación de acciones puedan acceder a él de manera sencilla.

2.7 Agente IA

El sistema DICYME integra Modelos de Lenguaje de Gran Escala (*Large Language Models*, LLMs) para proporcionar descripciones automatizada sobre el contenido de la UI de la aplicación, que ayudan a los usuarios a interpretar datos complejos de ciberseguridad. La Ilustración 8. Agente IA en la interfaz (esquina inferior derecha) y la

Ilustración 9. Ejemplo de explicación automática muestran la apariencia del botón y disposición del análisis que realiza la IA en la interfaz de la aplicación de DICYME.

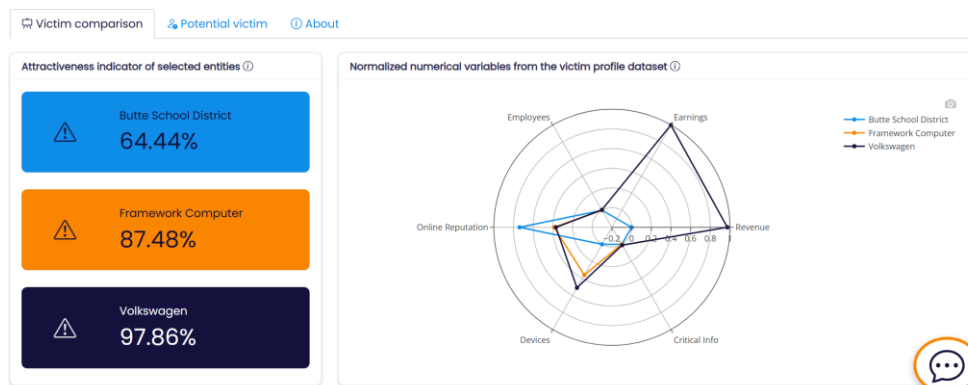


Ilustración 8. Agente IA en la interfaz (esquina inferior derecha)

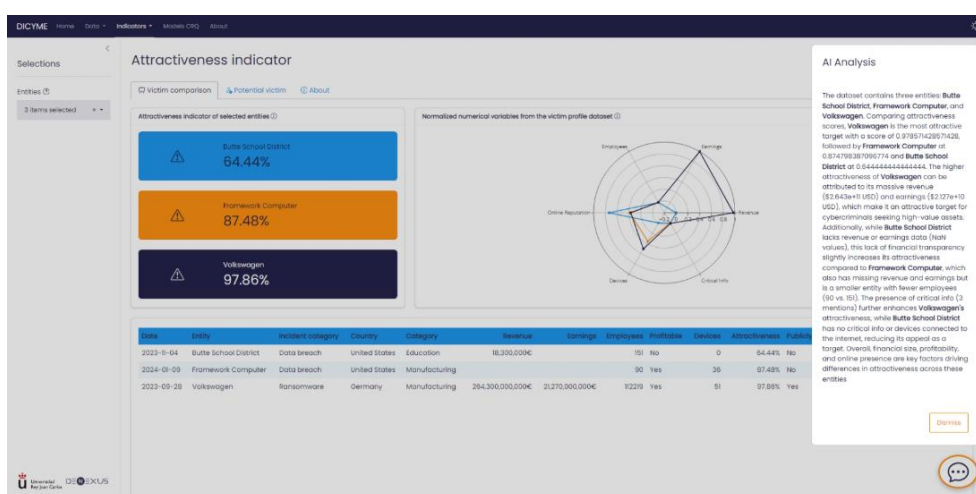


Ilustración 9. Ejemplo de explicación automática

3 FUNCIONAMIENTO DEL SISTEMA

En este prototipo se integran todos los resultados de datos y modelos en el sistema de cuantificación diseñado para este proyecto. De momento, únicamente se puede calcular el ciber riesgo para las compañías presentes en el conjunto de datos de perfil de víctimas desarrollado en la actividad 1. Sin embargo, como se está trabajando paralelamente en añadir la ejecución de los diferentes modelos al Sistema, este funcionamiento cambiará, pudiendo realizar una ejecución *end-to-end* para una compañía dadas sus características, ejecutando los modelos y combinándolos en la estimación del ciber riesgo.

3.1 Data. Sistema ETL

Los sistemas ETL se describen en los entregables E1.1, E1.2 y E1.3 del proyecto. A modo de ejemplo, la Ilustración 10 muestra el sistema implementado para la obtención y tratamiento de datos necesarios para el cálculo del indicador Atractivo.

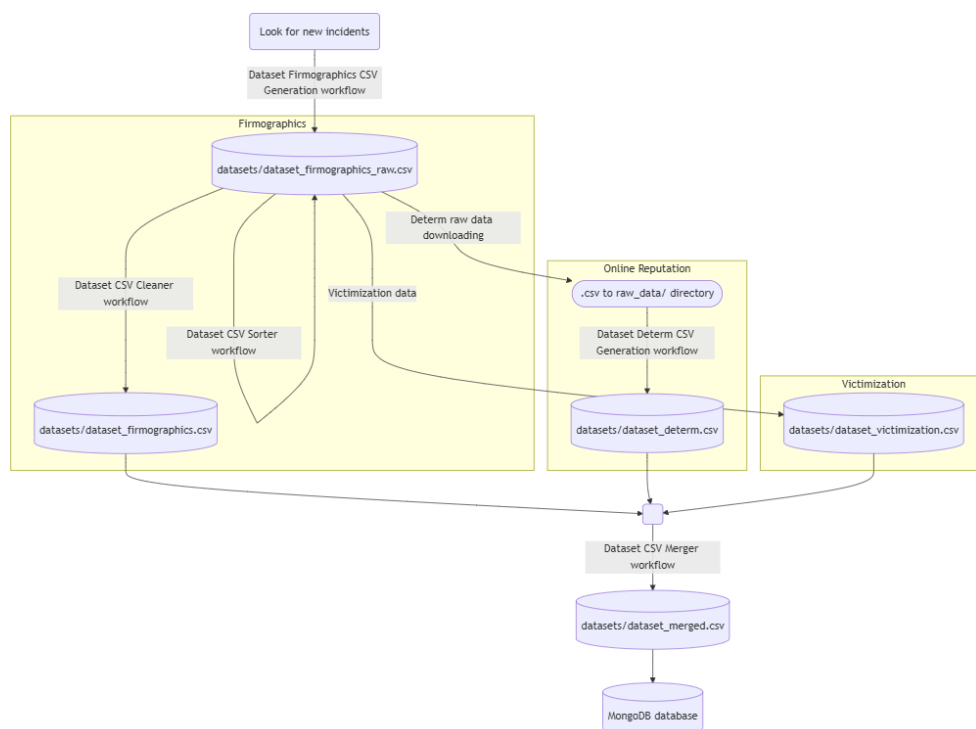


Ilustración 10. ETL de Atractivo

3.2 Indicadores. Modelos

Las transformaciones aplicadas a los datos recolectados se describen en el E1.4 del proyecto. Por ejemplo, en la Ilustración 11 se muestra la arquitectura del modelo utilizado para el mapeo entre las vulnerabilidades (CVE) and las técnicas de las matrices MITRE ATT&CK.

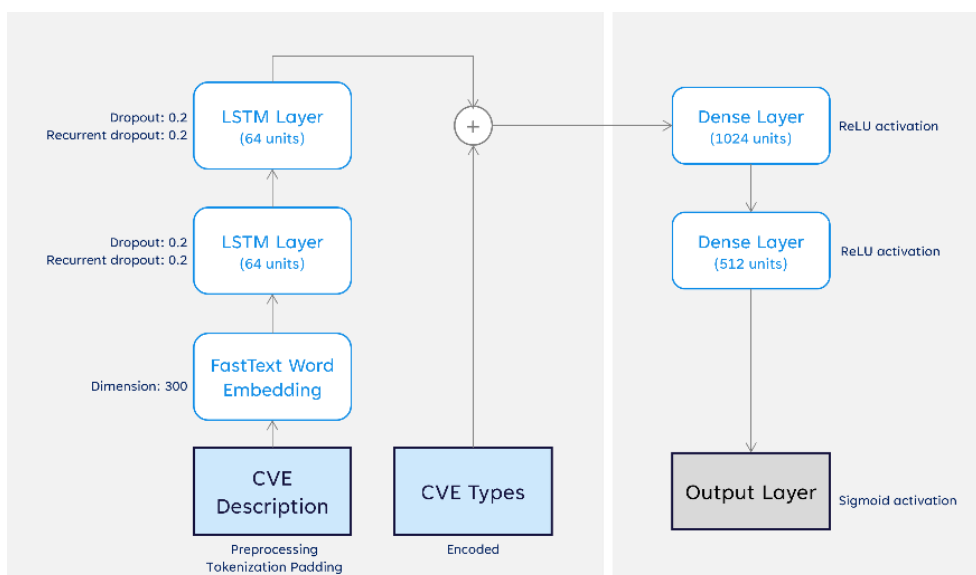


Ilustración 11. Arquitectura del modelo CVE2ATK

3.3 Calculadora de ciber riesgo

Propondremos un marco sólido y fácil de entender para cuantificar el riesgo cibernético en términos financieros. Este marco integra todos los datos e indicadores descritos en el proyecto, combinando datos dinámicos con perspectivas de expertos, ya que parte de la información podría no estar disponible públicamente.

DICYME utiliza algunos elementos de la taxonomía FAIR (como se muestra en la Ilustración 4), adaptados para estimar el riesgo cibernético al que se enfrenta una organización en un año. También utiliza la simulación de Monte Carlo para calcular la Pérdida Anual Esperada.

3.3.1 Frecuencia (Loss Event Frequency)

- (A) **Baseline.** Es el promedio de intentos de ciberataques que una organización enfrenta al año. Se obtiene a partir de informes de terceros por industria y región.

$$\text{base}_k(y-1, i)$$

- (B) **Conteo de incidentes** en los últimos 3 años en la industria y la región. Este recuento es una estimación mínima del número de intentos. Calculamos la tasa de crecimiento anual utilizando esta serie temporal corta.

$$\text{rate}_k = f(N_{y-3,i}, N_{y-2,i}, N_{y-1,i})$$

- (C) **Atractivo.** La probabilidad de VERDADERO, es decir, la probabilidad de ser víctima de un ciberataque es un número entre 0 y 1.

$$\text{attrac}_k$$

- (D) **Índice de Actores de Amenaza** de los últimos 3 años. Es la fuerza de una amenaza para comprometer a la organización. El índice muestra una tendencia y se utiliza un promedio ponderado. Su valor oscila entre 0 y 1.

$$\text{threat}_k = f(a_{y-3,i}, a_{y-2,i}, a_{y-1,i})$$

- (E) **Técnica de MITRE ATT&CK** que un CVE permite a los atacantes explotar. Lista de técnicas a las que la organización está más expuesta según su lista de vulnerabilidades. Utilizamos un índice ponderado de exposición. Cuanto más cercano al impacto esté, mayor será el peso de la técnica.

$$\text{expose}_k = \sum_h w_h \times (\text{cvss} * \text{epss}) / 10$$

- (F) **Perfil de Seguridad.** La probabilidad de que una amenaza comprometa a una organización depende también de su capacidad para resistirla (detectarla y responder a ella). Contamos con tres perfiles basados en datos de Denexus

$$\text{security}_k = sc_k$$

- (F1) **Threat Event Frequency**. Simular el número de intentos de ataque a partir de una distribución de Poisson con media N_k .

$$N_k = \text{base}_k \times \text{rate}_k \times \text{attrac}_k$$

- (F2) **Susceptibility**. Simular la probabilidad de éxito a partir de una distribución Beta con media p_k .

$$p_k = \frac{\text{threat}_k}{\text{expose}_k \times \text{security}_k}$$

- (G) *Loss Event Frequency*

$$\text{Frequency}_k = N_k \times p_k$$

- (H) **Exposición a técnicas de impacto según la lista CVE**. No todas las técnicas pueden causar el mismo impacto o efecto. El tipo de pérdida primaria causada por una técnica MITRE se deriva de eventos pasados.

3.3.2 Severidad (Loss Magnitude)

- (Ia) **Pérdida primaria - Tiempo de inactividad**. El sistema simula la duración de la interrupción; el resultado debe multiplicarse por el coste de una hora de inactividad (reportado por la organización).
- (Ib) **Pérdida primaria - Daños a equipos**. El sistema simula la magnitud de los daños. El resultado debe multiplicarse por el coste del equipo o la infraestructura en riesgo (reportado por la organización).
- (Ic) **Pérdida primaria - Extorsión**. El sistema simula la proporción de ingresos que se retiene como extorsión. Suele ser un valor pequeño, no superior al 1% de los ingresos.
- (Id) **Pérdida primaria - Daños humanos**. El sistema simula la gravedad de los daños. Debe multiplicarse por el coste estadístico de una constante de vida.
- (Ja) **Pérdida Secundaria - Investigación Forense**. El sistema simula las horas dedicadas a esa actividad. El resultado debe multiplicarse por el coste de una hora de investigación forense.
- (Jb) **Pérdida Secundaria - Daño Reputacional**. El sistema simula la magnitud del problema. El resultado debe multiplicarse por los ingresos como indicador de la pérdida de cuota de mercado.

- (Jc) **Pérdida Secundaria – Sanciones.** El sistema simula la magnitud del problema. El resultado debe multiplicarse por los ingresos, pero se debe aplicar un valor máximo (p. ej., 10 millones).
- (K) **Magnitud de la pérdida - Árbol por técnica de impacto.**

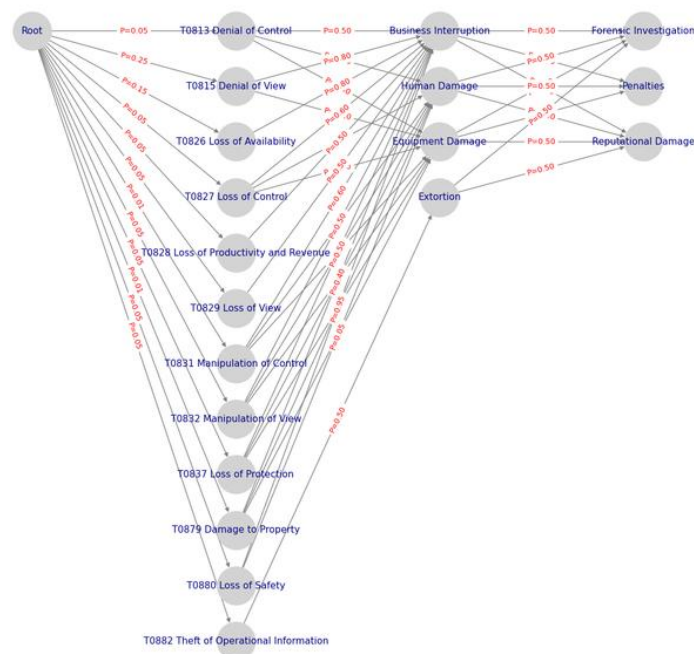


Ilustración 12. Modelo de severidad DICYME

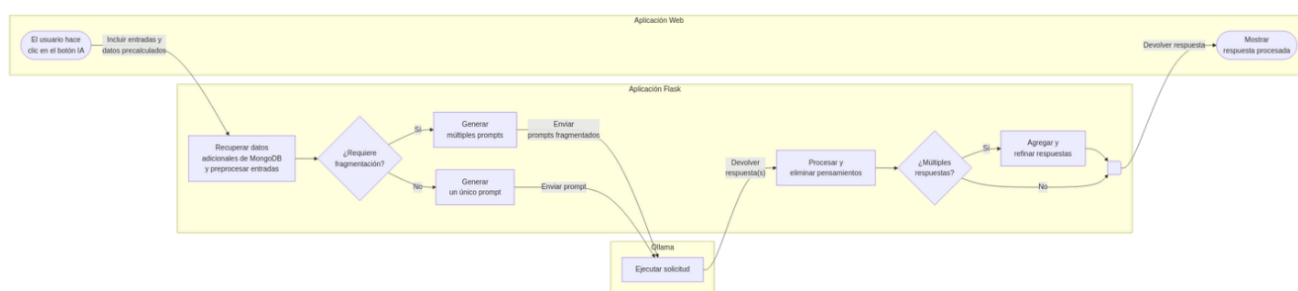
3.4 Agente IA

Para garantizar eficiencia, seguridad y fiabilidad, DICYME ha desarrollado una infraestructura local robusta que permite procesar y generar respuestas sin depender de servicios externos, es decir, DICYME emplea una infraestructura auto alojada dedicada. La configuración consta de:

- **Ejecución en servidor Ollama:** El modelo seleccionado se ejecuta localmente en un servidor dedicado, asegurando un control total sobre el procesamiento y eliminando la dependencia de APIs externas o servicios en la nube.
- **Entorno de contenedores:** Un contenedor **Docker** aloja una aplicación en **Flask (Python)**, desplegada a través de un **servidor WSGI**, lo que garantiza escalabilidad y mantenimiento.
- **Integración con base de datos local:** Una instancia de **MongoDB** almacena datos relevantes del proyecto, permitiendo al sistema recuperar información contextual de manera dinámica en cada solicitud.

Cuando un usuario realiza una solicitud, el sistema sigue un proceso estructurado en múltiples pasos para garantizar el rendimiento óptimo del modelo y la calidad de las respuestas:

1. **Recepción de la solicitud:** La API recopila las entradas del usuario y cualquier dato precomputado relevante de la aplicación web de DICYME.
2. **Enriquecimiento de la consulta:** La API consulta la base de datos MongoDB para obtener información adicional necesaria, asegurando que el modelo tenga un contexto completo.
3. **Preprocesamiento de datos:** Los datos recuperados se preprocesan junto con las entradas recibidas de la aplicación para garantizar que la respuesta refleje adecuadamente lo que el usuario está viendo, aplicando cualquier filtro seleccionado.
4. **Construcción del *prompt*:** Se genera un *prompt* adaptado, incorporando tanto los datos del usuario como los datos preprocesados, en un formato estructurado y optimizado para el procesamiento del modelo: Markdown.
5. **Manejo de consultas fragmentadas (si aplica):** Para solicitudes complejas que requieren análisis en profundidad, el sistema **divide la consulta en partes más pequeñas**, permitiendo que el modelo las procese de forma independiente antes de combinar los resultados. Esta técnica mejora la calidad del razonamiento al reducir la sobrecarga de contexto y mejorar el enfoque en subconjuntos de datos.
6. **Ejecución del modelo:** El *prompt* construido se envía al modelo alojado localmente, garantizando una inferencia de alto rendimiento en un entorno controlado.
7. **Agregación y refinamiento de la respuesta:** Si la solicitud fue fragmentada, las respuestas individuales se sintetizan en una respuesta final coherente.
8. **Devolución de la respuesta:** La salida final refinada se envía de vuelta a la aplicación web para su visualización por el usuario.



Un ejemplo de los *prompts* empleados para obtener el análisis del LLM se reproduce a continuación, utilizado para obtener información del repositorio de ciber incidentes de DICYME. Se puede comprobar la información incluida en el contexto, como las diferentes bases de datos que se están analizando, la unidad temporal de agrupación de los datos, así como instrucciones para el análisis y formato de la respuesta.

```

### **Cyber Incidents Temporal Trends Analysis**
You are an expert cybersecurity data analyst. Your task is to analyze the temporal trends of cyber-incidents based on the provided dataset.

### **Dataset Structure:**
The dataset is a table that contains information of a time series, containing the counts of the total number of recorded cyber-incidents for each cyber incidents repositories selected. Columns include:
- date: The date on which the cyber incidents were collected.
- Cyber incidents repository Count: The total number of cyber-incidents recorded for that period for each Cyber incidents repository.

### **Additional Data:
The names of the cyber incidents repositories are: ['ICSSTRIVE', 'TISAFE']
The time unit for grouping the data is: year

### **Analysis Instructions:
1. Cyber-Incident Trends Over Time:
    - Identify significant increases or decreases in the number of cyber-incidents.
    - Highlight spikes, drops, and periods of stability.
    - Compare trends across different time granularities (daily, monthly, quarterly, yearly).
2. Comparison Across Cyber incidents repositories:
    - Analyze differences in reported incidents among various cyber incidents repositories.
    - Identify which cyber incidents repositories report the highest and lowest number of incidents.
3. Upload Delay Analysis:
    - Detect possible delays in incident reporting by observing sudden peaks or missing data.
    - Compare reporting patterns among cyber incidents repositories.
4. Temporal Patterns:
    - Examine the seasonality of incidents: are there recurring peaks in certain months or quarters?
    - Analyze whether incident counts follow a long-term upward or downward trend.

### **Dataset in DataFrame format:
    date  ICSSTRIVE Count  TISAFE Count
0  2023             279           84
1  2024             55           41

### **Response Guidelines:
- The response must be a single concise paragraph (max 300 words).
- IMPORTANT. Group the data by each cyber incidents repository (Columns of the dataset).
- IMPORTANT. Analyze the cyber incidents by the date correctly.
- FOCUS on the date of the cyber incidents to analyze the time series.
- Analyze ALL cyber incidents repositories and highlight key patterns.
- DO NOT speculate on causes or motivations—focus on observable trends.
- Include specific numeric examples from the dataset.

```

4 INFORME DE RESULTADOS DE PRUEBA

Se ha implementado una serie de pruebas unitarias en la aplicación web de DICYME con el objetivo de validar el código, asegurar su correcto funcionamiento y verificar que las funciones devuelven el tipo de datos esperado. Estas pruebas cubren los siguientes aspectos clave:

- Verificación de tipos de datos: Se verifica que las funciones devuelven el tipo de datos correctamente como puede ser una lista, un DataFrame, etc.
- Verificación de clase: Se valida que ciertas funciones devuelvan un objeto de una clase en específico como puede ser un gráfico.
- Manejo de avisos: Se comprueba que, en determinados casos, las funciones generen los avisos deseados.

5 COMENTARIOS FINALES

En este documento se describe el estado actual de la aplicación web de DICYME para la estimación del ciber riesgo de una compañía. Concretamente, se describen todos los componentes del sistema, desde la recolegida de datos, la generación de informes en formato *pdf* y el uso de un agente IA que explica lo que el usuario está visualizando en la UI de la aplicación web.