

CIDAI

Centre of Innovation
for Data tech
and Artificial Intelligence

Promotors



Membres



Reseguimiento Red de Gas

Data: 11 maig 2023

www.cidai.eu

Índice de los contenidos





tras una reflexión interna, identificó la oportunidad de cierto **potencial de optimización** en el modelo de Reseguimiento

de sus redes de gas, donde ajustando las frecuencias de Reseguimiento en función de la condición del material de modo que se minimizase el **riesgo de fuga**, y aplicando los recursos allí dónde son más necesarios, en vez de reseguir una red nueva de materiales mas modernos que cumplen la normativa (ej. Polietileno) puesta en servicio hace 2 años o menos.

Ante esta necesidad, **Nedgia** lleva a cabo una iniciativa para poder analizar mediante los datos la posibilidad de cambiar las dinámicas de los mantenimientos en función del riesgo de fuga de gas.

En el marco de la inteligencia artificial, existen amplias posibilidades de generar valor gracias al uso apropiado de **técnicas de analítica avanzada**. Haciendo un correcto uso de los datos que proporciona el negocio, se generan análisis de comportamiento y de mejora. En este contexto, los datos de reseguimiento nos ofrecen la **visión de una nueva estrategia**.

SITUACIÓN ACTUAL

Actualmente el reseguimiento de redes y acometidas se realiza en base a 2 criterios el emplazamiento y la MOP para segmentar las redes y acometidas en frecuencias de 2 y 4 años de acuerdo a la regulación actual

CAMBIO

FRECUENCIAS

SITUACIÓN IDEAL

Al introducir el riesgo de fuga, así como otras variables como los materiales, diámetros, tipo de gas, etc., se puede definir una **nueva estrategia** en la frecuencia de los reseguimientos que permita tener **granularidad hasta 6 años** basadas en el riesgo asociado a los subtramos

Objetivos de la PoC

Nedgia, dentro del proceso de Reseguimiento de la red de gas se ha detectado una posible mejora en la planificación para lo cual se plantea la siguiente mejora a través de modelos predictivos utilizando IA:

Optimizar la frecuencia en la que realizar las tareas de mantenimiento, con el fin de minimizar el numero de visitas en función del riesgo de fuga, y de forma que en un periodo determinado se haya realizado un mantenimiento de toda la red.

- 1 Identificar el nivel de riesgo de las redes, para ayudar en la planificación del reseguimiento.
- 2 Realizar una clasificación de las redes de suministro en función del nivel de riesgo que permita una optimización de frecuencia de visitas a realizar.
- 3 Construir, en base a la clasificación, una herramienta de planificación del reseguimiento con la frecuencia adecuada para cada tipo de red.

Objetivos de la PoC

Con el objetivo de planificar de manera óptima el reaseguimiento de redes y acometidas en el marco del mantenimiento dinámico, se desarrollan **3 modelos analíticos secuenciales a nivel de subtramo**:

1. **Modelo de riesgo de fuga:** clasificación binaria (0: *No Fuga*, 1: *Fuga*) cuya predicción tiene formato de score (entre 0 y 1)
2. **Segmentación en clústers de frecuencias de reaseguimiento**, basados en scores de riesgo y tipologías
3. **Distribución en zonas de reaseguimiento y planificación anual** en función de frecuencias y cercanía entre subtramos



Modelo de scoring de riesgo de fuga

input: fugas detectadas (rsgs y recl.) y tipología de los subtramos

output: riesgo de fuga



Segmentación por riesgo de fuga

input: riesgo de fuga y tipología

output: frecuencia de reaseguimiento por agrupaciones



Distribución en zonas reaseguimiento

input: frecuencia de reaseguimiento

output: zonas de reaseguimientos y planes anuales

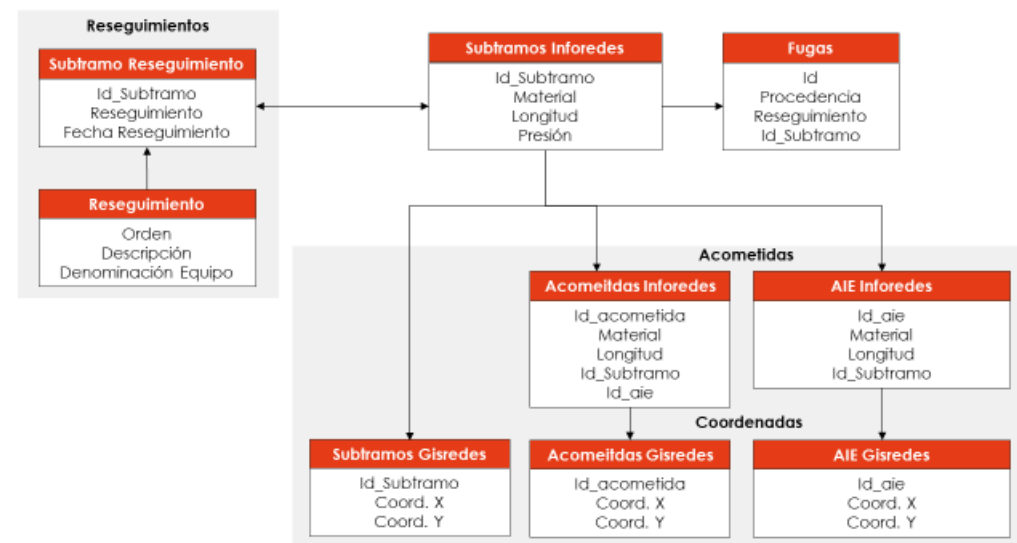
Entendimiento de negocio

Los datos recibidos para el modelado de las fugas en red y acometidas fueron **analizados y tratados para su correcta inclusión en el modelo**. Abajo los ficheros recibidos (tabla a la izquierda) y la relación entre ellos (flujograma a la derecha).

Datos recibidos:

- Características de subtramos, de las acometidas y acometidas interiores conectadas a ellos: propiedades y coordenadas
- Reseguimientos realizados a los subtramos incluyendo su planificación
- Fugas detectadas en los subtramos incluyendo las reclamaciones
- Datos de catálogos de materiales, presión, propiedad, etc.

Fugas	Total registros: 37,581. No contiene duplicados			
Infiores Subtramos	Total registros: 3.000.176	Total Subtramos: 1.547.433		
Gisredes Subtramos	Total registros: 1.494.871	Total Subtramos: 1.494.518		
Infiores Acometidas	Total registros: 1.494.871	Total Subtramos: 1.494.518		
Gisredes Acometidas	Total registros: 3.353.393	Total Subtramos: 1.494.518		
Infiores Acometidas Interior Enterrada	Total registros: 16.162	Total Subtramos: 13.667		
Gisredes Acometida Interior Enterrada	Total registros: 52.887	Total Subtramos: 13.667		
Relación Subtramos – Reseguimiento	Total registros: 8.483.347	Total Subtramos: 1.630.988		
Reseguimiento	Total registros: 20.313	Total reseguimientos: 4.948		
Catálogos		Diámetro	Presión	Propiedad Material



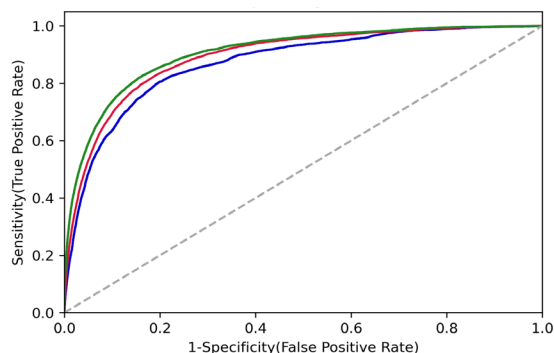
Riesgo de fuga

El modelo de **XGBoost** muestra resultados ligeramente superiores. Con una sensibilidad de 0.82 y un *lift* en el primer decil de 7.32. Los resultados por año son similares, siendo ligeramente mayor el lift de 2019.

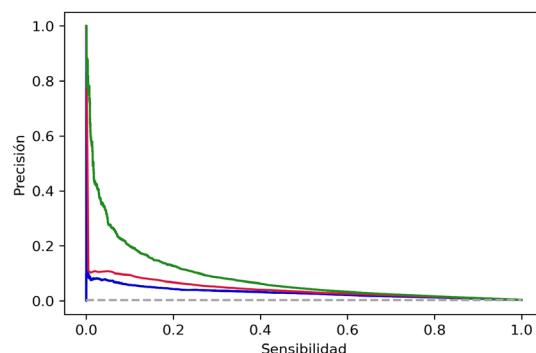
	Regresión Logística	Random Forest	XGBoost
• ROC-AUC	0.87	0.89	0.91
• Gini	0.74	0.78	0.81
• Recall (sensibilidad) ⁽¹⁾	0.76	0.81	0.82
• Lift global (todos los deciles)	4.56	4.64	5.11
• Lift 1º decil	6.32	6.89	7.32
• Lift 1º decil 2019	6.55	6.97	7.36
• Lift 1º decil 2020 ⁽²⁾	6.17	6.75	7.29
• Lift 1º decil 2021 ⁽²⁾	6.24	6.91	7.32

— LR
— RF
— XGB

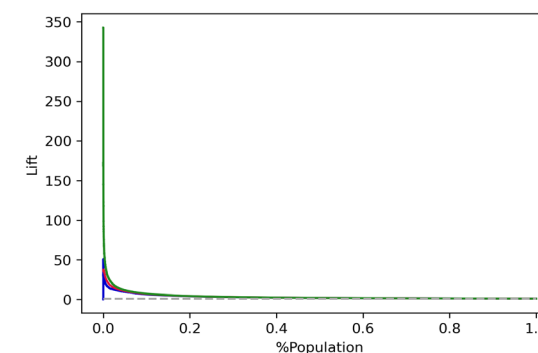
ROC



Recall



Lift



Nota: (1) la sensibilidad mide el porcentaje de aciertos de la clase *Fuga* (tasa de verdaderos positivo) y es una de las métricas prioritarias para este caso de uso (2) en los años 2020 y 2021 los datos de fugas no son completos a falta de 2 meses (2020 muy pocas reclamaciones por Covid y 2021 por datos extraídos en noviembre)

Riesgo de fuga

El análisis de poder discriminante permite identificar **aquellas variables más relevantes** para incluir en el modelo. Mientras que la combinación final de ellas en el modelo queda reflejada con el **análisis de importancia**.

Variables	Tipo	Poder Discriminante	Importancia en el modelo
Número de acometidas/AIEs	numérica	Tendencia creciente	alta
Tipo de gas	categorica dummy (x2)	Mayor para gas propano	alta
Presión (rangos)	categorica dummy (x5)	Tendencia creciente	alta
Material de subtramo y acometidas	categorica dummy (x5)	Relevancia	media
Boolean reclamación previa	boolean	Poca relevancia	media
Antigüedad	numérica	Tendencia creciente	media
Diámetro (mm)	numérica	Tendencia creciente "alterada"	media
Boolean nivel previo grave (1, 2)	boolean	Poca relevancia	baja
Longitud	numérica	Tendencia creciente	baja

Nota: el análisis de poder discriminante es bivalente (1 variable vs. target), mientras que en el modelo la interacción es multivalente y deben entenderse como una combinación

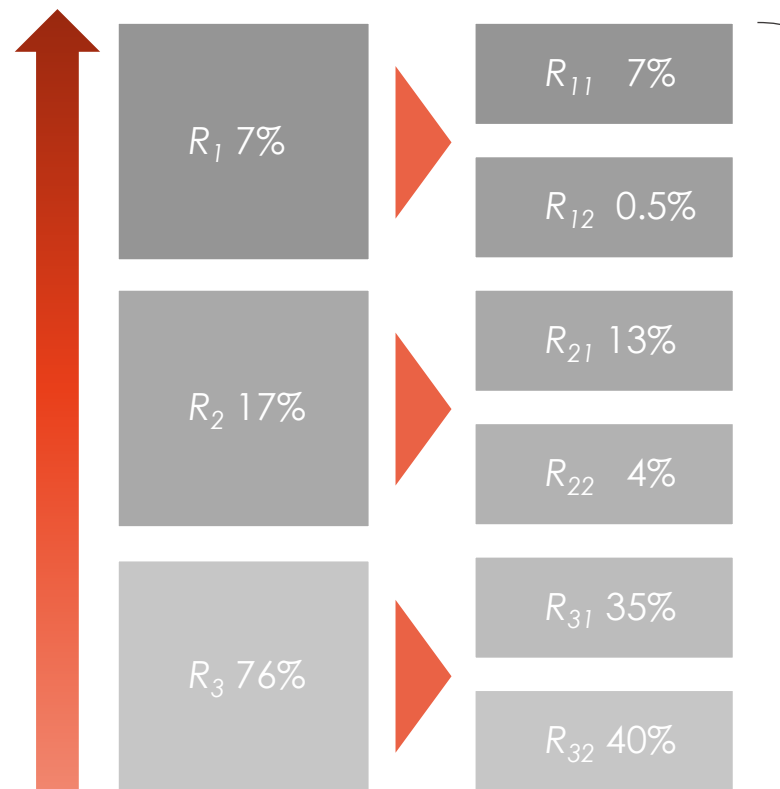
- (1) El modelo de XGBoost muestra los resultados más robustos, que además se mantienen relativamente constantes a lo largo de los 3 años del subconjunto de validación (2019 a 2021). Priorizamos este tipo de algoritmo.
- (2) Las variables de presión, tipo de gas y número de acometidas cobran mayor importancia en el modelo

Segmentación

Con el modelo de *K-means*, se generan 3 clústers automáticos por ser el número óptimo de grupos. Posteriormente se subdividen los clústers en 6 grupos finales siguiendo unas reglas sobre las propiedades de los subtramos. Las reglas quedan confirmadas con los perfilados.

De 3 clústers automáticos a 6 ordenados

+ riesgo

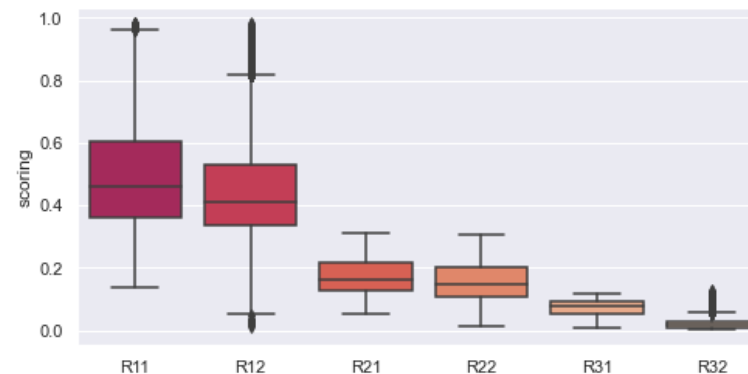
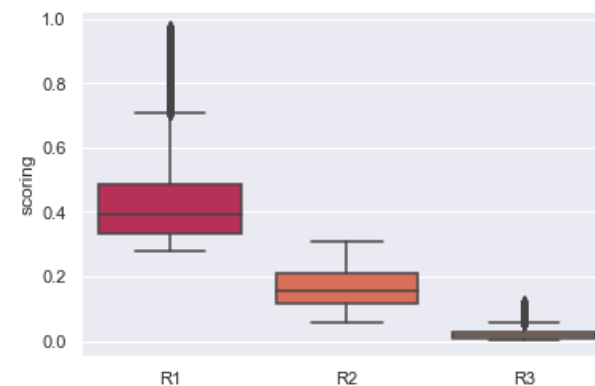


Reglas:

En cada clúster R_j , se establecen con mayor riesgo (R_{j1}) aquellos subtramos con:

1. Gas propano
2. Presión baja y media B
3. Material distinto a AO o PE

Distribución del scoring de fuga

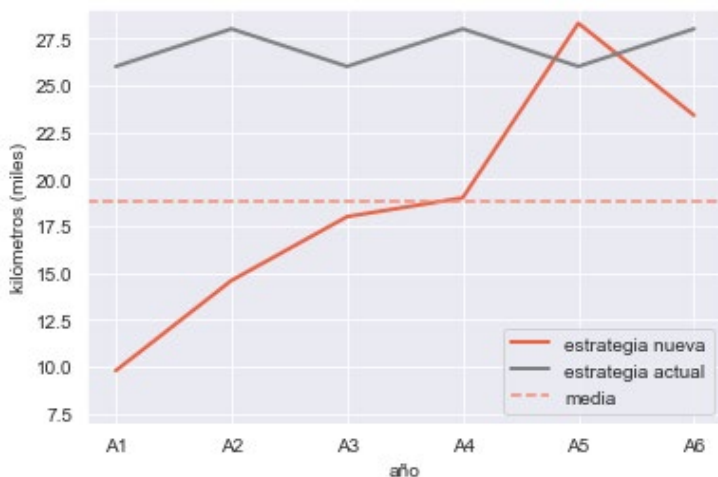


- riesgo

Optimización

El objetivo de la optimización es obtener una **planificación de las visitas más eficiente** y por ende **con menores costes asociados**. Esto se lleva a cabo gracias a la disminución total de kilometraje inspeccionado a lo largo de un ciclo (en este caso, de 6 años), tomando como input las frecuencias de la segmentación (basadas en el riesgo de fuga). Además, se incluye una componente geográfica para planificar por cercanía.

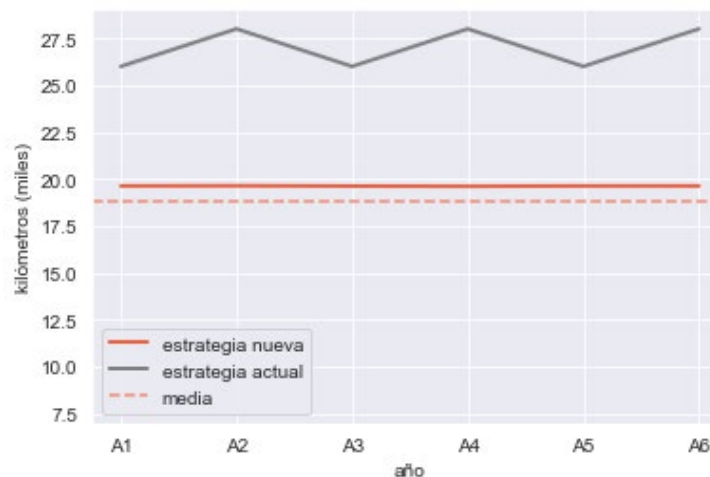
Sin optimizar la distribución



media anual

26.5k KMs

Tras la optimización



media anual

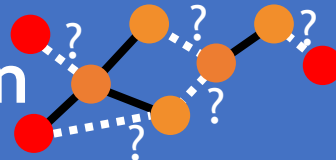
19.6k KMs

Ganancia

Mejor distribución: se aplanan los picos en los años 5 y 6

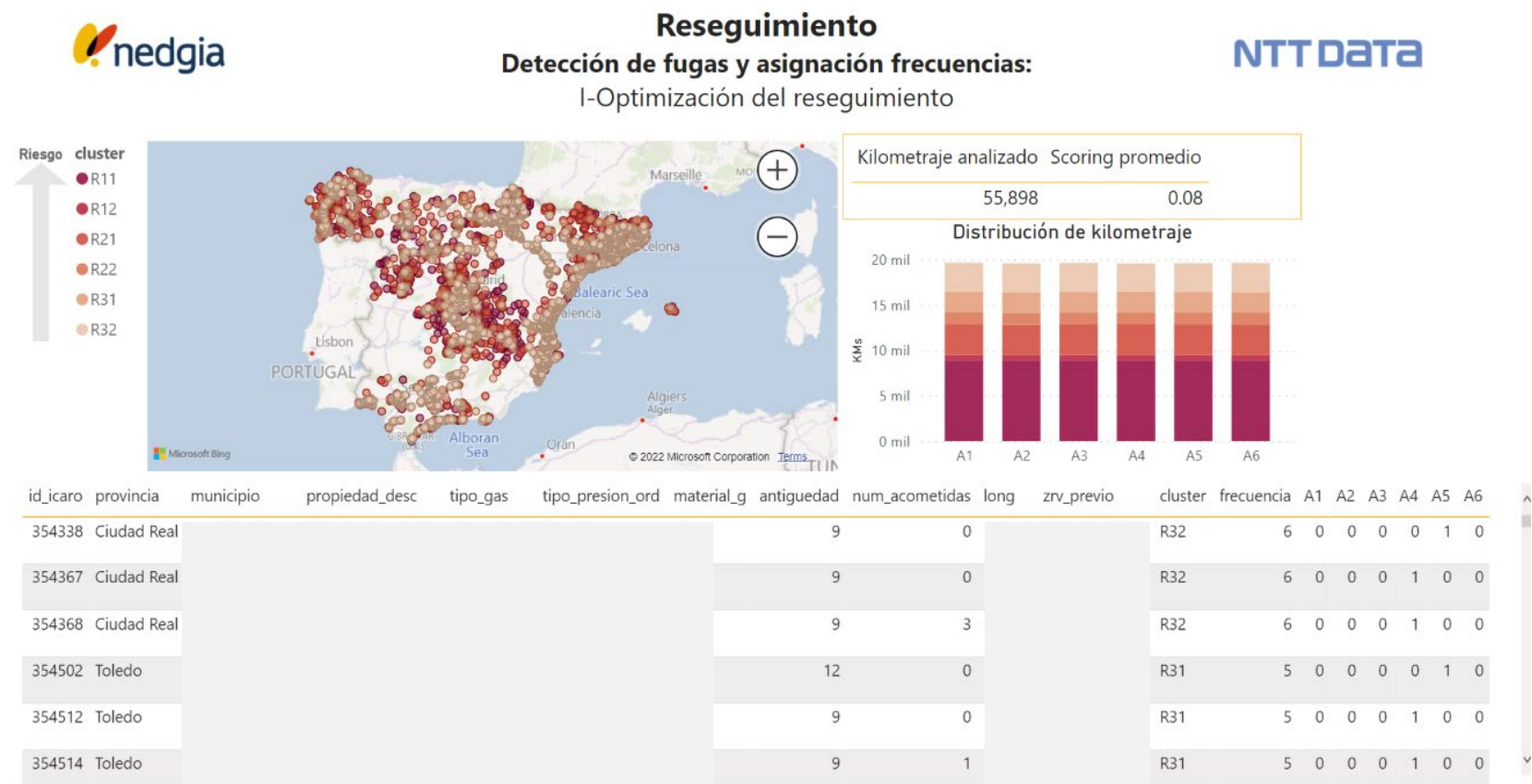
ahorro anual: 25%

7k KMs



Se ha realizado la parte de Visualización en Power Bi, de forma que se puedan consultar la información de los modelos.

Una primera página presenta **la asignación de subtramos a reseguimientos** en un ciclo de 6 años. Se muestra la distribución geográfica (por coordenadas en el mapa de la izquierda) y de kilometraje a lo largo de los 6 años. En la parte inferior, aparecen la tabla de subtramos, que puede ser filtrada por año de reseguimiento, provincia, municipio, tipo de gas, presión y material.



- Visión:
- Sin filtrar los datos en los genéricos

NOTAS: (a) los subtramos sin longitud o sin reseguimiento anterior se incorporan con valores estándar de un metro y reseguimiento el año anterior (b) WIP: tabla traspuesta

Conclusiones

- Como la mayor parte de los tramos tiene un riesgo pequeño, esto hace que puedan realizar seguimientos hasta en 6 años, lo que permite una **reducción del 25% de Kms anuales** de Reseguimiento, con el consiguiente **ahorro en costes en el mantenimiento de la red**
 - Introducir el riesgo de fuga por tramo permite una **mejor trazabilidad de las zonas mas problemáticas** que deben reseguirse con mayor frecuencia, y de **aquellas más estables** a reseguir con menor frecuencia
- Adicionalmente, y sin buscar este objetivo inicialmente, al realizar un Reseguimiento con mayor frecuencia de los tramos de mayor riesgo, se adelanta la detección de fugas, lo que se traduce en una **reducción del 20% de emisiones de gas** (efecto invernadero), lo que ayuda en la sostenibilidad de las tareas de mantenimiento

- El **grado de madurez** de estas técnicas es suficientemente **alto** como para plantearse su implementación en soluciones en ámbito empresarial.
- El uso de la IA en una base de datos en formato GIS puede ser de gran interés e impacto para la solución de los siguientes casos:



Optimización **de tareas**



Detección de fallos en una red con localización



Mejora de la sostenibilidad de las tareas



Detección de **materiales a sustituir**

CIDAI

Centre of Innovation
for Data tech
and Artificial Intelligence

Promotors



Membres



info@cidai-catalonia-ai.eu

www.cidai.eu