	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 1/13

ACTA DE REUNION

MODALIDAD: Virtual

DIA: Jueves 1 de Mayo de 2025

1. ASISTENTES

Apellido y Nombre	Organismo
Luis Reynoso	UNCo-DPCeIT
Mabel Rauque	Modernización Santa Cruz
Federico Ferella	Dir.Prov.Estadística Prov. Bs.As.
Leandro Rodriguez	Bunge & Born
Claudio Lezana	DiPIET-Chaco (ex IPECD)
Federico Tricarico	DGEyC SDE - CONICET

2. DESARROLLO

La reunión virtual duró hora y media.

Inicialmente se presentaron resultados del primer cuatrimestre 2025.

JORNADAS SIG E IDE Y JORNADAS IDERA


Se hizo difusión de las jornadas SIG en IDE de La Plata, organizadas por IDERA, los días 12 y 13 de junio de 2024 y de las XIX Jornadas IDERA + I Encuentro IDE de América del Sur a realizarse el 2, 3 y 4 de julio de 2025.

Federico Ferella quien es miembro de nuestro grupo y actualmente lidera la organización del evento en conjunto con la Secretaria Ejecutiva invitó a los miembros del grupo a enviar trabajos bajo las normas de la circular 3.

4 VIDEOS EDUCATIVOS DE IDERA

- El Video 1 muestra cómo utilizar el dataset de GEE Open buildings V3. El video muestra cómo utilizar el set OpenBuildings V3 de Google Earth Engine, este set fue obtenido aplicando ML. El video muestra la importancia de utilizar este set de datos por parte de los catastros, por ejemplo para analizar cuántas parcelas que el catastro tiene como baldías dejaron de serlo porque se han construido. Cantidad de vistas. 472 vistas en youtube url:


<https://youtu.be/JaomYeh6SO8?si=cSE0frdmikfkEkE4>

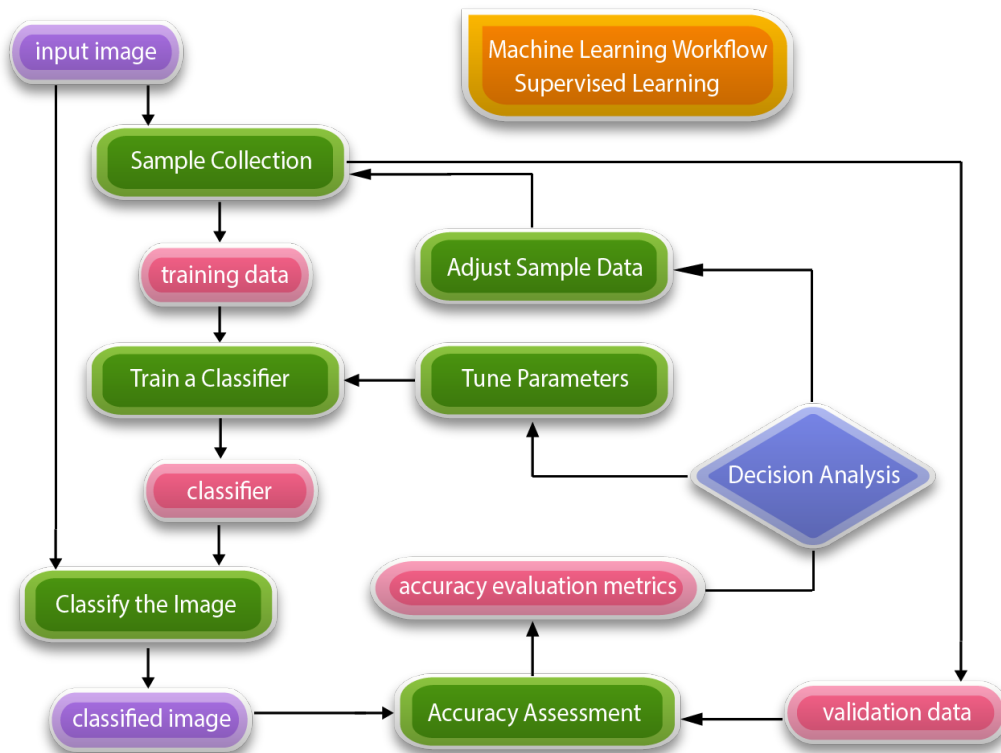
	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 2/13

Videos de Machine Learning aplicado a Información Geoespacial

- El video 2 muestra una Introducción a teledetección y Algoritmos de machine learning (ML) . Entrenamiento Supervisado y No supervisado. La teoría básica de ML y también muestra cómo aplicar Random Forest (en GEE) para el área metropolitana de rosario) en el entorno de Google Earth Engine: 40 minutos, +1K vistas url: <https://youtu.be/fk6atugR6ss?si=QCKRqN-j9vG3B4QI>
- El video 3 muestra como aplicar Máquina de soporte vectorial y Árboles de decisión (SVM y AD para Clasificación multiclase del área Metropolitana de Rosario) utilizando GEE y Python: 372 vistas url: <https://youtu.be/QBJtIQOPdfQ?si=Cnya6dNRfE8s0SXW>
- El video 4 muestra una Clasificación binaria: para detectar agua (Random Forest) en provincia Rio Negro y Neuquén. 671 vistas url: <https://youtu.be/-DMd6LXIEkE?si=aQ6bSISFpt9g-36o>

Los tres últimos videos propician el uso de un flujo de trabajo para aplicación de algoritmos de machine learning aplicado a información geoespacial. La figura siguiente detalla los pasos. Su aplicación garantizará resultados sistematizados que garanticen la calidad.

	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 3/13



Fuente: Elaboración Propia

CUMBRE IA + ENERGÍA + AGUA

El coordinador del GTTCDDG participó en las 2 Cumbres de IA + Energía + Agua, organizada por el Ministerio de Planificación de la provincia del Neuquén, con la participación de IA Lab de la UBA.


En la segunda cumbre Luis Reynoso coordinó una mesa sobre IAGEN IA generativa aplicada a la industria Energética (oil & gas): Mesa IA y agentes de IA en la gestión de Información y la toma de decisiones:

<https://minplanificacion.neuquen.gob.ar/assets/documents/mesas.pdf>

Los organizadores de la cumbre elaboraron 50 casos de uso de aplicación de IA aplicada a la industria energética, descritos con un formato estándar: resumen ejecutivo, podcast en castellano y podcast en inglés, etc. Los casos de uso se pueden encontrar en: <https://hubtechia.com.ar/>

Estos casos de uso pueden ser utilizados como ejemplo para el desarrollo de podcast para aplicación de IA en IG.

En el desarrollo de las mesas de trabajo de la cumbre se destaca la importancia de contar con datos geoespaciales de calidad provenientes de


	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 4/13

teledetección (satélite y drones), sensores IoT, generación de modelos (MDE, capas geológicas), la aplicación de estándares, y atender la necesidad de cambios en marcos legales.

Los organismos están empezando a aplicar IA para mejorar diferentes propósitos:

- detección de fraude, generación de valuaciones fiscales, y en la industria energética
- en la selección de pozos y su análisis de viabilidad, mejora de la extracción, producción y logística,
- detección de fugas de gas metano, contaminación de agua, cambios en el medio ambiente, cambios en vegetación



	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 5/13




Debemos compartir esas experiencias y casos de uso de IA en IG, por ello estamos escribiendo un libro en el GTT-CDG

LIBRO DE IA DEL GRUPO GTT-CDG

El libro se encuentra en desarrollo y se invita a todos los participantes del grupo y organismos que cuenten con casos de aplicación de IA a IG a publicar en el mismo. Hasta ahora tenemos estos capítulos en desarrollo:

- Paola Cardozo, envió el abstract ver Apéndice I.
- Leandro Rodriguez de Bunge & Born, detección de microbasurales
- Luis Reynoso, Flujo de trabajo aplicado a Imágenes satelitales para Clasificación supervisada multiclase y binaria en GEE
- Claudio Lezana (resumen de tecnicas ia en su maestría?)

¿Les parece poner como fecha de cierre de recepción de los capítulos en Agosto?

	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 6/13

MAPA DE LA OFERTA ACADÉMICA

Se comentó la actualización del mapa de la oferta académica y Análisis Varios (publicado en página de IDERA), a partir de la guía de carreras de 2025, que suministró del Ministerio de Educación Diccionario Nacional de Presupuesto e Información Universitaria. Ministerio de Capital Humano.


Se mostró dónde encontrar titulaciones sobre Ciencia de Datos e Inteligencia Artificial en el mapa (bajo la categoría Estadísticas)

<https://www.idera.gob.ar/index.php/comunidad-idera/mapa-oferta-academica>

PROPUESTA DE TRABAJO EN SUBGRUPO

Federico Tricarico propone trabajar el tema de grid.


PROPUESTA DE TRABAJO EN SUBGRUPO	
population grids	<p>Los population grids o datasets de población en cuadrícula son insumos esenciales para ubicar y cuantificar a la población con granularidad espacial muy fina, superando las limitaciones de los niveles administrativos tradicionales y permitiendo análisis, modelado y toma de decisiones más precisos en áreas como salud pública, gestión de desastres y planificación urbana [i][ii]. Dado que existen varios productos globales consolidados —WorldPop, GPW, GHS, LandScan, GRUMP, entre otros— que ofrecen datos teselados de población a escala mundial, podrían compararse y determinar los mejores. Las principales limitantes de estos productos provienen de las distintas resoluciones espaciales, la antigüedad de los censos de entrada (en torno a 2015) y la falta de (necesarios) ajustes locales, los que en casos críticos (ej. emergencias sanitarias o catástrofes) puede llevar a estimaciones imprecisas para un país o región específica [iii].</p> <p>Un population grid es un ráster o malla regular en la que cada celda contiene un valor de población estimada, derivado de la combinación de datos censales o</p>

	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 7/13

	<p>estimaciones de unidades de área (unidades administrativas más o menos desagregadas) y modelado estadísticas de cobertura de la tierra y cartografía/estimaciones dasimétricas.</p>
Definición y ventajas de un population grid	<p>Un population grid es un ráster o malla regular en la que cada celda contiene un valor de población estimada, derivado de la combinación de datos censales o estimaciones de unidades de área (unidades administrativas más o menos desagregadas) y modelado estadísticas de cobertura de la tierra y cartografía/estimaciones dasimétricas.</p>
Propuesta	<p>Comparar y evaluar el desempeño de los productos disponibles. Diseñar el desarrollo de un producto mejor. Buscar interesados. Explorar y exponer las iniciativas, experiencias y metodologías relevantes.</p> <p>Una etapa inicial puede contemplar el análisis de los patrones de asentamientos humanos con información de diversas fuentes y tipos. Enfocar inicialmente en el análisis y clasificación de cobertura de la tierra, y estimaciones de conteo de viviendas. Comparar y evaluar el desempeño de productos dedicados, y experiencias y casos exitosos</p>
Referencias	<p>[i] presentación de Tim Trainor US Census Bureau at UN-GGIM. Metodología de grado de urbanización. aquí y aquí</p> <p>[iii] Sobre la subestimación de población rural en Asia https://www.nature.com/articles/s41467-025-56906-7 . Sobre un producto regional específico https://www.nature.com/articles/sdata201545</p>

PARTICIPACIÓN EN CONGRESOS

- Leandro Rodriguez y Luis Reynoso participaron en el V Congreso Internacional de Geografía Urbana, Área temática: Sistemas de Información Geográfica y Análisis Espacial, Costa Rica, 24 al 29 de marzo de 2025 el título

	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 8/13

de su trabajo es: **Análisis de la valorización de lotes vacantes en el Partido de Pilar (Provincia de Buenos Aires)**

- Luis Reynoso envió dos artículos al congreso XIX CONFIBSIG 19 Conferencia Iberoamericana de Sistemas de Información Geográfica, Universidad de México, 9 al 13 de junio de 2025. Los mismos fueron aceptados <https://sites.google.com/view/xix-confibsig-2025/2da-convocatoria>
- Luis Reynoso envió dos ponencias a las XIX jornadas IDERA 2025
- Leandro Rodriguez envió una propuesta de taller y un short paper a las Jornadas IDERA 2025

INVITACIÓN A EVENTOS

<https://tecnap.neuquen.gov.ar/> 28, 29 y 30 de mayo de 2025 en Villa la Angostura. En agenda tematica sobre IA:


<https://tecnap.neuquen.gov.ar/agenda-2025/>

APÉNDICE I:

Estimación de la biomasa aérea con datos Sentinel-2 y bosque aleatorio: estudio de caso en el Parque Nacional El Rey (Argentina)

Paola Cardozo

Resumen: *El papel de los bosques nativos es de gran importancia en el contexto del cambio climático, no sólo como reservorios de carbono, sino también como proveedores de bienes y servicios ecosistémicos. El monitoreo de la vegetación en áreas protegidas es necesario para evaluar el cumplimiento de los objetivos de conservación. El Parque Nacional El Rey fue creado en 1948 y protege la diversidad biológica y valores culturales de la región de las Yungas centrales y de la cuenca del Río del Valle. En 1979 comenzó la colección, herborización e identificación de plantas del P. N. El Rey. Posteriormente se continuó con el registro de la flora en diversos estudios. Chalukian et al., (2019) describieron diez tipos de ambientes en el área protegida e identificaron 680 especies vegetales, con predominio de especies de la provincia fitogeográfica de la Yungas e ingresos de especies del Chaco serrano y Selva Pedemontana. Desde 2022 está en curso un proyecto para el relevamiento y la conservación de los bosques de queñoa (*Polylepis hieronymi*) en las zonas occidentales del área protegida. En los últimos tiempos, los sistemas basados en teledetección han sido ampliamente utilizados para dar seguimiento a los recursos forestales permitiendo así*

	ACTA DE REUNIÓN	Versión: 1
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Fecha: 01/05/2025
		Página 9/13

superar las limitaciones de los estudios de campo por los costos y mano de obra necesarias para su ejecución. La biomasa aérea (AGB) es la cantidad de materia orgánica presente en troncos, corteza, ramas y hojas. La misión Sentinel 2, comparado con otros datos ópticos obtenidos por teledetección, ha mostrado buen desempeño en la estimación de la AGB en distintos tipos de bosques del mundo. Este estudio se centra en el Parque Nacional El Rey (norte de Argentina) y tiene por objetivo estimar la biomasa aérea mediante la técnica de bosque aleatorio con el modelo de regresión Random Forest en la plataforma Google Earth Engine (GEE). Para ello se utilizarán estimaciones de la densidad de biomasa aérea de la Investigación de la dinámica de los ecosistemas globales (GEDI) que se utilizarán como variable predicha en el modelo de regresión. Como variables independientes se utilizarán datos de elevación y pendiente calculados a partir de un Modelo digital de elevaciones (DEM) y bandas espectrales e índices espectrales obtenidos de Sentinel-2. A los efectos de evaluar el ajuste y la precisión del modelo se utilizarán el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (RMSE) para cuantificar la incertidumbre. El capítulo propuesto brindará información de utilidad para dar continuidad a las acciones de conservación en el P. N. El Rey. Además, permitirá evaluar la utilidad de los datos Sentinel 2 para el modelado, predicción y mapeo de la biomasa aérea, entre otras contribuciones.


APÉNDICE 2 APORTES Y EXPERIENCIAS

2.1

Paola Cardozo- Estimación de la biomasa aérea con datos Sentinel-2 y bosque aleatorio: estudio de caso en el Parque Nacional El Rey (Argentina)


Resumen

El papel de los bosques nativos es de gran importancia en el contexto del cambio climático, no sólo como reservorios de carbono, sino también como proveedores de bienes y servicios ecosistémicos. El monitoreo de la vegetación en áreas protegidas es necesario para evaluar el cumplimiento de los objetivos

	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 10/ 13

de conservación. El Parque Nacional El Rey fue creado en 1948 y protege la diversidad biológica y valores culturales de la región de las Yungas centrales y de la cuenca del Río del Valle. En 1979 comenzó la colección, herborización e identificación de plantas del P. N. El Rey. Posteriormente se continuó con el registro de la flora en diversos estudios.

Chalukian et al., (2019) describieron diez tipos de ambientes en el área protegida e identificaron 680 especies vegetales, con predominio de especies de la provincia fitogeográfica de la Yungas e ingresos de especies del Chaco serrano y Selva Pedemontana. Desde 2022 está en curso un proyecto para el relevamiento y la conservación de los bosques de queñoa (*Polylepis hieronymi*) en las zonas occidentales del área protegida. En los últimos tiempos, los sistemas basados en teledetección han sido ampliamente utilizados para dar seguimiento a los recursos forestales permitiendo así superar las limitaciones de los estudios de campo por los costos y mano de obra necesarias para su ejecución. La biomasa aérea (AGB) es la cantidad de materia orgánica presente en troncos, corteza, ramas y hojas. La misión Sentinel 2, comparado con otros datos ópticos obtenidos por teledetección, ha mostrado buen desempeño en la estimación de la AGB en distintos tipos de bosques del mundo. Este estudio se centra en el Parque Nacional El Rey (norte de Argentina) y tiene por objetivo estimar la biomasa aérea mediante la técnica de bosque aleatorio con el modelo de regresión Random Forest en la plataforma Google Earth Engine (GEE). Para ello se utilizarán estimaciones de la densidad de biomasa aérea de la Investigación de la dinámica de los ecosistemas globales (GEDI) que se utilizarán como variable predicha en el modelo de regresión. Como variables independientes se utilizarán datos de elevación y pendiente calculados a partir de un Modelo digital de elevaciones (DEM) y bandas espectrales e índices espectrales obtenidos de Sentinel-2. A los efectos de evaluar el ajuste y la

	ACTA DE REUNIÓN	Versión: 1
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Fecha: 01/05/2025
Página 11/ 13		

precisión del modelo se utilizarán el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (RMSE) para cuantificar la incertidumbre. El capítulo propuesto brindará información de utilidad para dar continuidad a las acciones de conservación en el P. N. El Rey. Además, permitirá evaluar la utilidad de los datos Sentinel 2 para el modelado, predicción y mapeo de la biomasa aérea, entre otras contribuciones.

2.2


**CLAUDIO LEZANA - TESIS DE MAESTRÍA EN CIENCIA DE DATOS – ITBA.
(Abril/2025)**

Identificación y Conteo de Viviendas en Zonas Periurbanas, por medio de Imágenes Satelitales Ópticas aplicando técnicas de Clasificación de Machine Learning

Resumen:

En la actualidad, la identificación y conteo de viviendas, se realiza mediante censos nacionales o regionales. Estos censos, generalmente son llevados a cabo cada diez años, esta situación impide la disponibilidad de datos actualizados en el ámbito del área de Cartografía del Instituto Provincial de Estadística y Ciencia de Datos (IPECD) y por consiguiente afecta la capacidad de responder eficientemente a las demandas de la sociedad u organismos gubernamentales. Estas demandas abarcan todos los productos ofrecidos por la IPECD, como el Índice de Precios al Consumidor (IPC) y la Encuesta Permanente de Hogares (EPH); entre otros, pero desde la perspectiva de la Geo-estadística, es decir, centrándose en el análisis y la modelización de variables asociadas a la información espacial.

En busca de soluciones innovadoras, aprovechamos los avances en tecnologías de la información, hardware y software, como así también el auge de la Inteligencia Artificial (IA) y, específicamente, los algoritmos de aprendizaje automático o machine learning, sean supervisados o no supervisados, que ofrecen herramientas poderosas para identificar patrones en imágenes multi-espectrales y clasificar objetos geográficos de interés.

	ACTA DE REUNIÓN	Versión: 1
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Fecha: 01/05/2025
		Página 12/ 13

En el presente trabajo se aplicaron diversos algoritmos y técnicas de machine learning cuyo objetivo fue desarrollar una herramienta de apoyo que permita la identificación y conteo de viviendas en imágenes satelitales ópticas de zonas periurbanas del Área Metropolitana del Gran Resistencia (AMGR), en la provincia del Chaco, Argentina. Definiendo como la unidad de análisis al radio censal, establecido por el Instituto Nacional de Estadística y Censos (INDEC).

En el desarrollo de la investigación se llevaron a cabo varios experimentos: utilizando tres enfoques principales: En primer lugar el modelo de **Umbralización**, método básico de segmentación que permite distinguir objetos en una imagen mediante la aplicación de un umbral de intensidad. En segundo lugar el modelo **K-Nearest Neighbors (KNN)**, un algoritmo de clasificación supervisada que asigna una clase a una muestra basándose en la clase más común entre sus k vecinos más cercanos. Para finalizar el modelo **U-Net**, arquitectura de red neuronal convolucional (**CNN**) diseñada para tareas de segmentación de imágenes, que permite una clasificación precisa a nivel de píxel.

Estos métodos fueron evaluados y comparados utilizando métricas de desempeño estándar, como precisión, *recall*, F1-Score y *accuracy*, para determinar el más adecuado. Los resultados indicaron que U-Net es el modelo más robusto y equilibrado, con un F1-Score de 88.4%, destacándose en la identificación de viviendas. El modelo seleccionado fue posteriormente integrado con una base de datos geoespacial para su validación.

En conclusión, la aplicación de modelos de machine learning para la detección de viviendas en imágenes satelitales ópticas ha demostrado ser una herramienta poderosa y eficiente, más específicamente el modelo U-Net que para nuestro caso de uso ha mostrado una precisión aceptable, mejorando el proceso de actualización y procesamiento cartográfico en el IPECD. Aunque se encontraron desafíos, como la selección de una resolución espacial adecuada, la investigación proporcionó un conocimiento valioso sobre los modelos de machine learning y los requisitos de hardware necesarios.

link de Presentación:

<https://docs.google.com/presentation/d/1RLmugXA7tey2cwqEDIYXGblpUUsUALXg/edit?usp=sharing&oid=101435385764224595873&rtpof=true&sd=true>


2.3

LEANDRO RODRIGUEZ - Análisis de la valorización de lotes vacantes en el Partido de Pilar (Provincia de Buenos Aires)

Rodríguez Leandro Nicolás. Universidad Nacional de General Sarmiento

Inrodriguez@campus.ungs.edu.ar

Flores Andrea Pamela. Universidad Nacional de General Sarmiento

	ACTA DE REUNIÓN	Versión: 1 Fecha: 01/05/2025
	Grupo Técnico de Trabajo Ciencia de Datos Geoespaciales	Página 13/ 13

pflores@campus.ungs.edu.ar

Tema: 9. Sistemas de información geográfica y Análisis Espacial

Palabras Claves

Planificación urbana - Lotes Vacantes - Precios de Suelo - Sistemas de Información Geográfica - Geoestadística

Modalidad de presentación: Ponencia Oral

El conocimiento y monitoreo de los precios de suelo urbano representan un insumo esencial para la planificación del territorio. Sin embargo, la desactualización de la información constituye una problemática presente en gran parte de las ciudades latinoamericanas. En este contexto, la aplicación de técnicas geoestadísticas representa una oportunidad para describir e inferir las dinámicas del territorio y su reflejo en los mercados inmobiliarios. El objetivo de este trabajo es evaluar modelos estadísticos que expliquen los cambios en la valorización de lotes vacantes urbanos a partir de variables propias del entorno. El trabajo se desarrolla en el área urbana del municipio de Pilar -Buenos Aires, Argentina- el cual se caracteriza por presentar un crecimiento poblacional intercensal (2010-2022) superior a la media nacional, con un correlato en la expansión del área urbana. Por un lado, se aplican técnicas de interpolación Kriging para estimar los valores del suelo a partir de anuncios recolectados en portales web inmobiliarios. Luego, se seleccionan variables dependientes referidas a la infraestructuras de servicios urbanos para explicar el valor del suelo, mediante la aplicación de técnicas de regresión como Mínimos Cuadrados Ordinarios (OLS) y Regresión Ponderada Geográficamente (GWR). El total de lotes vacantes a analizar fue de 8588. La técnica de interpolación que mejores resultados presentó fue el Kriging exponencial, con parámetros de distancia adaptativa y de kernel gaussiano. La técnica de regresión GWR presentó modelos con un mejor desempeño principalmente en las zonas residenciales suburbanas, por fuera del centro urbano. Las variables que incidieron fuertemente en la explicación de los valores fueron Distancia a Barrios Populares, Distancia a Espacios Verdes y Distancia a Paradas de Colectivo, y explicaron el 80% de la variabilidad del modelo. Los resultados obtenidos, reafirman la importancia de considerar la variabilidad espacial al proponer modelos que expliquen la estructura de precios en lotes vacantes.