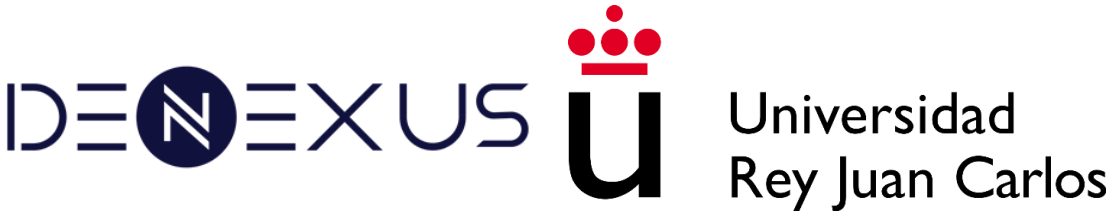
Convocatoria Colaboración Público-Privada CPP2021-009025

DICYME:

Dynamic Industrial CYberrisk Modelling based on Evidence

Iniciativa Conjunta de:



ENTREGABLE 2.2:
Segundo prototipo 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 del segundo prototipo de 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".

En el resto de secciones de este documento se abordan los siguientes temas:

1. Los nuevos datos empleados para la mejora de los modelos, empleando los resultados de la Actividad 1 "Sistema de recogida de evidencias".
2. Las evoluciones introducidas en los modelos desarrollados en el primer prototipo, correspondiente al entregable E2.1.
3. Los resultados de los nuevos modelos.
4. Otras peculiaridades de los modelos que permiten su mejora continua.
5. Qué datos y código se encuentran disponibles en este entregable (deriskGroup / DICyME Project · GitLab) y cómo pueden emplearse.

2 ATR2ATK: Atractivo de una organización a ciberataques

En el entregable E2.1 se introdujo un nuevo modelo de cuantificación del atractivo de una entidad: la posesión de atributos o características que causan el interés de la entidad para un atacante. El objetivo de este modelo es cuantificar la probabilidad de que una entidad sea víctima de un ciberataque. En este primer prototipo se emplearon atributos que definen a la empresa por su ubicación, tamaño, etcétera, lo que se denominó con datos firmográficos, junto a una combinación de dos métodos: minería de reglas con *FP-Growth* y un árbol de decisión.

2.1 Novedades del segundo prototipo

En este segundo prototipo, se han añadido dos nuevas categorías de atributos, que aportan mayor dinamismo, puesto que la información firmográfica es eminentemente estática por naturaleza: las entidades apenas cambian de ubicación, tampoco varían enormemente su tamaño con frecuencia, la categoría de industria, etc. Estas nuevas categorías son:

- **Reputación en línea**: se define como el nivel de aceptación y reconocimiento que una entidad tiene entre las personas en la red, medido a partir de lo que los usuarios, clientes o empleados escriben, comunican y comparten en Internet. Este indicador dinámico es clave para comprender cómo las interacciones y opiniones en redes y medios sociales impactan en el atractivo de la entidad, ya sea de manera positiva o negativa, incluso para posibles atacantes.

  La medición de la reputación en línea se realiza considerando la interacción (*engagement*) y el alcance (*reach*) tanto del contenido controlado por la entidad como de las menciones externas, utilizando un enfoque cuantitativo y dinámico que refleja los cambios a lo largo del tiempo. Esta fórmula, que asigna valores entre -0.5 (mala reputación) y 1 (buena reputación) se detallará en la documentación de los modelos y métricas recogida en el entregable E2.4.

Cabe mencionar que este indicador es el primero de los desarrollados en DICYME que se está utilizando parcialmente en el módulo NoA de DeRISK.

- **Victimización**: se refiere a la probabilidad de que una entidad sea considerada un objetivo atractivo por parte de adversarios. Esto puede deberse a su presencia en foros *underground* o sitios de la *dark web*, la exposición de información sensible debido a brechas de datos, o la percepción de facilidad para atacar su infraestructura.

  La medición se basa en la visibilidad de información de riesgo y la percepción de éxito, evaluadas a través de datos sobre brechas de información y dispositivos conectados a Internet, asignando un valor que varía entre 0 (baja victimización) y 2 (alta victimización) en función de si ambas variables tienen valores distintos de 0 o no.

Estas nuevas componentes aprovechan las nuevas evidencias descritas en el entregable E1.3 y recogidas en la Tabla 1. Nuevos indicadores incluidos en la base de datos DICYME (véase el entregable E1.3).

Nombre del atributo	Expresión para su cálculo
Reputación online	Fórmula basada en el estado del arte, detallada en el entregable E2.4.
Información crítica	Cantidad de apariciones del nombre de la entidad en *leaks* de *ransomware* diferentes.
Dispositivos visibles	Cantidad de dispositivos visibles en Internet asociados al nombre de la entidad.

Tabla 1. Nuevos indicadores incluidos en la base de datos DICYME (véase el entregable E1.3)

De esta manera, el atractivo pasa a estar formado por tres categorías, firmográficos, reputación en línea y victimización. Puesto que para la primera métrica se dispone de una probabilidad entre 0 y 1, fruto de la predicción con el árbol de decisión, y las dos nuevas ya disponen de un valor acotado, la combinación o *stacking* de las tres métricas se realiza empleando una regresión logística. Véanse la Ilustración 1. Concepto del modelo de atractivo y la Ilustración 2. Metodología a alto nivel del modelo de atractivo para más detalles sobre la combinación de las diferentes categorías y el funcionamiento del modelo.

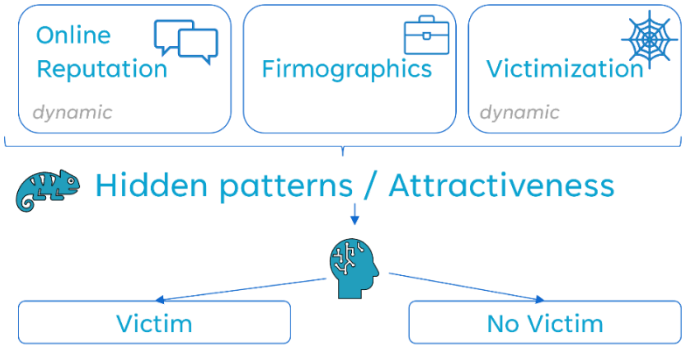


Ilustración 1. Concepto del modelo de atractivo

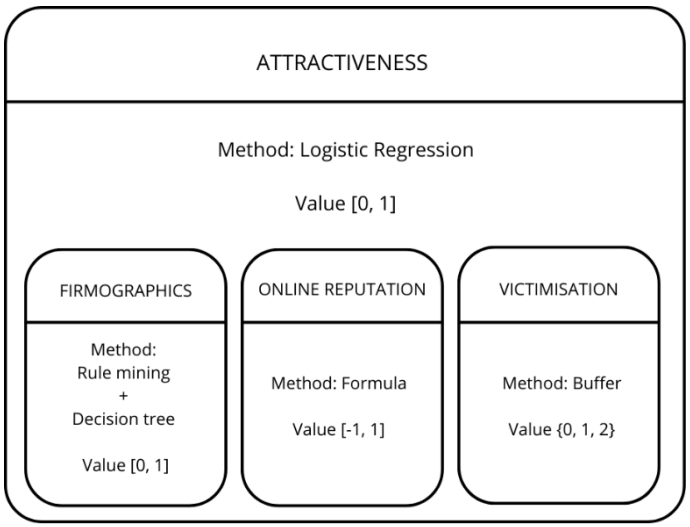


Ilustración 2. Metodología a alto nivel del modelo de atractivo

2.2 Resultados

Dados los nuevos aportes del sistema, pero aún conscientes de sus limitaciones y oportunidades de mejora, en las cuales se sigue trabajando para el futuro tercer prototipo, el modelo ha sido enviado como publicación científica a una revista y a un congreso internacional, encontrándose aún en proceso de revisión.

Es importante destacar que en este segundo prototipo se han resuelto diversos problemas en el código del primer prototipo, tanto en el modelo como en las transformaciones de los datos, por lo que los resultados de ambos no son directamente comparables. Los principales estadísticos para la evaluación del modelo se recogen en la Tabla 2. Resultados del modelo de atractivo con las nuevas categorías de datos.

Métrica	Datos de entrenamiento	Datos de prueba
Accuracy	0.75	0.68
Sensitivity	0.97	0.91
Specificity	0.18	0.11
F1-score	0.85	0.80

Tabla 2. Resultados del modelo de atractivo con las nuevas categorías de datos

A modo descriptivo y comparativo respecto al primer prototipo, se incluyen en la Tabla 3. Primeras 10 reglas aprendidas según su soporte las principales reglas, según su soporte, obtenidas mediante *FP-Growth*, sabiendo que el conjunto de entrenamiento de la minería de reglas contiene 388 entidades.

Regla	Support	Confidence	Lift	Count
{} => {Earnings=NA}	0.94	0.94	1.00	366
{Publicly_traded=FALSE} => {Earnings=NA}	0.87	1.00	1.06	338
{Earnings=NA} => {Publicly_traded=FALSE}	0.87	0.92	1.06	338
{} => {Publicly_traded=FALSE}	0.87	0.87	1.00	338
{} => {Employees=2}	0.62	0.62	1.00	240
{Employees=2} => {Earnings=NA}	0.60	0.98	1.03	234
{Earnings=NA} => {Employees=2}	0.60	0.64	1.03	234
{Employees=2, Publicly_traded=FALSE} => {Earnings=NA}	0.57	1.00	1.06	220
{Earnings=NA, Employees=2} => {Publicly_traded=FALSE}	0.57	0.94	1.08	220
{Earnings=NA, Publicly_traded=FALSE} => {Employees=2}	0.57	0.65	1.05	220

Tabla 3. Primeras 10 reglas aprendidas según su soporte

3 Cuantificación del ciber riesgo

En la memoria técnica de DICYME se planteó que los indicadores propuestos se integrarían en el producto DeRISK y que todas las pruebas y resultados finales se harían usando DeRISK como sistema de cuantificación del riesgo ciber de DICYME. Ese planteamiento no puede ser llevado a cabo porque las integraciones con DeRISK sólo las pueden ejecutar los ingenieros de *backend* de DeNexus, que no hacen parte del Proyecto y están 100% enfocados al plan de desarrollo propio de DeNexus. Por ese motivo, los indicadores se han probado en el entorno de desarrollo e integrado parcialmente en DeRISK.

- Los indicadores calculados a partir de datos internos de las organizaciones (*inside-out data*) han sido desarrollados por DeNexus y se encuentran integrados en DeRISK. Al tratarse de datos confidenciales de instalaciones reales, el equipo DICYME ha trabajado con una pequeña muestra de datos anonimizados y modificados, suficientes para plantear los indicadores.
- El Threat Actor Index (THRACT) se ha integrado con DeNexus pero usa dos fuentes de datos en lugar de sólo 1 (La fuente datos usada en DICYME es pública. La 2da base de datos usada por DeNexus no puede ser cedida a DICYME)
- El Atractiveness Index (ATR2ATK) se ha probado en desarrollo y está será integrado a medio plazo. Esta integración supondrá un aumento en la calidad de como DeNexus representa el atractivo de una empresa en el modelo, pero también un aumento factura en consumo de datos de otros proveedores.

- La primera versión del CVE2TTs está 100% integrada en DeRISK. Los nuevos entrenamientos de los algoritmos y mapeos se realizan directamente en DeRISK.

Para evitar bloquear el desarrollo de DICYME, el equipo ha acordado trabajar con un motor de cálculo alternativo a DeRISK, desarrollado ad-hoc para el Proyecto. Se trata de un *Cyber Risk Calculator* que adopta la taxonomía FAIR pero usa algoritmos propios para el cálculo de los componentes, de amanera análogo al modelo conceptual de DeRISK.

A continuación, se describe el *Cyber Risk Calculator* de DICYME.

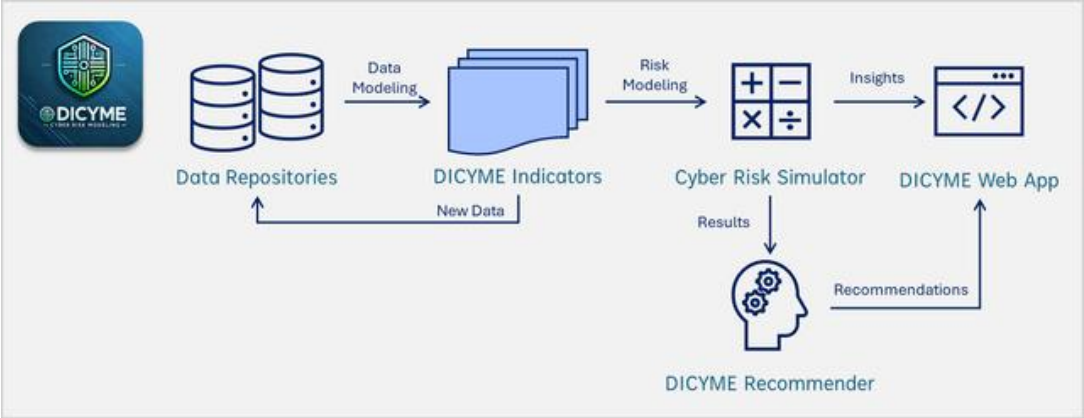


Ilustración 3

3.1 Enfoque FAIR

La **cuantificación del riesgo cibernético** (CRQ) es el proceso de evaluación y cálculo del impacto financiero potencial de las amenazas cibernéticas en una organización. Traduce los riesgos de ciberseguridad en términos relevantes para el negocio, lo que permite a las organizaciones comprender el valor monetario de su exposición al riesgo. En otras palabras, la cuantificación del riesgo cibernético (CRQ) calcula la exposición al riesgo y su impacto financiero potencial en una organización en términos relevantes para el negocio, lo que proporciona una forma para que las organizaciones impulsen la alineación entre la estrategia de seguridad y los objetivos comerciales.

FAIR (*Factor Analysis of Information Risk*) es un marco de análisis de riesgos cuantitativo que ayuda a las organizaciones a evaluar y cuantificar los riesgos cibernéticos en términos monetarios. Al analizar escenarios de riesgo específicos, el modelo FAIR estima la exposición potencial a pérdidas, lo que permite a las empresas comprender dónde son más vulnerables a los ciberataques. Reconocido como un estándar internacional, FAIR ayuda a las organizaciones a tomar decisiones informadas sobre la gestión de riesgos y las estrategias de mitigación.

La metodología FAIR cuantifica el riesgo de ciberseguridad desglosándolo en factores individuales mediante análisis estadísticos y probabilidades. Evalúa el riesgo a través de escenarios cuidadosamente definidos mediante la evaluación de la frecuencia probable de eventos de pérdida (Frecuencia de eventos de pérdida) y el impacto potencial de esas pérdidas (Magnitud de pérdida). Al combinar estos componentes, FAIR proporciona una comprensión integral y cuantitativa del riesgo en términos monetarios,

lo que permite a las organizaciones tomar decisiones de seguridad informadas basadas en datos.

Para cuantificar el riesgo, FAIR lo divide en dos componentes principales: Frecuencia de eventos de pérdida (LEF) y Magnitud de pérdida (LM).

- **Frecuencia de eventos de pérdida** (LEF): la frecuencia probable, dentro de un período de tiempo determinado, en que el agente de amenaza causará daño a un activo.
    - *Frecuencia de eventos de amenaza*: la frecuencia probable, dentro de un período de tiempo determinado, en que un agente de amenaza actuará contra un activo (intentos).
    - *Vulnerabilidad* o *Susceptibilidad*: la probabilidad de que un evento de amenaza se convierta en un evento de pérdida. También conocido como "susceptibilidad"
- **Magnitud de pérdida** (LM): la magnitud probable de la pérdida resultante de un evento de pérdida.
    - *Pérdida primaria*: las pérdidas primarias se incurren como resultado directo del evento de pérdida en sí, o de las reacciones de la parte interesada primaria al evento.
    - *Pérdida secundaria*: las pérdidas secundarias se incurren cuando las partes interesadas secundarias (partes externas) reaccionan al evento de pérdida, lo que provoca más pérdidas a la parte interesada primaria.

3.2 Enfoque DICYME

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

El marco DICYME utiliza algunos de los elementos de la taxonomía FAIR, 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.

Siguiendo la taxonomía FAIR, estimaremos el Riesgo como el **producto de la Frecuencia y la Magnitud de la Pérdida**. El análisis se realiza por empresa u organización.

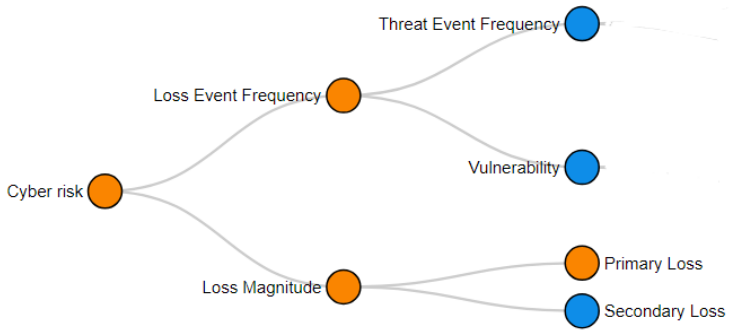


Ilustración 4

3.2.1 Frecuencia

Se definen indicadores para

- (A) *Línea de base*. Es la cantidad promedio de intentos de ataques cibernéticos que enfrenta una organización en un año. Se obtiene a partir de informes de terceros por industria y región.
- (B) *Recuento de incidentes en los últimos 3 años por industria y región*. El recuento es una estimación baja (mínima) del número de intentos. Calculamos la tasa de crecimiento anual utilizando esta serie temporal corta.
- (C) *Atractivo*. La probabilidad de ser víctima de un ciberataque es un número entre 0 y 1.
- (D) *Índice de actores de amenazas de los últimos 3 años*. Es la fuerza de una amenaza para comprometer a la organización. El índice mostrará una tendencia, se utiliza un promedio ponderado. Es un valor entre 0 y 1.
- (E) *Técnica MITRE ATT&CK que un CVE permite a los atacantes explotar*. Lista de técnicas a las que la organización está más expuesta en función de su lista de vulnerabilidades. Utilizamos un índice ponderado de exposición. Cuanto más cerca del impacto, mayor es el peso de la técnica.
- (F) *Perfil de seguridad*. La probabilidad de que una amenaza comprometa a una organización depende también de la capacidad de resistirla (detectarla, responder a ella).
- ***Loss Event Frequency = f(A,B,C,D,E,F)***

3.2.2 Magnitud

Se definen indicadores para

- (H) *Exposición a técnicas de impacto basadas en 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.
- (I) *Pérdida primaria: Tiempo de inactividad*. El sistema simula la duración de la interrupción, el resultado debe multiplicarse por el costo de 1 hora de tiempo de inactividad (reportado por la organización)
- (J) *Pérdida primaria: Daño a los equipos*. El sistema simula la magnitud del daño. El resultado debe multiplicarse por el costo del equipo o la infraestructura en riesgo (reportado por la organización)
- (K) *Pérdida primaria: Extorsión*. El sistema simula la proporción de los ingresos que se retienen como extorsión. Por lo general, se trata de un valor pequeño, no más del 1 % de los ingresos.
- (L) *Pérdida primaria: Daño humano*. El sistema simula la gravedad del daño. Debe multiplicarse por el costo estadístico de una vida (constante conocida).
- (M) *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.

- (N) *Pérdida secundaria: Daño reputacional*. El sistema simula la magnitud del problema. El resultado debe multiplicarse por los ingresos como indicador de la pérdida de participación en el mercado.
- (O) *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)
- *Loss Magnitude*. Simular el árbol con las 77 tuplas (Técnica, Pérdida primaria, Pérdida secundaria)

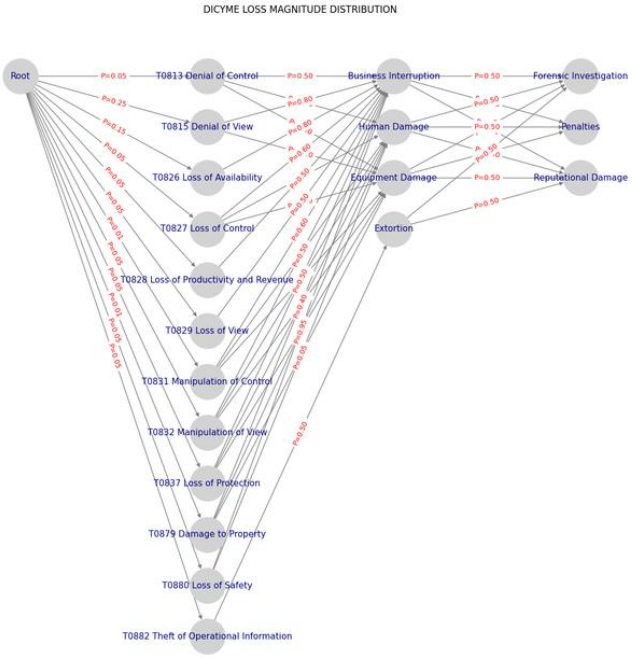


Ilustración 5

4 DATOS Y CÓDIGO

El código de este módulo se encuentra en el repositorio del proyecto: *deriskGroup / DICyME Project · GitLab* , dentro del directorio E.*2.2 Second prototype of measurement and estimation of probability and impact of a cyberincident*.

5 CONCLUSIONES Y SIGUIENTES PASOS

En DICYME se están desarrollando modelos de medida o estimación de factores relacionados con el ciber riesgo. Con relación a la primera fase de DICYME, los indicadores sugeridos se integran en un motor de cálculo de riesgo diseñado por DICYME.

El *Cyber Risk Calculator* propuesto en DICYME requiere una implementación detallada e integración compleja entre todas las fuentes de datos utilizadas y procesos implementados, por lo que el equipo DICYME aún se encuentra trabajando en ello.