Convocatoria Colaboración Público-Privada CPP2021-009025

DICYME:
Dynamic Industrial CYberrisk Modelling based on Evidence

Iniciativa Conjunta de:

DENEXUS

Universidad Rey Juan Carlos

ENTREGABLE 3.1:
Primer prototipo de interfaz gráfica y herramienta de soporte a la decisión

Coordinadores:

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

MINISTERIO DE CIENCIA E INNOVACIÓN

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

AGENCIA ESTATAL DE INVESTIGACIÓN

Contenido

1. INTRODUCCIÓN Y OBJETIVOS

Este documento resume los primeros pasos realizados en el desarrollo de la Actividad 3 (Sistema de Visualización y toma de decisiones), concretamente relativo a las tareas 3.1, 3.2 y 3.3, que ha resultado en el desarrollo del primer prototipo de interfaz gráfica y herramienta de soporte (entregable 3.1).

Este sistema de visualización se vale de los datos extraídos gracias a los módulos de extracción automática de datos desarrollados en la Actividad 1 (Sistema de Recogida de evidencias) y a los resultados de los módulos de medida/estimación de probabilidad e impacto de la Actividad 2 (Módulos de medida/estimación de probabilidad e impacto).

En las siguientes secciones se explica:

- Qué intereses tienen los operadores de infraestructuras industriales y los agentes del sector asegurador en relación con la visualización de indicadores de soporte a la toma de decisiones.
- Qué diseño se ha realizado para la visualización de datos e indicadores en relación con el soporte a la toma de decisiones.
- Qué proceso de desarrollo y despliegue se ha seguido para poner en funcionamiento el primer prototipo de la aplicación.
- Qué datos y código se encuentran disponibles en este entregable (deriskGroup / DICyME Project · GitLab) y cómo pueden emplearse.

2. CONTENIDO DEL SISTEMA DE VISUALIZACIÓN

En primera instancia, se ha diseñado un sistema de visualización genérico que pueda servir tanto a operadores de infraestructuras industriales como a aseguradoras. El planteamiento es que, a medida que se desarrollen prototipos y avance el desarrollo del sistema, se puedan crear módulos separados que muestren información específica para cada agente involucrado.

El sistema diseñado consta de tres módulos principales, a los que se añaden una primera página de resumen o visión general y una pestaña final de información sobre el proyecto y las fuentes de datos.

Respecto a los 2 módulos que representan el núcleo del primer diseño del sistema, el primero de ellos incluirá algoritmos de modelado del ciberriesgo trabajados en otras actividades del proyecto. Este tiene por nombre "*Modelando el ciberriesgo*". El segundo, denominado "*Datos para el ciberriesgo*", contiene representaciones de los diferentes conjuntos de datos generados en dichas actividades.

La aplicación descrita a lo largo de este entregable corresponde con la versión v0.1.0, cuyo desarrollo concluyó en Julio de 2024. Es una versión inicial que ha servido, además

de para dibujar bocetos y esquemas para lo que será el producto final, para construir toda la infraestructura y despliegue de la misma, de manera que se agilice el desarrollo de nuevas versiones del software. Asimismo, ha supuesto una oportunidad para que el equipo explore Shiny sobre el lenguaje de programación y análisis estadístico R, empleado también para el desarrollo de otros conceptos y entregables del proyecto. No obstante, al tratarse de un primer prototipo, el equipo del proyecto se ha reunido para comprobar el estado del mismo y plantear mejoras y cambios sobre él, que serán abordados en posteriores entregables, así como reestructuraciones del menú y nuevos módulos, entre otros, que ayudarán a explicar de una manera lógica y útil todo el trabajo desarrollado en el proyecto.

2.1 Modelando el ciberriesgo (Indicadores)

El propósito de este primer módulo es que sea dinámico, reactivo y amigable. Debe permitir la ejecución de los diferentes modelos e indicadores definidos en otras actividades del proyecto. Este módulo está en proceso de desarrollo y no se detalla el contenido de sus elementos en este primer prototipo de la interfaz gráfica.

Respecto a la estructura de la visualización, la primera pestaña que incluye es para los actores de amenaza. En concreto, se pretende poder introducir diferentes variables o parámetros que condicionen el indicador, para obtener el resultado en diferentes casos.

En el caso de la segunda pestaña, esta es similar a la anterior, pero en este caso empleando el modelo de atractivo definido en la Actividad 2, e introducido en el Entregable 2.1 del proyecto. En la pestaña se podrá calcular el atractivo de una empresa dadas sus características firmográficas (país, tipo de industria, facturación, o número de empleados).

La tercera y última pestaña del módulo es para el módulo de estimación de impacto definido también en el Entregable 2.1, que conecta CVEs con TTPs de MITRE. Esta página permitirá obtener esta relación, dados los parámetros de un nuevo CVE.

2.2 Datos para el ciberriesgo (Datos)

Este segundo módulo ofrece una forma de explorar los conjuntos de datos generados en el proyecto y empleados en el desarrollo de los modelos e indicadores mencionados. En concreto, por el momento se ha diseñado la visualización de dos conjuntos diferentes, aunque relacionados:

1. El primero de ellos es un conjunto de datos que combina la información de incidentes extraída de múltiples bases de datos públicas, empleando los procedimientos automáticos desarrollados en la Actividad 1 y detallados como parte del Entregable 1.2.
2. El segundo es el conjunto de incidentes enriquecido, detallado también en el Entregable 1.2, que contiene tanto incidentes como los datos firmográficos de las víctimas.

3. DISEÑO DEL SISTEMA DE VISUALIZACIÓN

Se han realizado diferentes previsualizaciones de los módulos y pestañas enunciadas en la sección anterior. Estos han servido a los equipos para debatir y preparar el mejor diseño posible, así como poner en común ideas de elementos que se podían incluir o agrupar. Además, también se han desarrollado visualizaciones preliminares de los conjuntos de datos, mediante herramientas como Databricks, que han permitido generar ideas del contenido que se quería representar en la interfaz gráfica. La siguiente figura (ver Ilustración 1) es un ejemplo de estos *mockups*, ilustrando en concreto la estructura de menús de la herramienta y la pestaña de cálculo del atractivo de una entidad.

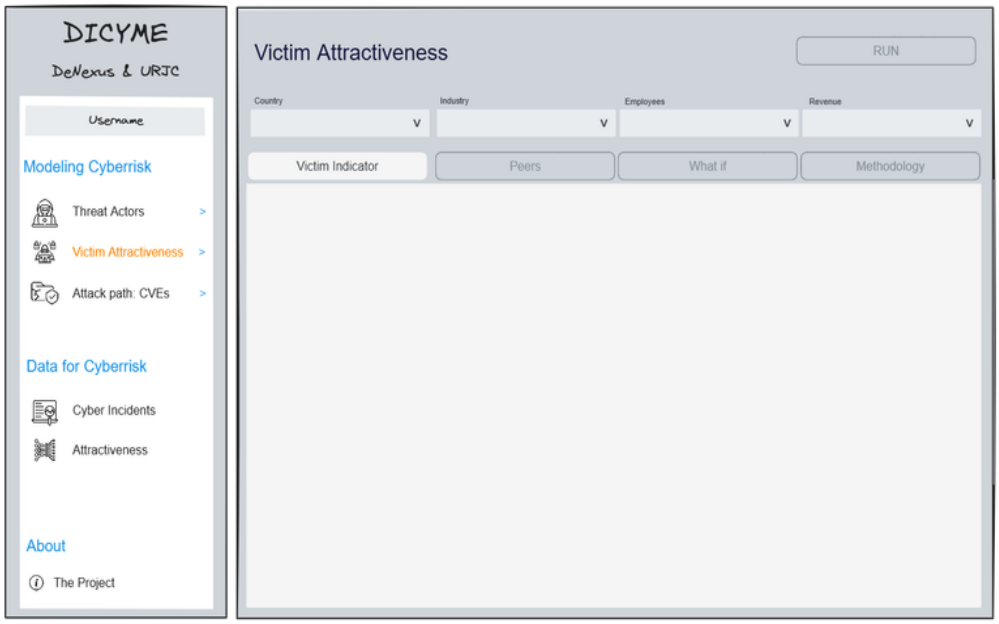

Ilustración 1 - Diseño del sistema de visualización.

Además, se han recopilado los manuales de identidad visual corporativa de los dos socios del proyecto, de manera que la herramienta desarrollada incluya los elementos posibles, acercándola al estilo de las dos entidades.

4. DESARROLLO DEL SISTEMA DE VISUALIZACIÓN

Este primer prototipo de interfaz gráfica se ha desarrollado empleando el lenguaje de programación R. En concreto, se ha empleado el paquete Shiny, que permite desarrollar y desplegar aplicaciones web interactivas.

Dentro del primer prototipo, se ha comenzado desarrollando el módulo de datos para el ciberriesgo. En primer lugar, para la pestaña de ciberincidentes (ver Ilustración 2), se ha creado un gráfico de barras apiladas que muestra, para cada base de datos pública, la cantidad de incidentes acontecidos en una ventana temporal. Este periodo puede ser modificado mediante un selector, que permite escoger los valores: "*Día*", *"Semana"*, *"Mes"*, *"Cuatrimestre"*, *"Año"*. En función de la selección, el gráfico cambia dinámicamente, obteniendo los periodos naturales y modificando la representación.

Igualmente se ha realizado para datos enriquecido para la métrica de atractivo ((ver Ilustración 3), pero esta vez agrupando los eventos por tipo de incidente, incluyendo los no incidentes. Además, en este caso también se ha incluido un mapa coroplético que muestra la cantidad de incidentes recogidos por país.

Ambas pestañas incluyen un apartado que muestra en formato tabla el conjunto de datos completo, con filtros para cada columna y posibilidad de ordenarlas ascendente y descendentemente, así como un campo de búsqueda libre. De esta manera, se puede explorar los conjuntos de datos completos, buscar entidades específicas, o incidentes (ver Ilustración 4).

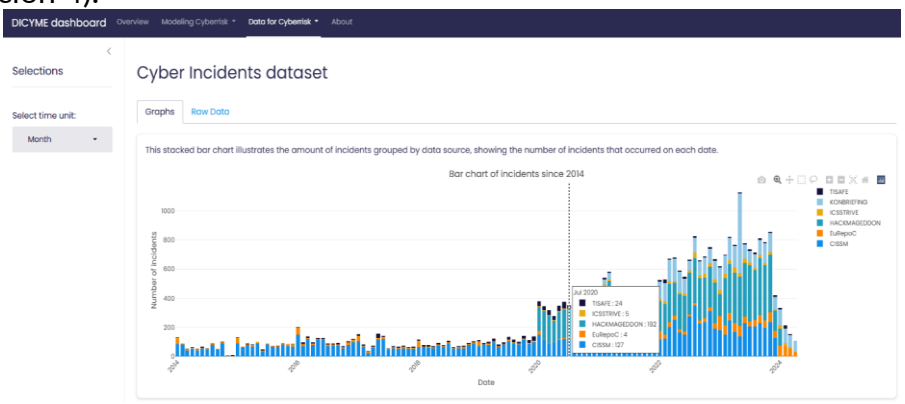

Ilustración 2 - Pestaña con gráficos del conjunto de datos de ciberincidentes.

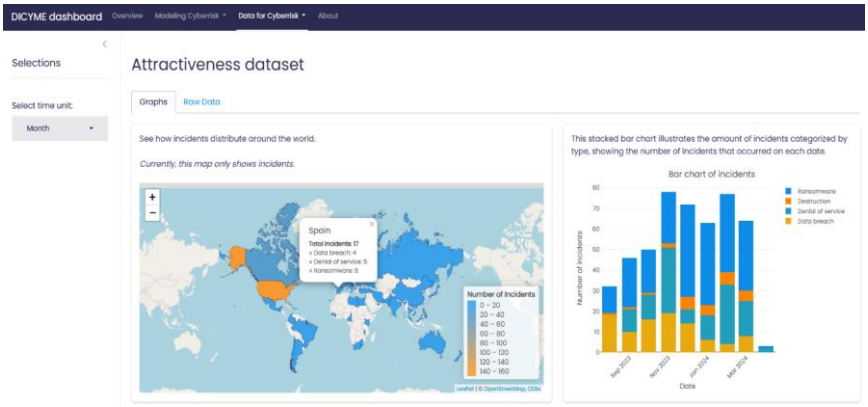

Ilustración 3 - Pestaña con gráficos del conjunto de datos de atractivo.

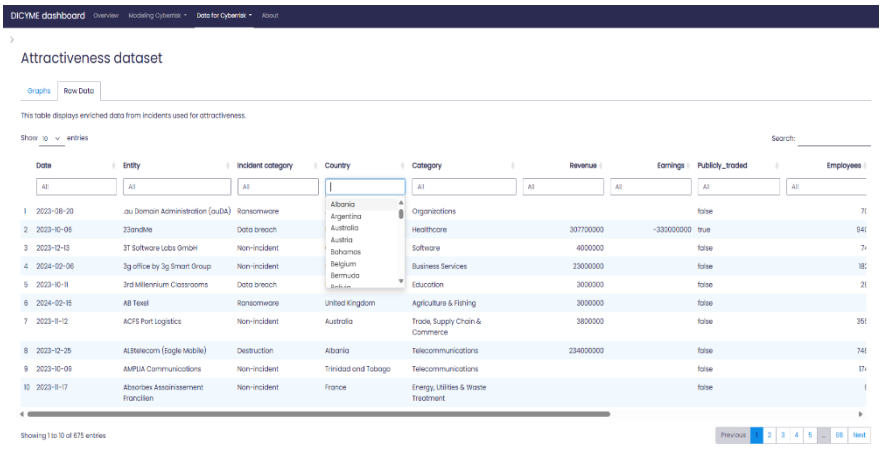

Ilustración 4 - Pestaña con tabla del conjunto de datos de atractivo.

5. TECNOLOGÍA Y DESPLIEGUE

Como ya se ha introducido en la sección 2, la tecnología escogida para el desarrollo de la interfaz gráfica es una aplicación mediante la librería Shiny desarrollada en el lenguaje R. Esta permite desarrollar aplicaciones web interactivas y dinámicas sin un conocimiento avanzado en desarrollo web, con una arquitectura cliente-servidor. El navegador del usuario muestra los resultados generados en el servidor, que ejecuta código R.

Este tipo de aplicaciones permiten manejar grandes volúmenes de datos y realizar cálculos complejos en tiempo real como respuesta a entradas del usuario, pudiendo personalizar los gráficos y visualización de manera dinámica. Además, se pueden integrar con otros paquetes de R, ampliando sus capacidades y potencia para el desarrollo de aplicaciones web avanzadas.

Para mejorar la calidad del código en la aplicación Shiny, se utiliza `lintr`, una herramienta de análisis estático para R que identifica errores de sintaxis y posibles problemas semánticos, y proporciona informes detallados para que los desarrolladores puedan tomar medidas correctivas. Esto es beneficioso porque ayuda a mantener un código limpio y consistente, facilita la detección temprana de errores, y mejora la legibilidad y mantenibilidad del código.

Además, toda la aplicación se ha desarrollado como un paquete de R en sí misma, lo que significa que está diseñada y estructurada como un conjunto de funciones y herramientas que pueden ser fácilmente instaladas y utilizadas dentro del entorno de R. Entre las ventajas que esto aporta, se encuentra el uso de `renv` para el control de versiones de las dependencias. Esta herramienta permite gestionar las versiones de los paquetes de R utilizados en la aplicación, asegurando que el entorno de desarrollo sea reproducible y consistente por otros desarrolladores y minimizando los problemas de compatibilidad.

Los datos utilizados en la aplicación se obtienen de la base de datos MongoDB que fue desplegada previamente en los servidores de la URJC, en la que se almacenan los datos e indicadores generados a lo largo del proyecto. El acceso a la misma está restringido y únicamente se permite desde la red VPN de la URJC, a la que tiene acceso también el equipo de DeNexus, y desde los puestos de trabajo de la red interna de los miembros del equipo.

En dichos servidores se ha desplegado también la aplicación basada en Shiny, dentro de un contenedor Docker, y con otro contenedor dispuesto con un proxy Apache que proporciona cifrado con el protocolo SSL. Para ello, se emplea un certificado emitido por GÉANT para el dominio del servidor.

Cabe destacar que todo el proceso de despliegue se ha automatizado con GitHub Workflows y se realiza cada vez que el equipo de desarrollo publica una nueva versión de la aplicación en la rama principal de la herramienta de control de versiones.

Una vez detalladas la tecnología y proceso del despliegue, es preciso mencionar la URL de acceso a la aplicación, aunque la versión disponible puede ser posterior a la descrita en este entregable: https://gondor.etsii.urjc.es:3866/ .

6. DATOS Y CÓDIGO

En el enlace deriskGroup / DICyME Project · GitLab, dentro del directorio *E.3.1 First prototype of a graphical interface and decision support tool*, se puede encontrar el directorio `dicymeviz-0.1.0/` que contiene el código de la aplicación en formato paquete de R. El archivo `README.md` detalla la estructura y modo de ejecución de la aplicación. Caben destacar algunos elementos del código:

- `renv.lock` contiene las versiones de todos los paquetes de R empleados, así como sus dependencias y versiones de las mismas.
- `R/` contiene el código R con las diferentes funciones que componen la aplicación.
- `.github/workflows/` contiene el código con las automatizaciones del despliegue y análisis estático.
- `DESCRIPTION` contiene la información del paquete, versión, autores, etc.

Aunque las versiones de todas las dependencias aparecen especificadas en el archivo renv.lock y se pueden restaurar de manera sencilla en cualquier entorno de desarrollo, cabe destacar las siguientes:

- `R 4.4.1`
- `Shiny 1.8.1.1`
- `bslib 0.7.0.9000`

7. CONCLUSIONES Y SIGUIENTES PASOS

En este punto del proyecto se ha desarrollado un primer prototipo de una aplicación web mediante la tecnología Shiny en R. Este sistema permite visualizar los datos generados en el proyecto de manera sencilla e interactiva desde una aplicación web. Para ello, se ha conectado el servidor de la aplicación con la base de datos MongoDB alojada en los sistemas de la URJC.

En siguientes iteraciones de este prototipo se pretende añadir nuevos selectores de rangos temporales para modificar dinámicamente los gráficos y series temporales ya creados en este prototipo, nuevos gráficos y paneles de visualización que permitan representar otras dimensiones de los datos, visualización de indicadores e integración para cuantificar el ciberriesgo, así como todos los cambios que ocurran en futuros prototipos de las Actividades 1 y 2.

Además, se implementará todo el módulo 1, Modelando el ciberriesgo, que integrará y permitirá la ejecución de los modelos de estimación de probabilidad e impacto definidos en otras actividades.