

BARRERAS PARA LA IMPLEMENTACIÓN EFECTIVA DE LA IA EN LA PREDICCIÓN DE OBRAS PÚBLICAS

ELENA USOBIAGA¹, PATRICIA MOLINA-COSTA¹, BEGOÑA ISPIZUA¹, JOSÉ LUIS IZKARA²

¹ TECNALIA, BASQUE RESEARCH AND TECHNOLOGY ALLIANCE (BRTA), ESPAÑA

² UNIVERSIDAD DE DEUSTO, ESPAÑA

PALABRAS CLAVE

*Inteligencia artificial (IA)
Planificación urbana
Predicción de obras públicas
Calidad de los datos públicos
Intercambio de datos
Ciudades inteligentes*

RESUMEN

Los problemas de calidad de los datos públicos suelen obstaculizar las aplicaciones de inteligencia artificial (IA) en la planificación urbana, afectando la aplicabilidad, eficacia y resultados de los modelos. En este artículo se presenta un estudio de caso sobre la predicción de obras públicas afectadas por la infraestructura de red y su impacto a nivel municipal. Este caso ha servido para identificar barreras y limitaciones en la adopción de la IA en este subdominio, permitiendo formular un conjunto de recomendaciones para mejorar la producción y el intercambio de datos públicos, sentando las bases para futuros modelos de IA en la predicción de obras públicas. Al abordar estos desafíos, las ciudades pueden aprovechar al máximo el potencial de la planificación y toma de decisiones urbanas impulsadas por la inteligencia artificial.

Recibido: 06 / 08 / 2025

Aceptado: 31 / 10 / 2025

1. Introducción

La inteligencia artificial (IA) es una tecnología clave que cada vez cobra más relevancia en los entornos urbanos. El auge de los *local digital twins* (LDT), junto con la IA, abre nuevas oportunidades para la experimentación y la predicción en el campo de la planificación urbana (Villanueva-Merino et al., 2024). De hecho, la integración de la IA en la planificación urbana ha dado lugar a la aparición de las «ciudades inteligentes», entornos urbanos que aprovechan la capacidad de la IA para analizar grandes cantidades de datos y apoyar la toma de decisiones informadas. Los avances tecnológicos han impulsado la evolución de los sistemas urbanos inteligentes, y ahora se está produciendo un cambio hacia ciudades que aprenden y mejoran por sí mismas (Wu, 2025).

Las ciudades se enfrentan a importantes retos a la hora de implementar la IA debido a la falta de integración entre departamentos, lo que da lugar a información dispersa o incompleta que a menudo queda aislada en los diferentes departamentos, lo que dificulta el intercambio y la coordinación eficientes de datos (Urrutia-Azcona et al., 2021). Además, en el ámbito de la planificación urbana, el acceso a datos desagregados de alta calidad con metadatos relevantes sigue siendo un reto importante (Urra-Uriarte et al., 2024). Para superar estos obstáculos, es esencial adoptar una visión centralizada del intercambio de datos que dé prioridad a los datos precisos y procesables, en consonancia con las necesidades y requisitos de los usuarios (Villanueva-Merino et al., 2024).

Este artículo se centra específicamente en las posibilidades y limitaciones de la aplicación de la IA para predecir las obras en los espacios públicos, combinando datos públicos con datos privados sobre el agua, el gas, la electricidad, las telecomunicaciones y la infraestructura de la red municipal, con el fin de minimizar los posibles incidentes o interrupciones causados por las obras públicas (como, por ejemplo, los cortes prolongados en el suministro de electricidad, agua o redes de fibra óptica, así como las interrupciones en los sistemas de distribución de gas, que plantean importantes riesgos para la seguridad y retos operativos).

En concreto, esta investigación utiliza datos privados de Inkolan, una organización que proporciona información digital de redes urbanas y proveedores de servicios. Esta organización opera en España, por lo que el ámbito de aplicación se enmarca dentro de este país. Tras analizar la disponibilidad de datos públicos en algunos municipios, el estudio de caso se centra en la ciudad de Madrid, trabajando con datos abiertos sobre obras públicas finalizadas. El estudio de caso nos permite abordar las dificultades encontradas en la aplicación de modelos de IA en el campo de la predicción de obras públicas en espacios públicos, con el objetivo de extraer conclusiones y recomendaciones de mejora que faciliten la producción de datos de alta calidad para mejorar la aplicabilidad de futuros modelos de IA en este subdominio de la planificación urbana. La estructura del artículo incluye una revisión bibliográfica del estado del arte, una descripción del estudio de caso y la metodología empleada, un análisis de los resultados obtenidos, y concluye con una sección de discusión que aborda algunas recomendaciones de mejora, así como conclusiones generales.

2. Estado actual de la técnica

2.1. Requisitos de datos para la aplicación de la IA

El crecimiento exponencial de las capacidades y aplicaciones de la IA en los últimos años puede atribuirse a la convergencia de tres factores clave. En primer lugar, el rápido aumento de la potencia de procesamiento, impulsado en parte por los principios descritos en la Ley de Moore (Garg, 2021; Moore, 2006). En segundo lugar, la proliferación de datos digitales procedentes de múltiples fuentes (Gao et al., 2024; Liu et al., 2018). Por último, el desarrollo de algoritmos avanzados de aprendizaje automático, en particular modelos de aprendizaje profundo basados en redes neuronales artificiales con miles de millones de parámetros interconectados (Damioli et al., 2025; Lu et al., 2018). En conjunto, estos factores han transformado la IA en una tecnología de uso general con amplias aplicaciones.

Los datos son el elemento esencial para el funcionamiento de los sistemas de inteligencia artificial. Los modelos de aprendizaje automático se entrenan, validan y optimizan a través de los datos. La calidad, la cantidad y la diversidad de esos datos influyen directamente en el rendimiento, la precisión y la generalización de las aplicaciones de IA. Numerosos estudios han propuesto diferentes dimensiones de calidad de los datos en función de sus objetivos de evaluación y de las etapas del ciclo de vida del conjunto de datos (Gong et al., 2023). Un modelo práctico y ampliamente utilizado tiene en cuenta cinco

dimensiones clave: precisión, relevancia, exhaustividad, actualidad y coherencia (Fernández Álvarez y Garaizar Sagarminaga, 2024). La precisión implica que los datos deben representar fielmente los fenómenos que describen; la relevancia requiere que los datos sean pertinentes para el problema que se estudia; la integridad garantiza que no falten registros críticos; la actualidad se refiere a la disponibilidad de datos actualizados; y la coherencia garantiza la coherencia entre las fuentes y a lo largo del tiempo. Las deficiencias en cualquiera de estas dimensiones pueden dar lugar a modelos menos precisos, menos sólidos y menos generalizables, mientras que las mejoras sistemáticas en la calidad de los datos aumentan la fiabilidad y el valor de las soluciones de IA (Mohammed et al., 2025). Por esta razón, la gestión activa de la calidad de los datos se considera un pilar fundamental para el desarrollo de una inteligencia artificial fiable y eficaz.

Estos principios cobran especial importancia en ámbitos complejos como la planificación urbana, que es intrínsecamente multidimensional, a largo plazo e implica diferentes dimensiones y múltiples partes interesadas, entre ellas las administraciones públicas, los ciudadanos y los agentes privados. El proceso de planificación urbana es dinámico y depende del contexto, por lo que requiere conjuntos de datos heterogéneos, como información geoespacial y catastral, indicadores socioeconómicos y medioambientales, datos de sensores de infraestructuras de IoT, imágenes satelitales y aportaciones de la participación ciudadana (por ejemplo, encuestas o redes sociales). La integración y el control de calidad de fuentes tan variadas son fundamentales para el éxito de la implementación de soluciones urbanas basadas en la IA.

2.2. Enfoques basados en la inteligencia artificial para la planificación urbana

La IA está transformando cada vez más la planificación urbana, especialmente en el diseño y la gestión de los espacios públicos. Las herramientas y los algoritmos de IA se utilizan para analizar grandes conjuntos de datos, predecir tendencias urbanas, optimizar el uso del suelo y mejorar la calidad y la accesibilidad de los espacios públicos. Estas tecnologías ayudan a los planificadores a tomar decisiones más informadas, eficientes y creativas, pero también plantean nuevos retos relacionados con la ética, la gobernanza de los datos y la participación de la comunidad.

En este campo, herramientas como el aprendizaje automático y el aprendizaje profundo se aplican ampliamente para el reconocimiento de patrones, la clasificación y la predicción del crecimiento urbano, el uso del suelo y los impactos ambientales. Algoritmos como los bosques aleatorios, las redes neuronales convolucionales (CNN) y las redes generativas adversarias (GAN) se utilizan comúnmente para analizar datos de observación de la Tierra y simular patrones urbanos (Anwar y Sakti, 2024; Chaturvedi y de Vries, 2021; He y Chen, 2024; Park et al., 2023). Además, los sistemas de información geográfica (SIG) integrados con IA permiten mejorar el análisis espacial, lo que contribuye a la optimización de las infraestructuras públicas, la movilidad y las condiciones microclimáticas en los espacios urbanos (Anwar y Sakti, 2024; Korobeinikova et al., 2024).

La IA también se ha utilizado para abordar cuestiones relacionadas con la edad, como la creación de entornos urbanos adaptados a las personas mayores. Se han utilizado modelos basados en datos, incluidas redes neuronales convolucionales profundas, para evaluar las deficiencias en el diseño urbano y proponer mejoras (Delgado-Enales et al., 2022). Las herramientas de planificación urbana basadas en la IA, como el Simulador del Índice de Amigabilidad para las Personas Mayores (AFIS), integran el reconocimiento de imágenes, el aprendizaje automático y la hibridación de metaheurísticas y la optimización basada en la simulación para identificar problemas de accesibilidad y proponer soluciones eficaces (Delgado-Enales et al., 2022; Villanueva-Merino et al., 2024). Otros sistemas de planificación y apoyo a la toma de decisiones combinan datos objetivos con la participación de las partes interesadas, lo que permite a los responsables políticos evaluar las intervenciones con mayor transparencia y legitimidad hacia una planificación urbana sostenible (Delgado Enales, 2024; Park et al., 2023; Shulajkovska et al., 2024). Además, las herramientas de diseño basadas en la IA, como las redes neuronales y los algoritmos de agrupación, pueden generar esquemas de diseño personalizados, como paletas de colores para espacios públicos, adaptados a la cultura local, el clima y las preferencias de los usuarios (Xiao Ting Cheng et al., 2025).

Sin embargo, la adopción de la IA en la planificación urbana también se enfrenta a importantes retos. Las limitaciones técnicas incluyen la escasa infraestructura de datos, los complejos requisitos de los modelos y la calidad de los datos. Las barreras organizativas, como la resistencia al cambio, la falta de

colaboración interdisciplinaria y la insuficiente experiencia y formación de los planificadores (Bibri et al., 2024; Koutra e Ioakimidis, 2023; Rjab et al., 2023). La literatura crítica hace hincapié en que los algoritmos de big data e IA suelen funcionar como «armas de destrucción matemática» (O’Neil, 2016), ya que su opacidad y los sesgos estructurales inherentes a sus conjuntos de datos (como la raza, el género o la edad) amplifican las desigualdades sociales (D’Ignazio y Klein, 2020). Esta cuestión también se aplica a los sistemas con intervención humana, en los que las pruebas demuestran que la exposición a decisiones algorítmicas sesgadas puede modificar y consolidar el juicio humano, convirtiendo la supervisión en un canal que perpetúa y transfiere los propios sesgos de la IA en lugar de corregirlos (Agudo et al., 2024; Vicente y Matute, 2023). Para mitigar estos riesgos, la explicabilidad, la gobernanza y la transparencia son fundamentales. Los métodos de IA explicable (*Explainable AI*) (XAI) permiten a los seres humanos interpretar el funcionamiento y el impacto de los modelos, proporcionando información sobre su lógica, sus limitaciones y sus posibles sesgos. Dada la relevancia pública de la planificación urbana, garantizar la interpretabilidad y la rendición de cuentas de los sistemas de IA es esencial para la confianza ciudadana (Molina-Costa, 2024).

En conclusión, la IA ofrece oportunidades para mejorar la eficiencia, la personalización y la participación de la comunidad en la planificación urbana. La IA puede automatizar tareas rutinarias, mejorar las previsiones y optimizar la planificación espacial, lo que da lugar a ciudades más sostenibles y resilientes (He y Chen, 2024; Mashhood et al., 2023; Zheng et al., 2023). Sin embargo, es esencial abordar los retos y limitaciones asociados a su adopción. Invertir en infraestructura de datos, fomentar la colaboración interdisciplinaria y establecer marcos de gobernanza éticos y transparentes son elementos clave para maximizar las oportunidades que ofrece la IA en la planificación urbana.

Una aplicación específica en la que la IA puede tener un impacto particular es la coordinación de las obras públicas. Las obras públicas tienen un impacto en las actividades económicas de la zona afectada, ya que dificultan el acceso a las empresas. Esto también causa inconvenientes a los ciudadanos, que sufren una reducción de la accesibilidad, una menor disponibilidad de espacios públicos y alteraciones en la calidad del aire y el ruido (Correia y Roseland, 2022). En el caso concreto de las obras públicas, la predicción de este tipo de actividades puede ayudar a mejorar su coordinación (Hamann et al., 2023; Pericault et al., 2023). Esta coordinación es fundamental para aumentar la eficiencia, armonizar a las distintas partes interesadas y reducir el uso de recursos, evitando la apertura repetida de la misma calle en diferentes momentos.

3. Materiales y métodos

3.1. Descripción del caso de uso

El objetivo principal de este caso práctico era explorar las posibilidades de la IA para ofrecer un nuevo enfoque a la predicción y el impacto de las obras en el espacio público basándose en la combinación de datos públicos de los ayuntamientos y datos privados ofrecidos por Inkolan. Inkolan es una organización formada por la mayoría de los principales operadores de servicios públicos de España, que facilita el suministro de información digital sobre agua, gas, electricidad, telecomunicaciones e infraestructuras de redes municipales.

Esta información es clave para llevar a cabo obras en el espacio público y evitar cortes e incidentes en las redes de servicios. Como parte de su actividad, Inkolan cuenta con más de 20 años de historial de descargas de datos y sus metadatos (quién realiza la descarga, cuándo, con qué finalidad, para qué ubicación, etc.). Esta información se considera propicia para la aplicación de la IA a nivel urbano. Esta investigación se centra en dos objetivos:

- Como objetivo principal, desarrollar un algoritmo de IA que prediga las obras realizadas en el espacio público basándose en la información de descarga de Inkolan.
- Como objetivo secundario, desarrollar un algoritmo de IA que pueda predecir las descargas de Inkolan basándose en datos históricos.

A la luz del ámbito de aplicación, es fundamental contextualizar los datos utilizando variables urbanas que ayuden a enriquecer el análisis y a comprender mejor los impactos espaciales de cada obra pública.

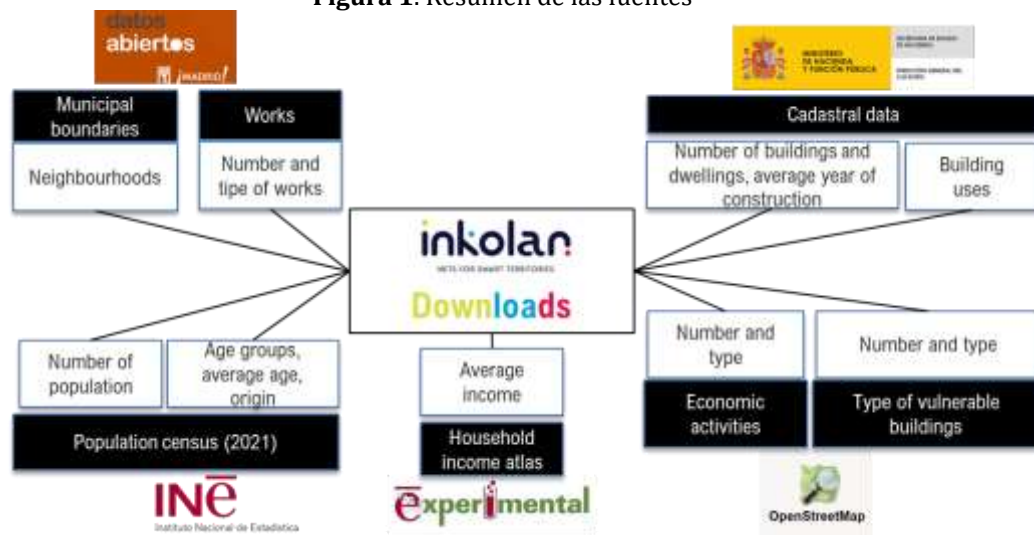
3.2. Fuentes de datos

Este estudio combina fuentes de datos públicos y privados georreferenciados para explorar la viabilidad de aplicar modelos de IA para predecir las obras en el espacio público y analizar su impacto urbano. La geolocalización o la posibilidad de georreferenciarlos es fundamental para respaldar las comparaciones espaciales necesarias para el proyecto.

La principal fuente de datos para el análisis es el conjunto de datos internos de Inkolan, que consiste en metadatos asociados a las descargas de información sobre infraestructuras subterráneas. Cada registro incluye diversos atributos, como el perfil de la entidad solicitante, el propósito de la solicitud, si está vinculada a una obra pública, la fecha de ejecución prevista y un límite geoespacial (denominado «cerca») que define el área geográfica de interés. Además, el estudio incorpora múltiples fuentes de datos públicos, agrupadas en dos categorías:

- Datos de obras públicas: estos conjuntos de datos describen la ejecución de intervenciones en el espacio público e incluyen licencias y registros de obras completadas. Cuando se georreferencian adecuadamente (por ejemplo, con coordenadas o direcciones), pueden vincularse a los registros de Inkolan para validar y entrenar modelos predictivos.
- Datos contextuales urbanos: estos conjuntos de datos proporcionan información relevante sobre el entorno físico, social y económico en el que se realizan las obras públicas. Ayudan a enriquecer la interpretación de las predicciones y a analizar los posibles impactos. Las fuentes incluyen límites administrativos, datos censales, información catastral, indicadores de ingresos familiares y datos geoespaciales abiertos.

Figura 1. Resumen de las fuentes



Fuente: Elaboración propia, 2025.

Todas estas fuentes estaban interconectadas con los registros de descargas de Inkolan, tal y como se muestra en la figura 1. La tabla 1 resume las diferentes fuentes de datos utilizadas en este estudio, junto con sus principales características y capacidades de geolocalización.

Tabla 1. Resumen de las fuentes de datos utilizadas en el estudio.

Fuente	Categoría	Tipo de datos	Atributos clave	Geo-rreferenciado
Registros de descarga de Inkolan	Datos privados	Metadatos sobre solicitudes de infraestructura digital	Identidad del usuario, tipo de usuario, marca de tiempo, finalidad, indicador de relación con el trabajo, fecha prevista de trabajo, perfil de la empresa, geovalla	Sí
Ayuntamiento / Opendata	Obras terminadas en datos de espacios públicos	Permisos y registros de obras	Tipo de obra, fecha de ejecución, ubicación	No
Instituto Nacional de Estadística	Datos contextuales urbanos	Censo de población (2021)	Demografía, densidad de población y distribución por edades según el censo	Sí
Datos catastrales	Datos contextuales urbanos	Información sobre edificios	Uso del edificio, superficie, año de construcción, número de viviendas	Sí
Atlas de ingresos familiares	Datos contextuales urbanos	Indicadores socioeconómicos	Ingresos familiares por distrito censal	Sí
OpenStreetMap (OSM)	Datos contextuales urbanos	Datos geoespaciales abiertos de infraestructura y servicios urbanos	Actividades económicas, tipo de edificios	Sí
Límites municipales	Datos contextuales urbanos	Límites administrativos	Nombres, códigos, límites de distritos y barrios	Sí

Source: Own elaboration, 2025.

3.3. Metodología

La metodología adoptada en este estudio se enmarca dentro de un enfoque de métodos mixtos para la investigación en ciencia de datos urbanos, tal y como propone Batty (2013) en The New Science of Cities. Este enfoque reconoce la necesidad de combinar la recopilación y el análisis de grandes volúmenes de datos urbanos con la contextualización cualitativa y la participación de diferentes actores urbanos. De este modo, facilita una comprensión más profunda y realista de los fenómenos urbanos, incorporando tanto la dimensión técnica como la social. La integración de datos cuantitativos georreferenciados con los conocimientos aportados por las partes interesadas permite diseñar modelos predictivos y soluciones de inteligencia artificial más adecuados a la complejidad y heterogeneidad del entorno urbano.

La metodología seguida consta de varios pasos, tal y como se presenta en la figura 2 y se explica a continuación.

Figura 2. Metodología



Fuente: Elaboración propia, 2025.

En primer lugar, se lleva a cabo un ejercicio de recopilación de datos. Se contacta con Inkolan para que proporcione datos sobre registros de descargas y metadatos asociados. Paralelamente, se establecen colaboraciones con los ayuntamientos de Zaragoza y Ermua para obtener datos sobre obras en espacios públicos. Sin embargo, debido a problemas de calidad y exhaustividad de los datos, el proyecto acaba basándose en datos abiertos de la ciudad de Madrid. Se utilizan conjuntos de datos adicionales disponibles públicamente del Instituto Nacional de Estadística y del Catastro Nacional para contextualizar los registros de descargas y los datos sobre obras, lo que pone de relieve la importancia de las fuentes de datos abiertos para superar las limitaciones y garantizar la disponibilidad de los datos.

En segundo lugar, toda la información recopilada se preprocesa georreferenciando los datos que carecen de coordenadas precisas a través de la API de Google Maps e interrelacionando los datos para proporcionar contexto mediante geoalgoritmos desarrollados a través de FME Form (una herramienta ETL geoespacial para tareas de procesamiento e integración de datos). Por ejemplo, las descargas de Inkolan se vinculan a su barrio correspondiente y se enriquecen con información contextual, como datos sobre edificios y población. Las obras públicas también se localizan por barrio y se relacionan espacialmente con las descargas de Inkolan. Este cuidadoso preprocesamiento es crucial para abordar las inconsistencias de los datos y garantizar la alta calidad, la interrelación y la geolocalización del conjunto de datos necesarios para el rendimiento eficaz de los modelos de IA, especialmente teniendo en cuenta los retos comunes de la heterogeneidad y la incompletitud de los datos en las aplicaciones del mundo real.

En tercer lugar, se lleva a cabo un análisis exploratorio exhaustivo de los datos para identificar patrones y tendencias dentro del conjunto de datos de Inkolan. Además, se realizan análisis para examinar las relaciones entre los datos de descargas de Inkolan y los datos de obras públicas, incluidas las correlaciones espaciales y temporales, con el fin de identificar posibles asociaciones y dependencias para los modelos predictivos.

En cuarto lugar, basándose en el análisis anterior, se desarrolló un modelo predictivo ARIMA para pronosticar las descargas de Inkolan utilizando patrones históricos. Este algoritmo utiliza las frecuencias de descarga recientes para estimar el número de descargas futuras que se espera que se produzcan. Los parámetros del modelo (p , q , d) se estimaron mediante un proceso de optimización que selecciona automáticamente los mejores valores basándose en criterios como el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC), lo que garantiza el equilibrio entre la precisión y la generalización. Además, se desarrolla un panel de control completo en Power BI para facilitar la visualización de patrones y tendencias en la relación entre las descargas, las obras públicas y sus factores contextuales asociados.

En quinto lugar, se realiza una síntesis de los resultados, que ofrece una visión general completa de los resultados del estudio, las limitaciones encontradas y el potencial de desarrollo y aplicación futuros.

Como paso final y con el objetivo de contrastar las conclusiones obtenidas, se llevó a cabo una serie de entrevistas con las partes interesadas pertinentes para comprender mejor la relación entre las descargas de Inkolan, la solicitud de licencias de construcción y la posterior ejecución de las obras. Estas entrevistas sirvieron para validar los resultados y evaluar la viabilidad de continuar la investigación sobre la predicción de las intervenciones en el espacio público.

4. Resultados

4.1. Recopilación de datos: lagunas y carencias

Como se menciona en la metodología, la disponibilidad de datos relacionados con licencias de obras públicas o registros de construcción se contrastó con los municipios de Zaragoza y Ermua. Algunas de las dificultades encontradas en este proceso incluyen: la complejidad de identificar interlocutores organizativos adecuados, la diversidad de formatos de datos (en algunos casos, a pesar de estar digitalizados, la información relevante se encontraba dispersa en diferentes documentos PDF adjuntos a un expediente de construcción) y, lo más relevante y común a ambos, el volumen insuficiente de registros que dificulta la aplicación efectiva de algoritmos de IA (entre 100 y 500 registros totales, dependiendo del registro consultado).

Ante este hecho, se analizaron los datos disponibles junto con conjuntos de datos públicos que pudieran proporcionar un volumen suficiente para apoyar la implementación de modelos de IA, identificando una oportunidad prometedora en la ciudad de Madrid. Se seleccionó un conjunto de datos de permisos de construcción completados llevados a cabo en el dominio público, que abarcaba obras de alcantarillado, reparaciones de emergencia, canalizaciones y excavaciones programadas entre 2017 y 2021. Este conjunto de datos contenía más de 55 000 registros, incluyendo el mes y el año de finalización, el tipo de obra y una dirección postal de referencia para cada proyecto. Sin embargo, carecía de información adicional relevante, como la fecha de inicio, la duración, el alcance de la obra o la identidad del diseñador o promotor involucrado.

Si bien el volumen de datos era adecuado para el análisis exploratorio y el desarrollo inicial del modelo, esto pone de relieve la necesidad de equilibrar la cantidad y la calidad de los datos, así como la importancia de integrar múltiples fuentes de datos para enriquecer la comprensión contextual y mejorar la precisión de los modelos predictivos. Mediante la combinación de conjuntos de datos de diferentes orígenes —como datos abiertos de Madrid, registros municipales y fuentes de terceros como Inkolan—, este estudio tenía por objeto superar las limitaciones de los conjuntos de datos individuales y mejorar la solidez del análisis global.

4.2. Preprocesamiento de datos

El preprocesamiento de los datos de obras públicas consistió en adaptar y unificar archivos de diferentes años en un único archivo, geocodificar los registros basándose en la dirección postal indicada para cada obra y, posteriormente, limpiar los datos para descartar los registros sin georreferenciación asignada o mal geolocalizados (los que habían sido geolocalizados fuera del límite municipal). Además, la base de datos de obras se relacionó espacialmente con la base de datos de descargas de Inkolan para identificar las obras contenidas dentro del perímetro de una o varias descargas.

Por otra parte, las descargas de Inkolan, además de relacionarse con los proyectos, también se relacionaron con información contextual para identificar los barrios en los que se producen, así como con algunos indicadores contextuales relacionados con la población potencialmente afectada en caso de un proyecto de construcción dentro del perímetro de descarga, los equipos sensibles afectados, la información sobre el nivel de ingresos de los residentes o las actividades económicas que pueden verse afectadas. El objetivo de incluir estos datos era enriquecer los análisis posteriores y proporcionar contexto para la visualización de los resultados.

4.3. Principales conclusiones del análisis de datos

Tras el preprocesamiento de los datos, se obtuvieron dos conjuntos de datos completos: uno con obras públicas y otro con descargas de Inkolan, ambos enriquecidos con atributos espaciales y temporales.

Sin embargo, como se ha mencionado anteriormente, no existe ningún campo común entre las dos bases de datos que permita establecer una relación uno a uno, es decir, que vincule una obra determinada con una descarga específica. Por lo tanto, el primer paso consistió en definir criterios para aproximar esta asociación uno a uno entre obras y descargas.

Inicialmente, cada descarga se vinculó a todas las obras cuyas geolocalizaciones se encontraban dentro del polígono especificado en el registro de descarga correspondiente, ya que cualquier otra asociación habría carecido de coherencia lógica. Posteriormente, se calculó la distancia euclidiana entre el centroide del polígono y las ubicaciones de las obras, generando para cada descarga una lista clasificada de obras potencialmente relacionadas en función de la proximidad espacial.

Además, el conjunto de datos de descargas incluía la fecha de descarga, junto con las estimaciones proporcionadas por los usuarios sobre las fechas de inicio y finalización de las obras. Es importante destacar que solo la fecha de descarga era totalmente fiable. Por el contrario, el conjunto de datos de obras públicas proporcionaba información verificada sobre las fechas de finalización de los proyectos. Por esta razón, se introdujo un segundo criterio, basado en la proximidad temporal, para refinar el proceso de asociación. Esta proximidad se calculó restando la fecha de descarga de la fecha de finalización de la obra.

No obstante, aún no se había establecido una relación uno a uno, que vinculara una sola obra con una sola descarga. Para solucionar este problema, se formularon algunas hipótesis que sirvieran de guía para la creación de relaciones uno a uno. Esto implicó definir umbrales para la diferencia de tiempo máxima permitida entre la fecha de finalización de una obra y la fecha de descarga correspondiente, así como para la distancia euclidiana máxima permitida entre sus centroides.

Una relación uno a uno se produce cuando una obra se encuentra dentro de un único polígono de descarga y ese polígono no contiene otras obras. Sin embargo, esto es poco común dado el tamaño de los conjuntos de datos: aproximadamente 50 000 obras en 4 años frente a 40 000 descargas en una década. Por lo tanto, se analizaron tres escenarios para definir estos criterios:

1. Descargas con una sola obra dentro del polígono: una muestra de 4154 registros mostró que dichas descargas se localizaban a una distancia máxima de 75 metros de la obra asociada y solían producirse hasta 1500 días antes de la finalización de la obra.

2. Obras dentro de un único polígono: de las 1827 obras analizadas, no surgieron criterios de filtrado adicionales, pero los umbrales de distancia y tiempo del caso 1 se cumplieron en general.
3. Coincidencias estrictas uno a uno: solo 45 casos cumplían la condición estricta de una obra dentro de un polígono, sin solapamientos en otros lugares. Las distancias rara vez superan los 100 m (en su mayoría por debajo de los 75 m), mientras que las diferencias de tiempo varían ampliamente entre -2000 y 3000 días. Debido al pequeño tamaño de la muestra, no se consideró que este patrón fuera generalizable. Sin embargo, combinando esto con el caso 1, el umbral de tiempo se ajustó a [-800, 1500] días para tener en cuenta las diferencias de fecha negativas observadas en estas coincidencias.

Como conclusión tras este análisis, se determinó que una obra se asociaría con una descarga si se encontraba a menos de 75 metros del centroide de la zona y tenía una diferencia de fecha entre -800 y 1500 días.

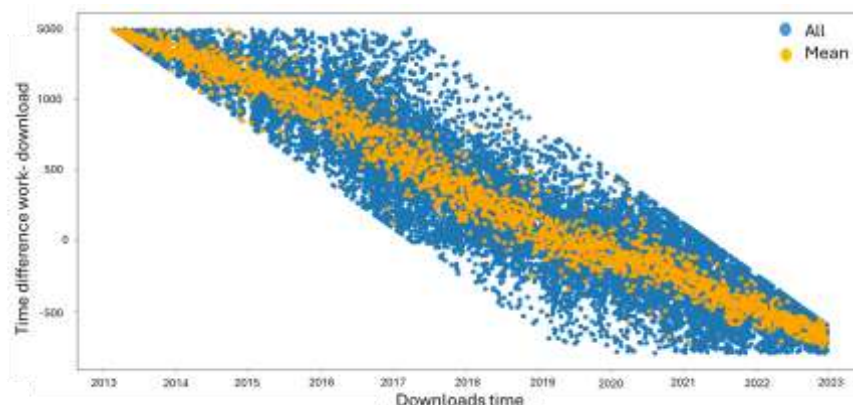
4.4. Resultados del modelado y la visualización

Una vez preprocesados los datos y definidas las relaciones entre ellos, se siguieron dos enfoques: identificar las obras asociadas a cada descarga y predecir futuras descargas. En las siguientes subsecciones se explican estos modelos en detalle.

4.4.1. Patrones de relación entre obras y descargas

A partir del análisis anterior, se determinó que, en general, una obra puede completarse entre 800 días antes y 1500 días después de que se realice una descarga. Utilizando esta información, se perfeccionó el proceso de emparejamiento entre descargas y obras, y se examinó la evolución temporal de las descargas en comparación con la diferencia de tiempo con respecto a las obras asociadas. La figura 3 muestra, para cada fecha de descarga, la diferencia en días entre esa fecha de descarga y la fecha de finalización de las obras asociadas temporal y espacialmente.

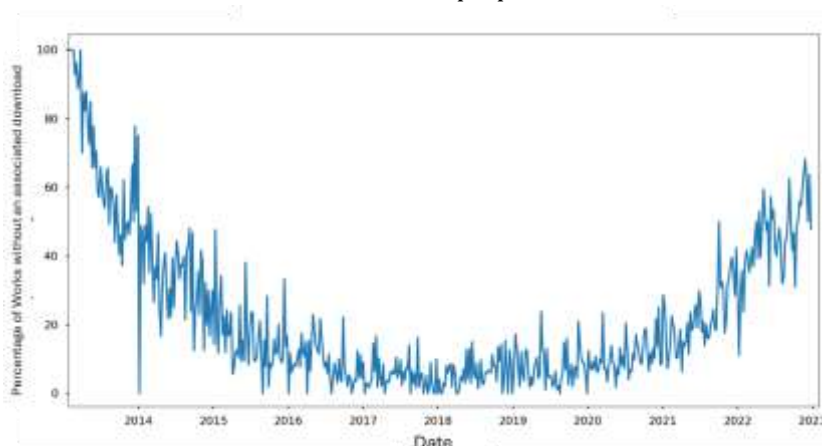
Figura 3. Fechas de descarga en comparación con el tiempo restante hasta la finalización del trabajo



Fuente: Elaboración propia, 2025.

El análisis reveló una tendencia lineal: con el tiempo, la diferencia entre la fecha de descarga y la fecha de finalización del proyecto correspondiente disminuye. Esto se debe a un sesgo en los datos: como se mencionó anteriormente, las descargas abarcan desde 2010 hasta 2022, mientras que las finalizaciones de proyectos solo se registran entre 2017 y 2021, lo que da lugar a asociaciones más frecuentes entre descargas recientes y proyectos pasados. Por consiguiente, aún no se puede confirmar una relación directa. No obstante, al aplicar filtros espaciales y temporales, algunas descargas no tenían obras coincidentes, lo que permitió calcular el porcentaje de descargas sin coincidencia a lo largo del tiempo, como se ilustra en la figura 4.

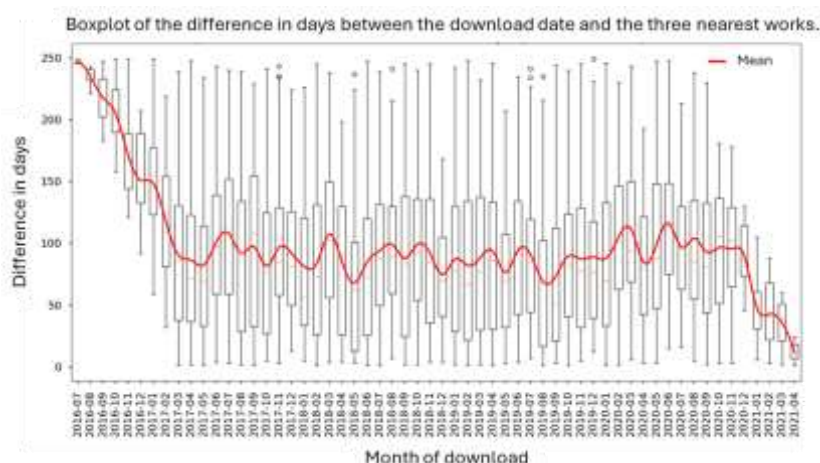
Figura 4. Porcentaje de descargas sin una obra asociada
Fuente: Elaboración propia, 2025.



Para seguir perfeccionando la relación entre el trabajo y la descarga, se seleccionaron las tres obras más cercanas para cada descarga y se calculó el tiempo medio hasta las fechas de finalización de las obras. Esto proporcionó un valor de «días hasta la finalización» para cada descarga. Los resultados siguieron mostrando una tendencia a la baja debido a los diferentes rangos de fechas entre los conjuntos de datos, lo que llevó a la aplicación de un filtrado más estricto. Sin embargo, dado que el tipo de descarga afecta al tiempo, los datos se dividieron en descargas relacionadas con el trabajo y descargas relacionadas con el proyecto, que son características especificadas dentro del propio conjunto de datos. El análisis de frecuencia indicó lo siguiente:

- Las descargas de trabajo suelen producirse entre 0 y 250 días antes del inicio del proyecto, con centroides dentro de los 75 m. Como se muestra en la figura 5, la serie se mantuvo estacionaria entre 2017 y 2021, con un margen de error intercuartílico de aproximadamente 100 días (unos 3 meses), lo que sugiere coincidencias uno a uno razonablemente fiables.

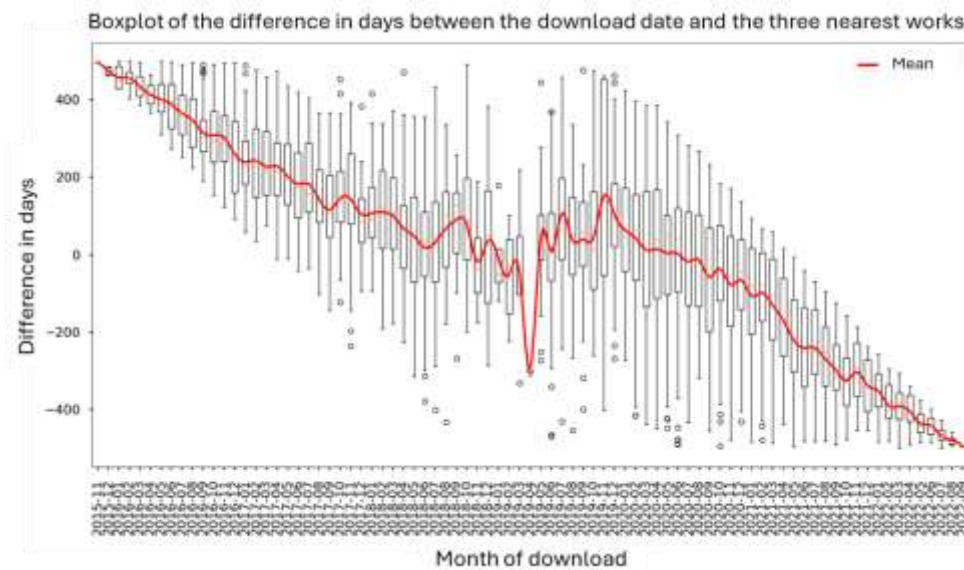
Figura 5. Diagrama de caja de la diferencia en días entre las fechas de las descargas de tipo laboral y las fechas de las tres obras más cercanas.



Fuente: Elaboración propia, 2025.

- Descargas de proyectos: suelen producirse entre 500 días antes y 500 días después del inicio del proyecto, también en un plazo de 75 minutos. Sin embargo, no se observó un patrón temporal estable (véase la figura 6), lo que hace que la correspondencia uno a uno no sea fiable. Con más datos sobre la finalización de proyectos y rangos intercuartílicos más estrechos, en el futuro se podrían desarrollar modelos predictivos o autorregresivos.

Figura 6. Diagrama de caja de la diferencia en días entre las fechas de las descargas por tipo de proyecto y las fechas de las tres obras más cercanas.



Fuente: Elaboración propia, 2025.

4.4.2. Descargar modelo predictivo

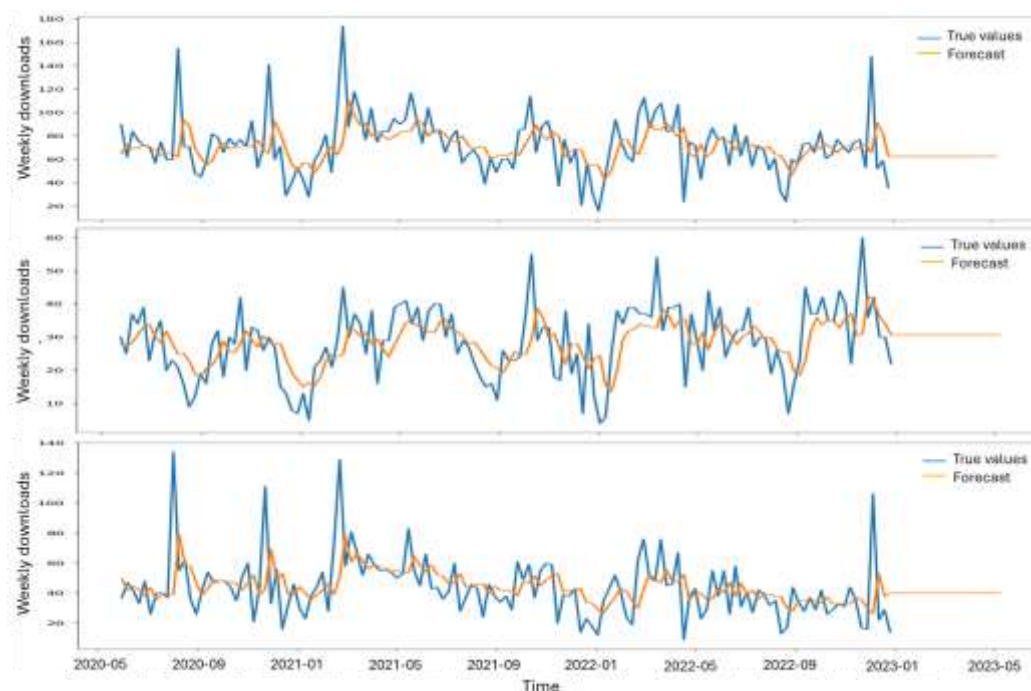
El objetivo de este segundo enfoque de modelización era pronosticar el número futuro de descargas de Inkolan utilizando datos históricos. Para ello, se desarrolló un modelo ARIMA (media móvil autorregresiva integrada) (Shumway y Stoffer, 2005). Este modelo de series temporales capturó la dinámica temporal de la actividad de descarga y demostró su capacidad para generar previsiones de valores futuros basadas en información de puntos temporales anteriores.

Se realizaron tres experimentos para evaluar el rendimiento predictivo del modelo. En primer lugar, se analizaron y pronosticaron las tendencias temporales de todo el conjunto de datos de descargas. En segundo lugar, debido a las diferencias significativas en las frecuencias de descarga entre las descargas de tipo proyecto y las de tipo trabajo, el conjunto de datos se estratificó en estas dos categorías y se evaluaron modelos predictivos separados para cada una de ellas. En todos los experimentos, el conjunto de datos se dividió en subconjuntos de entrenamiento y prueba, utilizando el 80 % de los datos para el entrenamiento del modelo y reservando el 20 % para la evaluación fuera de la muestra.

Los resultados (véase la figura 7) indicaron una gran precisión predictiva para los valores de descarga intermedios. Sin embargo, debido a la naturaleza de media móvil del modelo ARIMA, no fue capaz de pronosticar con precisión los picos extremos —casos en los que el recuento de descargas se desviaba significativamente de la media y la desviación estándar— con un RMSE de 23,95. Además, la precisión predictiva mejoró al modelar subconjuntos de datos más específicos, con valores RMSE de 9,21 para las obras y 20,30 para los proyectos.

Por último, el conjunto de datos de obras se desagregó aún más por tipo, una característica que también está presente en los datos. Como era de esperar, el aumento de la especificidad mejoró la precisión de la previsión. No obstante, el trabajo con subconjuntos más pequeños dio lugar a una reducción del tamaño de las muestras para el entrenamiento, lo que podría limitar la capacidad del modelo para generalizar a datos no vistos.

Figura 7. Predicción de descargas futuras utilizando el modelo ARIMA para el conjunto de datos completo y, por separado, para las descargas por tipo de obra y por tipo de proyecto.

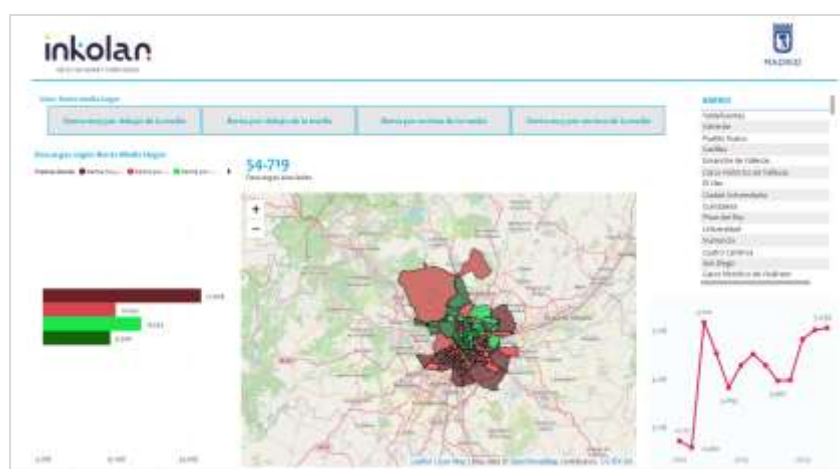


Fuente: Elaboración propia, 2025.

4.4.3. Panel de visualización

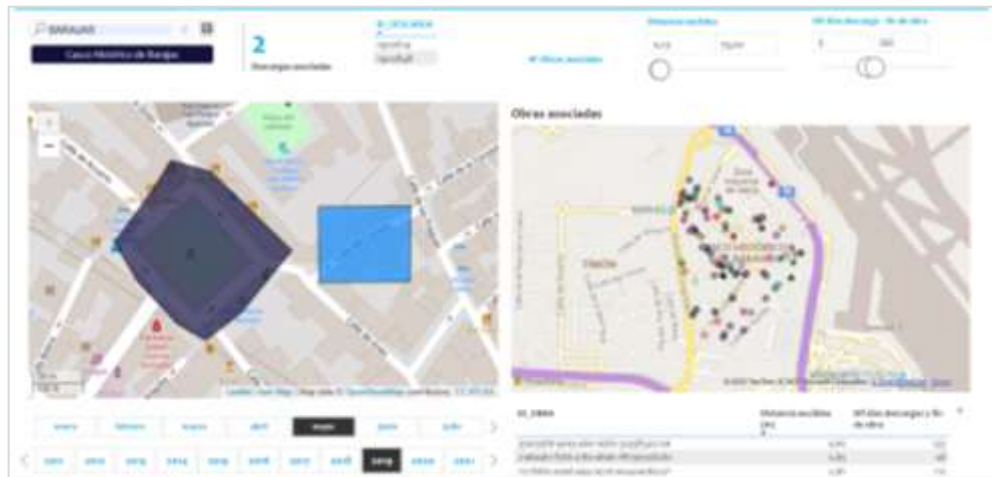
Para la visualización de los resultados, tanto los datos de construcción como los datos de descargas se integran con la información contextual asociada y las conclusiones obtenidas sobre los patrones de relación entre las obras y las descargas en un panel desarrollado con PowerBI. El panel consta de varias pestañas a través de las cuales se puede analizar por separado la información sobre las descargas de Inkolan y la construcción de datos abiertos, incorporando diferentes posibilidades para relacionar la información de las descargas de Inkolan con su contexto (para analizar la influencia de los ingresos de los habitantes o el período de construcción de los barrios de la ciudad en las descargas, véase la figura 8) y una pestaña específica que permite relacionar las descargas de Inkolan con sus posibles obras en función de los parámetros definidos a partir del análisis exploratorio (véase la figura 9).

Figura 8. Ejemplo de pestaña del panel de control que relaciona los ingresos de los habitantes con las descargas de Inkolan por barrio.



Fuente: Elaboración propia, 2025.

Figura 9. Ejemplo de pestaña del panel de control que vincula las descargas de Inkolan con los trabajos de datos abiertos.



Fuente: Elaboración propia, 2025.

4.5. Síntesis de resultados

Sin embargo, no se pudo consolidar un modelo predictivo de obras públicas basado en las descargas de Inkolan. Por lo tanto, se avanzó en la relación de cada descarga con una obra en función de criterios de fecha y posición. En caso de existir una relación directa (uno a uno) entre descargas y obras, sería posible determinar la diferencia temporal esperada para futuras descargas. En este sentido, debido a la falta de esta relación directa, se formularon las siguientes hipótesis:

1. Una obra se asocia a una descarga si se encuentra dentro del polígono, con una distancia desde el centroide inferior a 75 m y con una diferencia temporal con respecto a la descarga de entre -800 y 1500 días. Esto sugiere que la obra se completa entre 2 años antes de la descarga y 4 años después.
2. Para las descargas de tipo «obra», se observó que las descargas se producen en un intervalo de 0 a 250 días antes de la fecha de inicio de la obra. Además, se constató que los centroides deben estar a menos de 75 m entre sí.
3. Por el contrario, para las descargas de tipo «proyecto», las descargas se producen en un intervalo de 500 días antes a 500 días después de la fecha de inicio del proyecto. Del mismo modo, se descubrió que los centroides deben estar a menos de 75 m entre sí.

Tras estas aproximaciones de emparejamiento, se analizó la diferencia en días entre la descarga y los posibles trabajos asociados. En resumen, se obtuvieron los siguientes resultados:

- Con un campo común entre las bases de datos de descargas y trabajos, se puede definir una relación uno a uno, lo que permitiría predecir cuándo se completará el trabajo relacionado con la descarga.
- Se conoce la proporción de descargas que no darán lugar a obras, basándose en las fechas y distancias máximas definidas.
- Se proporciona un intervalo de tiempo para cuando se espera que se complete la obra con una fecha de descarga.

Estos resultados sugieren que, aunque no se pudo consolidar un modelo predictivo de obras públicas basado en las descargas de Inkolan, se lograron avances significativos en la relación entre las descargas y las obras. Estos hallazgos pueden ser útiles para mejorar la planificación y la gestión de las obras públicas.

4.6. Contraste con las partes interesadas

Dados los resultados obtenidos al intentar relacionar el archivo de descargas de Inkolan con el archivo de obras, se consideró necesario realizar una serie de entrevistas con agentes clave para comprender mejor la relación entre la descarga y la obra, el proceso desde la solicitud de la licencia hasta la ejecución de la obra, los agentes implicados y las posibilidades que esto puede abrir para la continuación de la investigación en este campo.

Se realizaron entrevistas de contraste con departamentos de gestión de licencias de obras de dos ayuntamientos (Bilbao y Madrid) y con responsables de la ejecución de obras de socios de Inkolan en diferentes territorios de España. Además, se revisan las normas de gestión de obras de 10 municipios españoles relacionados con las áreas en las que operan los agentes entrevistados.

Los resultados del análisis proporcionaron algunas conclusiones relevantes. En primer lugar, cabe mencionar que no existe un procedimiento único para la concesión de licencias de obras en España, ya que difieren de un municipio a otro. En segundo lugar, el plazo entre la solicitud de la licencia y la fecha de concesión depende de muchas variables, entre ellas la agilidad del municipio y, en función de los efectos de la obra, puede requerir la concesión de otras licencias/autorizaciones vinculadas al proyecto (costera, de carreteras, provincial, industrial, etc.). Por lo general, el ayuntamiento espera a que el solicitante tenga todos los permisos necesarios antes de conceder la licencia. No existen autorizaciones integradas ni ventanillas únicas. Los permisos deben solicitarse a cada organismo, lo que puede llevar hasta un año o más (dependiendo del organismo, el municipio, etc.). En tercer lugar, actualmente no existe un identificador común que permita asociar una descarga de Inkolan con un registro municipal de obras. En cuarto lugar, en el proceso de descarga-obra intervienen varios agentes y, en ocasiones, el estudio o la preparación de la documentación para la solicitud de la licencia de obras se subcontrata a una empresa (lo que suele implicar la descarga de Inkolan para realizar los estudios de impacto), lo que significa que la entidad que solicita la licencia puede no ser la misma que la que realizó la descarga. En ocasiones, pueden existir tres entidades con diferentes funciones (diseñador, promotor y contratista) vinculadas al proceso de descarga y obra, y no siempre es el diseñador quien realiza la descarga. Se dan situaciones en las que la descarga se realiza posteriormente para comprobaciones o incluso en diferentes momentos del proceso, y/o por diferentes agentes.

5. Discusión

El estudio destaca los retos críticos que plantea la aplicación de modelos de IA para predecir las obras públicas basándose en los datos disponibles, y hace hincapié en la exhaustividad de los datos como la limitación más significativa. La ausencia de identificadores comunes que vincularan las descargas de Inkolan con los registros municipales dificultó el establecimiento de relaciones uno a uno, lo que requirió el uso de criterios espaciales y temporales para asociar los datos. Aunque las demás dimensiones de calidad (precisión, relevancia, puntualidad y coherencia) se abordaron en general mejor, también presentaron limitaciones, como errores de geolocalización, discrepancias temporales entre los conjuntos de datos y la falta de estructuras estandarizadas. Estas deficiencias afectan a la eficacia de los modelos y análisis predictivos. Los resultados subrayan que no basta con disponer de grandes volúmenes de datos, sino que es importante garantizar que los datos disponibles sean de alta calidad. Sin embargo, teniendo en cuenta que la transparencia y la equidad son elementos clave en las decisiones urbanísticas impulsadas por la IA, incluso los conjuntos de datos precisos y coherentes pueden perpetuar las desigualdades sistémicas si reflejan sesgos históricos. Las consultas a las partes interesadas y el análisis normativo confirman la dificultad de obtener datos de calidad en el contexto de la planificación urbana. Este trabajo ofrece un enfoque inicial para abordar estos retos y orienta las futuras mejoras en la integración y la calidad de los datos.

Como se destaca en la revisión de la literatura, una limitación clave en la planificación urbana es la dependencia de conjuntos de datos existentes que no fueron diseñados originalmente con fines analíticos o predictivos. En consecuencia, estos conjuntos de datos a menudo no cumplen los requisitos de la investigación contemporánea y las aplicaciones tecnológicas. Este estudio subraya la necesidad de generar datos específicamente adaptados a los objetivos de la investigación, lo que requiere la colaboración entre múltiples partes interesadas, un principio que ya se ha destacado como esencial en la gobernanza y la planificación urbanas.

En cuanto a los retos identificados por las administraciones públicas en la producción de datos, este documento propone una serie de recomendaciones para facilitar la integración de algoritmos de IA en futuras iniciativas de planificación urbana. En primer lugar, es esencial digitalizar los datos desde el principio, asegurándose de que estén estructurados teniendo en cuenta su uso analítico futuro. En particular, la información sobre los permisos de construcción y su seguimiento debe recopilarse en formatos digitales que permitan su exportación sin problemas a bases de datos. Los esfuerzos de digitalización también deben dar prioridad a la transparencia, por ejemplo, documentando las fuentes, las limitaciones y los posibles sesgos de los conjuntos de datos. Esto se ajusta a los principios de la IA

explicable (XAI), que hacen hincapié en que los procesos de toma de decisiones algorítmicas sean accesibles para las partes interesadas sin conocimientos técnicos.

Dada la complejidad del proceso de concesión de permisos y construcción, es fundamental interrelacionar sistemáticamente las distintas etapas, incluidos los procedimientos, las fechas, los incidentes, los retrasos (por ejemplo, solicitudes de modificaciones o información adicional), los resultados (por ejemplo, denegaciones, aprobaciones u obras no ejecutadas), la magnitud de la obra (presupuesto, duración estimada) y la geolocalización del emplazamiento (con coordenadas y, preferiblemente, un perímetro definido). Además, es importante registrar las funciones de las partes interesadas implicadas en cada etapa. Este enfoque estructurado puede ayudar a identificar patrones de desigualdad, como retrasos o rechazos desproporcionados en barrios marginados, lo que puede indicar sesgos algorítmicos o procedimentales. Además, documentar las funciones de las partes interesadas mejora la rendición de cuentas al aclarar quién es responsable de la calidad de los datos y de la toma de decisiones en cada paso.

La estructuración y el formato adecuados de estos datos permitirían el desarrollo de una base de datos sólida sobre obras públicas, lo que facilitaría la identificación de hitos clave en el proceso, desde la solicitud del permiso hasta la ejecución. Además, dado que la presentación de planes de impacto de la red de servicios suele ser obligatoria, sería beneficioso registrar sistemáticamente si dichos planes se incluyen en la solicitud del permiso y, en caso afirmativo, anotar sus códigos de referencia. Esto favorecería la integración de los registros públicos con los datos descargados de Inkolan. Sin embargo, esta integración debe ir acompañada de salvaguardias para evitar el refuerzo de los sesgos. Por ejemplo, si los modelos de IA dan prioridad a los permisos en zonas con valores inmobiliarios más elevados debido a la disponibilidad de datos, los planificadores deben auditar activamente estos resultados para garantizar una asignación equitativa de los recursos.

Un requisito fundamental para el desarrollo de un algoritmo predictivo es la disponibilidad de un conjunto de datos que vincule claramente las obras con sus descargas asociadas, lo que sirve como referencia para perfeccionar los patrones de relación y entrenar el algoritmo. Esto podría permitir la predicción de futuras solicitudes de permisos basadas en los datos de descarga, la estimación de la duración de las obras y el análisis de si una descarga determinada dará lugar a una obra real, en función de sus características. Sin embargo, las implicaciones éticas de tales predicciones son significativas.

Sin embargo, debido a la complejidad del proceso y a la variabilidad de los plazos, influidos por factores como la rapidez de tramitación municipal, la magnitud y el tipo de impacto de la obra, la validez del algoritmo dependería del contexto específico. Su precisión puede variar en función de la naturaleza de la obra (por ejemplo, su magnitud o impacto), y su aplicación a otros contextos requeriría una adaptación basada en datos relevantes a nivel local. Esta especificidad contextual pone de relieve la importancia de involucrar de forma proactiva a las partes interesadas locales en la validación de los resultados para mitigar los posibles sesgos algorítmicos.

6. Limitaciones del estudio y orientaciones para futuras investigaciones

Este estudio presenta varias limitaciones que deben tenerse en cuenta. En primer lugar, la escasez de datos públicos disponibles sobre licencias de obras públicas y registros de construcción limitó considerablemente el alcance del análisis. En segundo lugar, la ausencia de procedimientos estandarizados para la concesión de licencias de obras en los municipios españoles introdujo variabilidad en la recopilación e interpretación de datos. Esta falta de armonización complica las comparaciones entre jurisdicciones y limita el potencial de soluciones escalables. Por último, la naturaleza fragmentada actual de los procesos de concesión de permisos, que requieren autorizaciones separadas de múltiples organismos, introduce complejidad en el seguimiento y el análisis de los flujos de trabajo, ya que no existe un sistema centralizado o integrado para supervisar el progreso o los resultados.

Para abordar estas limitaciones, las investigaciones futuras podrían centrarse en las siguientes áreas:

1. Estandarización de la recopilación de datos: Establecer marcos unificados para recopilar y compartir datos de obras públicas entre municipios, incluyendo identificadores comunes para vincular sistemas dispares (por ejemplo, registros municipales y bases de datos externas).

2. Aprovechamiento de fuentes de datos indirectas: Explorar fuentes de datos alternativas (por ejemplo, licitaciones públicas, registros de adquisiciones) como sustitutos de las licencias de obras de construcción.
3. Mejora de la infraestructura de datos: investigar plataformas digitales centralizadas o «ventanillas únicas» para agilizar las solicitudes de permisos y reducir la fragmentación administrativa. Los proyectos piloto podrían poner a prueba flujos de trabajo integrados y evaluar su impacto en los tiempos de tramitación y la transparencia.
4. IA y análisis predictivo: explorar modelos de IA para predecir retrasos, identificar cuellos de botella en los procedimientos o detectar patrones de desigualdad (por ejemplo, rechazos desproporcionados en barrios marginados). La ampliación de los conjuntos de datos mediante la colaboración municipal y la generación de datos sintéticos, en consonancia con los principios de la IA explicable (XAI), mejoraría la transparencia y la confianza.
5. Análisis de políticas y gobernanza: abordar las barreras sociopolíticas para la armonización, como la resistencia de las partes interesadas, las discrepancias legales y las disparidades de recursos. Las reformas políticas podrían incentivar la colaboración entre organismos y establecer mecanismos de rendición de cuentas para la calidad de los datos y la toma de decisiones.

Al abordar estas limitaciones y ampliar el alcance de la investigación, los estudios futuros pueden contribuir a crear sistemas de gestión de obras públicas más equitativos, eficientes y basados en datos.

7. Conclusiones

Este estudio ha puesto de manifiesto la complejidad del proceso de concesión de permisos de obras públicas en España, así como la ausencia de un procedimiento unificado y estandarizado en todas las jurisdicciones. Además, ha subrayado la importancia fundamental de la colaboración interinstitucional entre todas las partes implicadas en el proceso. Asimismo, se ha destacado la necesidad de generar datos específicos para cada finalidad, haciendo especial hincapié en la necesidad de establecer un vínculo sistemático y fiable entre los registros de descargas de Inkolan y los registros municipales de obras públicas.

El desarrollo y la aplicación de la inteligencia artificial (IA) en el contexto de las obras públicas y la planificación urbana han demostrado un gran potencial. Sin embargo, el éxito de estas aplicaciones depende en gran medida de la calidad y la disponibilidad de los datos. Del análisis realizado en este estudio se desprende que los datos necesarios para la implementación de la IA deben cumplir con determinados requisitos de calidad, como la precisión, la pertinencia, la exhaustividad, la puntualidad y la coherencia. Estos requisitos son esenciales para garantizar que los datos utilizados para entrenar los modelos de IA sean fiables y representativos de los escenarios del mundo real que pretenden replicar o predecir.

Una de las principales conclusiones de este estudio es la necesidad de disponer de datos estructurados y digitalizados desde el principio. La información relacionada con los permisos de construcción y su seguimiento debe recopilarse en formatos digitales que permitan una integración perfecta en las bases de datos. Este enfoque estructurado facilita la identificación de los hitos clave del proceso, desde la solicitud del permiso hasta su ejecución, y favorece el desarrollo de bases de datos sólidas sobre obras públicas.

Además, el estudio destaca la importancia de registrar sistemáticamente información relevante, como los planes de impacto de la red de servicios, las funciones de las partes interesadas y los datos de geolocalización. Estos detalles son cruciales para crear un conjunto de datos completo que pueda servir como base para entrenar algoritmos predictivos. La integración de los registros públicos con datos de plataformas como Inkolan mejora aún más la utilidad y el alcance de los datos.

A pesar de los posibles beneficios, el estudio también identificó varios retos, entre ellos el volumen insuficiente de registros disponibles en los municipios de Zaragoza y Ermua. La escasez de datos dificultó la aplicación eficaz de los algoritmos de IA, lo que puso de relieve la necesidad de prácticas de recopilación de datos más amplias y estandarizadas.

En conclusión, aunque la IA tiene el potencial de revolucionar la planificación urbana y la gestión de las obras públicas, su implementación exitosa requiere un esfuerzo concertado para mejorar la calidad, la disponibilidad y la estandarización de los datos. Abordar estos retos será esencial para aprovechar todo el potencial de la IA y lograr un desarrollo urbano más eficiente y eficaz.

8. Agradecimientos

Los autores desean expresar su agradecimiento a INKOLAN y a sus socios, así como a los municipios y partes interesadas que han participado en este proceso.

Referencias

- Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, 9(1). <https://doi.org/10.1186/s41235-023-00529-3>
- Anwar, M. R., & Sakti, L. D. (2024). Integrating Artificial Intelligence and Environmental Science for Sustainable Urban Planning. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 5(2), 179-191. <https://doi.org/10.34306/itsdi.v5i2.666>
- Batty, M. (2013). *The New Science of Cities*. The MIT Press. <https://www.jstor.org/stable/j.ctt9qf7m6>
- Bibri, S. E., Huang, J., Jagatheesaperumal, S. K., & Krogstie, J. (2024). The synergistic interplay of artificial intelligence and digital twin in environmentally planning sustainable smart cities: A comprehensive systematic review. *Environmental Science and Ecotechnology*, 20, 100433. <https://doi.org/10.1016/j.esec.2024.100433>
- Chaturvedi, V., & de Vries, W. T. (2021). Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Science*, 5(3), 68. <https://doi.org/10.3390/urbansci5030068>
- Correia, C. R., & Roseland, M. (2022). Addressing Negative Externalities of Urban Development: Toward a More Sustainable Approach. *Urban Science*, 6(2), 38. <https://doi.org/10.3390/urbansci6020038>
- Damioli, G., Van Roy, V., Vertesy, D., & Vivarelli, M. (2025). Is artificial intelligence leading to a new technological paradigm? *Structural Change and Economic Dynamics*, 72, 347-359. <https://doi.org/10.1016/j.strueco.2024.12.006>
- Delgado Enales, I. (2024). *Advanced optimization and data modeling techniques to improve accesibility and thermal comfort in urban plannning*. [Tesis Doctoral, Universidad del País Vasco] <http://hdl.handle.net/10810/73521>
- Delgado-Enales, I., Molina-Costa, P., Osaba, E., Urrea-Uriarte, S., & Del Ser, J. (2022). Improving the Urban Accessibility of Older Pedestrians using Multi-objective Optimization. *2022 IEEE Congress on Evolutionary Computation (CEC)*, 1-8. <https://doi.org/10.1109/CEC55065.2022.9870432>
- D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press. <https://doi.org/10.7551/mitpress/11805.001.0001>
- Fernández Álvarez, L., & Garaizar Sagarminaga, P. (2024). *La inteligencia artificial explicada a todos los públicos (castellano)*. A Fin de Cuentos.
- Gao, R. X., Krüger, J., Merklein, M., Möhring, H.-C., & Váncza, J. (2024). Artificial Intelligence in manufacturing: State of the art, perspectives, and future directions. *CIRP Annals*, 73(2), 723-749. <https://doi.org/10.1016/j.cirp.2024.04.101>
- Garg, P. K. (2021). Overview of Artificial Intelligence. En L. Sharma y p. Kumar Garg (Eds.) *Artificial Intelligence*. Chapman and Hall/CRC.
- Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology*, 162, 107268. <https://doi.org/10.1016/j.infsof.2023.107268>
- Hamann, K. R. S., Bertel, M. P., Ryszawska, B., Lurger, B., Szymański, P., Rozwadowska, M., Goedkoop, F., Jans, L., Perlaviciute, G., Masson, T., Fritsche, I., Favaro, T., Hofer, A., Eisenberger, I., Gutschi, C., Grosche, C., Held, J., Athenstaedt, U., & Corcoran, K. (2023). An interdisciplinary understanding of energy citizenship: Integrating psychological, legal, and economic perspectives on a citizen-centred sustainable energy transition. *Energy Research & Social Science*, 97, 102959. <https://doi.org/10.1016/j.erss.2023.102959>
- He, W., & Chen, M. (2024). Advancing Urban Life: A Systematic Review of Emerging Technologies and Artificial Intelligence in Urban Design and Planning. *Buildings*, 14(3), 835. <https://doi.org/10.3390/buildings14030835>
- Korobeinikova, A., Danilina, N., & Teplova, I. (2024). Planning Public Space Climate Comfortability: A GIS-Based Algorithm for the Compact Cities of the Far North. *Land*, 13(11), 1763. <https://doi.org/10.3390/land13111763>
- Koutra, S., & Ioakimidis, C. S. (2023). Unveiling the Potential of Machine Learning Applications in Urban Planning Challenges. *Land*, 12(1), 83. <https://doi.org/10.3390/land12010083>
- Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. (2018). Artificial Intelligence in the 21st Century. *IEEE Access*, 6, 34403-34421. <https://doi.org/10.1109/ACCESS.2018.2819688>

- Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018). Brain Intelligence: Go beyond Artificial Intelligence. *Mobile Networks and Applications*, 23(2), 368-375. <https://doi.org/10.1007/s11036-017-0932-8>
- Mashhood, M., Salman, H., Amjad, R., & Nisar, H. (2023). The Advantages of Using Artificial Intelligence in Urban Planning – A Review of Literature. *JOURNAL OF STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH*, 5(2), 1-12. <https://doi.org/10.52700/scir.v5i2.125>
- Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2025). The effects of data quality on machine learning performance on tabular data. *Information Systems*, 132, 102549. <https://doi.org/10.1016/j.is.2025.102549>
- Molina-Costa, P. (2024). *La planificación urbana en la era de la digitalización*. In *Ciudad digital Conectividad y (des)aceleración*. CATARATA. <https://www.catarata.org/libro/ciudad-digital-150546/>
- Moore, G. E. (2006). Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Society Newsletter*, 11(3), 33-35. <https://doi.org/10.1109/N-SSC.2006.4785860>
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.
- Park, C., No, W., Choi, J., & Kim, Y. (2023). Development of an AI advisor for conceptual land use planning. *Cities*, 138, 104371. <https://doi.org/10.1016/j.cities.2023.104371>
- Pericault, Y., Viklander, M., & Hedström, A. (2023). Modelling the long-term sustainability impacts of coordination policies for urban infrastructure rehabilitation. *Water Research*, 236, 119912. <https://doi.org/10.1016/j.watres.2023.119912>
- Rjab, A. B., Mellouli, S., & Corbett, J. (2023). Barriers to artificial intelligence adoption in smart cities: A systematic literature review and research agenda. *Government Information Quarterly*, 40(3), 101814. <https://doi.org/10.1016/j.giq.2023.101814>
- Sanchez, T. W., Fu, X., Yigitcanlar, T., & Ye, X. (2024). The Research Landscape of AI in Urban Planning: A Topic Analysis of the Literature with ChatGPT. *Urban Science*, 8(4), 197. <https://doi.org/10.3390/urbansci8040197>
- Shulajkovska, M., Smerkol, M., Noveski, G., & Gams, M. (2024). Enhancing Urban Sustainability: Developing an Open-Source AI Framework for Smart Cities. *Smart Cities*, 7(5), 2670-2701. <https://doi.org/10.3390/smartcities7050104>
- Shumway, R. H., & Stoffer, D. S. (2005). *Time Series Analysis and Its Applications*. Springer.
- Urra-Uriarte, S., Herranz-Pascual, K., Molina-Costa, P., Martin Roncero, U., & Glodeanu, A. (2024). An Age-Friendly Neighbourhood Index as a Long-Term Urban Planning Decision-Making Tool. *Urban Science*, 8(3), 127. <https://doi.org/10.3390/urbansci8030127>
- Urrutia-Azcona, K., Usobiaga-Ferrer, E., De Agustín-Camacho, P., Molina-Costa, P., Benedito-Bordonau, M., & Flores-Abascal, I. (2021). ENER-BI: Integrating Energy and Spatial Data for Cities' Decarbonisation Planning. *Sustainability*, 13(1), 383. <https://doi.org/10.3390/su13010383>
- Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports*, 13(1), 15737. <https://doi.org/10.1038/s41598-023-42384-8>
- Villanueva-Merino, A., Urra-Uriarte, S., Izkara, J. L., Campos-Cordobes, S., Aranguren, A., & Molina-Costa, P. (2024). Leveraging Local Digital Twins for planning age-friendly urban environments. *Cities*, 155, 105458. <https://doi.org/10.1016/j.cities.2024.105458>
- Wu, S. Z. (2025). *The AI City*. Springer Nature. <https://doi.org/10.1007/978-981-96-2560-4>
- Xiao Ting Cheng, Ming Wang, & XingQi Fan. (2025). Generation of Personalized Urban Public Space Color Design Scheme Assisted by Artificial Intelligence. *International Journal of High Speed Electronics and Systems*, 7(2). <https://doi.org/10.1142/S0129156425401615>
- Zheng, Y., Lin, Y., Zhao, L., Wu, T., Jin, D., & Li, Y. (2023). Spatial planning of urban communities via deep reinforcement learning. *Nature Computational Science*, 3(9), 748-762. <https://doi.org/10.1038/s43588-023-00503-5>