

empatia

Sistema de Analítica
Geoespacial para la
Priorización de Áreas
Objetivo de Política
Social y Búsqueda Activa

empatia.la

Escrito por: Alejandro Noriega Campero, Karla Macías Gómez, Alfredo Carrillo del Campo, Alvaro Farias Velasco, Roger G. Rivero Morales

Diseño y diagramación: Ápice Estudio

Licencia Internacional Pública de Atribución/Reconocimiento-NoComercial-SinDerivados 4.0 de Creative Commons



Este proyecto tiene como objetivo resolver problemas públicos en la región América Latina y Caribe (LAC por sus siglas en inglés) en el marco del proyecto global Artificial Intelligence for Development (AI4D) del Centro Internacional de Investigaciones para el Desarrollo (IDRC por sus siglas en inglés). El proyecto es dirigido por ILDA y Centro Latam Digital.

Sistema de Analítica Geoespacial para la Priorización de Áreas Objetivo de Política Social y Búsqueda Activa de Beneficiarios

1. Introducción

1.1 Problema de investigación

Uno de los principales retos para quienes están a cargo de formular las políticas públicas destinadas a combatir la pobreza y la desigualdad es encontrar a sus beneficiarios. Es por ello que se hacen grandes esfuerzos para diseñar e implementar herramientas y mecanismos para identificar a las poblaciones en condición de pobreza. Se han construido índices, como los de marginación o pobreza multidimensional, que capturan las múltiples carencias que puede enfrentar una persona. También, mapas de pobreza que ubican poblaciones en vulnerabilidad socioeconómica, e incluso la búsqueda activa de beneficiarios potenciales que realizan los trabajadores sociales en campo, por mencionar solo algunos.

Sin embargo, la pobreza es un fenómeno complejo y ciertas condiciones y circunstancias pueden invisibilizar a poblaciones marginadas; y lo que es más grave aún, excluirlas de las políticas sociales creadas para mejorar su calidad de vida. Por ejemplo, los recursos de los gobiernos en materia de política social suelen estar destinados a zonas que son consideradas prioritarias en razón de su nivel de marginación. Pero, ¿qué pasa con un hogar que esté en condición de marginación pero dentro de una zona no marginada? o cuando los trabajadores sociales deben realizar visitas a asentamientos irregulares para evaluar su incorporación a programas de regularización, pero no tiene cómo identificar zonas de las que no hay indicios en los registros formales.

1.1.1 Políticas sociales, combate a la pobreza y sistemas de focalización

Las políticas sociales protegen a las personas y a las familias abarcando todos los sectores del desarrollo social. Incluyen programas de transferencias monetarias, de becas, vivienda, sistemas de salud y cuidado infantil subsidiados, seguridad alimentaria, pensiones para trabajadores no formales, apoyo a los microempresarios, e incluso asistencia jurídica subvencionada a los acusados en los tribunales.

Estas políticas tienen un papel central en las estrategias de la mayoría de los países en desarrollo para el combate a la pobreza y la redistribución de la riqueza. De acuerdo al Banco Mundial, los países en desarrollo gastan en promedio el 1.5% del PIB en este tipo de políticas. Por ejemplo, en el caso de las transferencias, éstas reducen en 45% la brecha de pobreza y 36% de quienes las reciben superan la pobreza extrema [1].

Sin embargo, su operacionalización, desde la definición de la población objetivo hasta la ejecución del pago o prestación de un servicio, a menudo enfrenta desafíos sustanciales. Por ejemplo, fuentes de información escasas debido a la alta prevalencia de mercados económicos informales, grandes poblaciones, impactos negativos por desastres, y sistemas de gestión e información desactualizados.

En el caso específico de los sistemas de focalización, estos sufren imprecisiones, ineficiencias, acceso desigual, y falta de transparencia. Esto se traduce en dos tipos de errores: población que recibe apoyos a pesar de no estar en una condición de vulnerabilidad (error de inclusión) y población en condición de vulnerabilidad que no recibe beneficios (error de exclusión). Se estima que en América Latina ambos tipos de errores son de más del 30% [2].

Si las políticas sociales son la columna vertebral de nuestras estrategias para lograr sociedades más equitativas y prósperas, entonces el hecho de no implementar una focalización precisa, justa, eficiente y transparente puede obstaculizar gravemente la capacidad y las esperanzas de la sociedad para alcanzar tales objetivos.

1.1.2 Focalización geográfica

En su mayoría, los sistemas de focalización se han basado en indicadores de pobreza como el ingreso, gasto, consumo y activos de los hogares y/o en indicadores sociales basados en características demográficas de las personas y/o geográficos. Para su implementación se utilizan mecanismos que van desde la selección de beneficiarios basada en la comunidad, indicadores *proxy*, *means-testing*, la autoselección, y hasta la focalización geográfica [3].

Dentro de esta última categoría, se encuentran los mapas de pobreza que en un inicio se elaboraron con base en datos de encuestas como fuente para las estimaciones de

pobreza. Posteriormente, con el objetivo de generar estimaciones con mayor resolución geográfica, se implementaron metodologías que combinan el detalle de la información generada en encuestas con la cobertura exhaustiva de los censos. Sin embargo, aún con esta combinación de fuentes de datos, usualmente la resolución más alta que se alcanza es a nivel manzana [4].

En la actualidad, “nuevos métodos prometedores que combinan encuestas de hogares tradicionales con fuentes no-tradicionales (como datos de teléfonos celulares, datos satelitales y datos de texto) están creando oportunidades para mapear la pobreza con mejor resolución y a mayor escala”. Estos métodos permiten realizar estimaciones con una alta resolución (a nivel predio, por ejemplo), con los datos más actualizados y a gran escala [5].

Adicionalmente, la focalización geográfica ha evolucionado de mapas que identificaban áreas con las tasas de pobreza más altas o con la mayor cantidad de personas en pobreza, a mapas que pueden combinar diversas capas de información. Lo que hace posible, por ejemplo, visualizar la intersección entre vulnerabilidad socioeconómica y vulnerabilidad a desastres naturales.

Finalmente, si los mapas y capas de información generadas son desplegados en una plataforma web interactiva es posible realizar análisis geoespaciales que facilitan a los tomadores de decisiones la simulación y generación de recomendaciones para la planeación y optimización de políticas públicas.

Justamente, en la Ciudad de México (CDMX) donde la pobreza tiene una distribución heterogénea y el 32.6% de la población se encontraba en condición de pobreza (según datos de 2020), aplicar esta metodología de fuentes combinadas es idónea para la elaboración del mapa de pobreza.

1.1.3 Medición de la pobreza en México

Dos de las principales mediciones de pobreza en México son el índice rezago social del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), y el índice de marginación del Consejo Nacional de Población (CONAPO). Estos índices se elaboran con base en características demográficas, socioeconómicas y culturales provenientes del Censo por lo que resumen información con un nivel de confiabilidad y riqueza inigualables.

Sin embargo, las encuestas y los censos presentan retos importantes. Para enfrentarlos, este proyecto combinará el uso de los índices de medición de pobreza y los resultados del Censo de Población y Vivienda 2020 (Censo 2020) con información sensorial remota, específicamente imágenes satelitales diurnas. El contraste en términos de costos, vigencia y resolución espacial de ambos tipos de fuentes de datos se detalla en la Tabla 1.1.

Tabla 1.1: Comparativa de fuentes de datos tradicionales vs no-tradicionales

	Fuentes de datos tradicionales	Fuentes de datos no-tradicionales
Costo	Las fuentes de datos tradicionales como encuestas y censos son mecanismos muy costosos para los países en desarrollo. Por ejemplo, el costo estimado del Censo 2020 en Brasil era de casi 500 millones dólares. De hecho, en el contexto de la pandemia por Covid-19, éste fue pospuesto por razones sanitarias y su presupuesto destinado a atender la crisis de salud. [6]	El mercado de imágenes satelitales ofrece tanto opciones gratuitas como una amplia gama de costos, dependiendo del proveedor y características de las imágenes, que se encuentran por debajo de los de una encuesta o censo. Por ejemplo, para el Área Metropolitana de Sao Paulo que cuenta con 1,700 km ² y 20 millones de habitantes, un costo aproximado promedio sería de 30 mil dólares (en comparación con los 500 millones que cuesta el censo). Cabe señalar que el mercado de imágenes satelitales se encuentra en un proceso de expansión importante por lo que se espera que los costos sigan disminuyendo.
Vigencia	Los datos provenientes de encuestas y censos suelen estar desfasados. El gasto que implica para los gobiernos ocasiona que transcurran largos períodos de tiempo entre levantamientos. Por ejemplo, en México la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) se realiza cada dos años y el Censo cada diez. Así, aunque por ahora existe la ventaja de tener los resultados del Censo en su edición 2020, con el paso del tiempo habrá un desfase entre estos y la cambiante realidad del país.	Las imágenes satelitales se actualizan constantemente y estos cambios se reflejan casi de manera inmediata sobre todo para áreas densamente pobladas como lo son las grandes concentraciones urbanas.
Resolución Espacial	El diseño muestral de las encuestas permite realizar estimaciones a nivel estatal mientras que la muestra censal del Censo 2020 proporciona datos a nivel manzana. Por lo tanto, aun cuando se cuenta con datos actualizados, la información no es exhaustiva ni permite llegar a un nivel de resolución espacial mayor (predio).	La información sensorial se captura para vastas regiones, de manera exhaustiva y a una alta resolución espacial (de 15 cm), lo que permite contar no sólo con información a gran escala y completa, sino con un alto nivel de resolución.

1.2 Objetivo

En el marco de EmpatIA, el objetivo de Centro Latam Digital y Prosperia es generar un mapa de pobreza y asentamientos informales de la CDMX mediante una metodología de fuentes combinadas (tradicionales y no-tradicionales) que posibilite estimaciones y análisis con los datos más actualizados, a un alto nivel de resolución espacial y a gran escala.

Para ello, se llevaron a cabo los siguientes procesos:

1. Mapeo de la pobreza en la CDMX a nivel de celdas de 50 mts. x 50 mts., mediante el uso de imágenes satelitales y un modelo de inteligencia artificial.
2. Mapeo de la pobreza en la CDMX a nivel predio mediante el uso del catastro, imágenes satelitales y un modelo de inteligencia artificial.
3. Mapeo de riesgo de inundación en la CDMX.
4. Disposición de los mapas resultantes en una plataforma web interactiva para la simulación y generación de recomendaciones para la planeación y optimización de políticas e infraestructura social.

Posteriormente, a través de esta plataforma web interactiva de apoyo a toma de decisiones, se llevó cabo el análisis de los mapas resultantes para tres casos de uso de política pública:

1. **Detección de asentamientos irregulares.** Para los propósitos de este proyecto, *asentamiento irregular* se define como aquella área con construcción habitacional detectada con imágenes satelitales y sin registro en el catastro del Sistema de Información Geográfica de la CDMX.
2. **Detección de islas de pobreza e islas de pobreza a nivel predio.** Para los propósitos de este proyecto se define como aquella área con una alta probabilidad de pobreza que está contenida dentro de un área más grande de vulnerabilidad baja. (véase la Figura 1.1)
3. **Detección de áreas con intersección de vulnerabilidad socioeconómica y vulnerabilidad a desastres naturales.** En particular para este proyecto se analizó el riesgo de inundación.

En las siguientes secciones se describen a detalle la metodología y resultados. La sección de conclusiones discute el impacto potencial y valor de este tipo de metodologías para

apoyo de la toma de decisiones y el diseño de políticas, así como avenidas de trabajo hacia adelante.

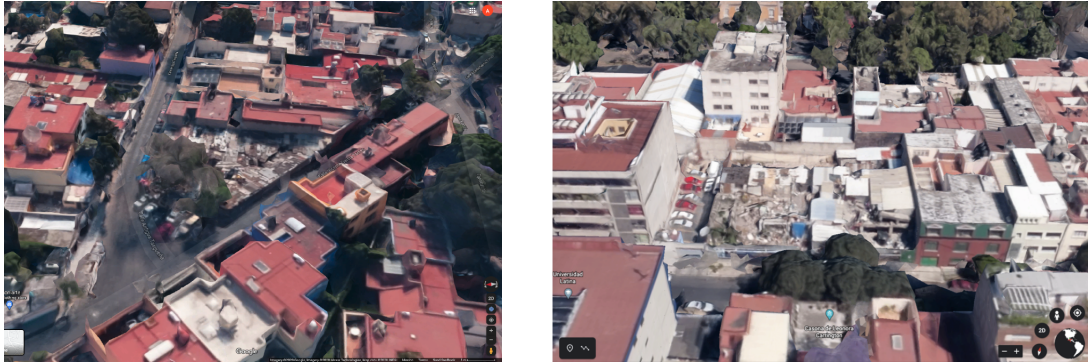


Figura 1.1: Ejemplos de Islas de Pobreza Escondidas

2. Desarrollo del proyecto

En el presente capítulo se describe la metodología seguida en cada una de las etapas del proyecto. En la sección 2.1, se enlistan los datos requeridos en cualquiera de sus formas: encuestas, tablas, imágenes satelitales, bases de datos geoespaciales. En la sección 2.2, se describe cada una de las fases del proyecto, así como las bases de datos utilizadas. En fases avanzadas del proyecto ya no se ocupan los datos en su forma original, sino una derivación o transformación de los mismos. Finalmente, en la sección 2.3 se muestran algunas estadísticas que nos ayudan a evaluar la precisión de los modelos.

2. 1 Datos requeridos

1. **Imágenes satelitales.** Se solicitaron imágenes satelitales diurnas de alta resolución (15cm), con canales de color RGB (rojo, verde y azul), para el área definida por la unión de las AGEBs urbanas del Censo 2020 comprendidas dentro de la Zona Metropolitana del Valle de México (ZMVM) (ver Figura 2.1).¹

¹ Área Geoestadística Básica (AGEB): Extensión territorial que corresponde a la subdivisión de las áreas geoestadísticas municipales. Constituye la unidad básica del Marco Geoestadístico Nacional y, dependiendo de sus características, se clasifican en dos tipos: rural o urbana. <https://www.inegi.org.mx/app/glosario/default.html?p=localidades>

2. **Principales resultados del Censo de Población y Vivienda 2020 a nivel manzana (Censo 2020).** El Censo de Población y Vivienda 2020 fue realizado durante marzo de 2020 por entrevistadores que visitaron miles de hogares para obtener la información de sus características demográficas, socioeconómicas y culturales. Los microdatos del censo a nivel hogar no son públicos. No obstante, se brinda acceso abierto a una base de datos resumida de los principales resultados a nivel de manzana de la CDMX. La base de datos cuenta con información de variables; por ejemplo: población total de la manzana, número de hogares, conteo de resultados de variables como número de habitaciones, recámaras, tipos de materiales del piso y paredes, energía eléctrica, agua entubada, sanitario, drenaje, entre otros. Fue esta base de datos la que se usó para el proyecto.²
3. **Índice de rezago social a nivel de localidad.** El índice de medición de la pobreza del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) resume cuatro indicadores de carencia social (educación, salud, servicios del hogar y espacios habitacionales) en uno solo.³
4. **Índice de marginación 2020 a nivel de AGEB.** Índice de medición de la pobreza del Consejo Nacional de la Población (CONAPO) que permite clasificar, por área geográfica, la intensidad de carencia y exclusión social de la población.⁴
5. **Capa geoespacial de polígonos de manzanas.** Base de datos geoespacial que contiene los polígonos de cada manzana en la CDMX. Las bases de datos geoespaciales también les llamaremos capas en lo que resta del documento. Se le dice así porque cada capa se puede desplegar en un mapa, de forma superpuesta.⁵
6. **Capa geoespacial de polígonos de predios o lotes.** Base de datos geoespacial para la división de predios en la CDMX descargada del Sistema Abierto de Información Geográfica de la Ciudad de México (SIG).⁶
7. **Riesgo de inundación a nivel de AGEB.** Base de datos geoespacial de riesgo de inundación. El riesgo se representa como un rango de probabilidad de inundación de al menos una vez en los siguientes 5 años.⁷

² INEGI: <https://www.inegi.org.mx/programas/ccpv/2020/>

³ CONEVAL: https://www.coneval.org.mx/Medicion/IRS/Paginas/Indice_Rezago_Social_2020.aspx

⁴ Gobierno de México: <https://www.gob.mx/conapo/documentos/indices-de-marginacion-2020-284372>

⁵ Marco Geoestadístico y Censo 2020: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463807469>

⁶ Sistema de Información Geográfica de la Ciudad de México: <https://sig.cdmx.gob.mx/datos/>

⁷ Portal de Datos Abiertos - Gobierno de la Ciudad de México:
<https://datos.cdmx.gob.mx/dataset/atlas-de-riesgo-inundaciones/resource/393b3552-48f2-4208-8798-82dd610a87d9>

2.2 Metodología

1. **Recolección de imágenes satelitales de alta resolución.** Se solicitaron y almacenaron fotos recientes de toda la Zona Metropolitana del Valle de México. La ZMVM comprende toda la CDMX y selectos AGEBS del Estado de México. Desde el cielo, esta área se percibe como una gran mancha urbana. Se solicitó toda la ZMVM en aras de tener más datos y así generar modelos más precisos. Las fotos fueron capturadas entre 2019 y 2020 en su vasta mayoría.
2. **Procesamiento y geo-referenciación de las imágenes satelitales.** Se hace un procesamiento de mapeo de cada foto al punto exacto de la tierra donde corresponde.

Para ello, usamos *rasters* que se guardan con la proyección EPSG:4326, la cual es comúnmente utilizada.

3. **Inventario e integración de las bases de datos.** Se organizan las imágenes satelitales en forma de una malla gigantesca de celdas que cubre a toda la ZMVM. Esta malla se guarda como una capa geoespacial. De esta forma es posible hacer un inventario de toda la información de imágenes almacenadas. Un sistema así es necesario para procesamientos de búsqueda geográficas. Por ejemplo, la rápida identificación de la imagen que cubre una coordenada dada, dentro de un espacio de búsqueda de millones de imágenes.
4. **Creación de ambiente virtual, nube de datos, repositorio de código, y máquinas para cómputo en la nube.** Se prepara un ambiente donde los científicos de datos pueden trabajar en paralelo. También, se crea un repositorio privado en *github* para almacenar y compartir todo el código del proyecto. Se configuran las máquinas de *Amazon Web Services* para el cómputo en la nube.
5. **Preprocesamiento de los datos del Censo 2020.** Se hace un preprocesamiento de la encuesta de Resultados del Censo a nivel de manzana y su correspondiente capa geoespacial del marco geoestadístico. Después, se cruza la información de ambas bases de datos, obteniendo una única base con los polígonos de las manzanas y toda la información de la encuesta del censo.
6. **Cálculo del índice de vulnerabilidad socioeconómica del Censo 2020 a nivel manzana.** Es necesario definir un índice de vulnerabilidad socioeconómica como la variable a predecir por el modelo de inteligencia artificial. Para producir el modelo de pobreza se generan dos muestras de hogares; una muestra contiene imágenes con hogares en alta vulnerabilidad y la otra contiene baja vulnerabilidad. Después, los modelos aprenden a reconocer los patrones de cada una de las muestras de acuerdo con la vulnerabilidad. A esta fase se le llama entrenamiento

de modelos. Después, los modelos pueden predecir con precisión el nivel de vulnerabilidad para cualquier imagen. En realidad, el modelo predice la probabilidad de vulnerabilidad de la imagen, en vez de su nivel de vulnerabilidad, lo cual es una diferencia sutil. Finalmente, el modelo entrenado se aplica para absolutamente cada imagen de la CDMX para poder generar las capas de pobreza.

Explicada ya la necesidad de un índice de vulnerabilidad, existen posibles elecciones de este. Los índices de marginación y rezago social son muy relevantes y de acceso público. Sin embargo, estos se publican a nivel AGEB y localidad, respectivamente; mientras que el presente proyecto requiere predecir a un nivel micro-territorial (celdas de 50 mts x 50 mts), es decir a una resolución más alta. Afortunadamente, los Resultados Principales del Censo 2020, que también son de acceso abierto, permiten llegar a una resolución mayor que localidad o AGEB: a nivel de manzana. Así, con base en dichos resultados, se construyó un índice de vulnerabilidad socioeconómica a nivel de manzana, el cual estará estrechamente ligado a nuestra definición de pobreza.

La construcción del índice considera las siguientes dimensiones:

- a) *Vulnerabilidad del hogar*: Vivienda con más de un cuarto, más de dos cuartos, más de un dormitorio, con piso de tierra, con servicios básicos, agua entubada, drenaje, electricidad.
- b) *Vulnerabilidad en activos*: Si el hogar cuenta con: autos, celulares, computadora, internet, lavadora, refrigerador, teléfono, entre otras.
- c) *Vulnerabilidad económica*: Si el hogar cuenta con población económicamente activa, ocupada, o en asistencia escolar si aplica.

Para el cálculo del índice de vulnerabilidad, para cada una de las variables enlistadas, se genera su correspondiente variable porcentual a partir del número de hogares que satisfacen dichos criterios dividido entre el número total de hogares censados. Después, el índice se construye como un promedio ponderado, otorgando el doble de peso a la dimensión de vulnerabilidad del hogar, pues tiene una correlación más directa con las imágenes satelitales. De esta manera, la dimensión de vulnerabilidad del hogar tiene un peso de 50 %, mientras que la vulnerabilidad en activos y económica el 25 % cada una. Dentro de cada dimensión, todas las variables consideradas tienen el mismo peso.

7. Obtención y procesamiento de las capas intra-manzana del catastro del SIG de la CDMX. Para cada alcaldía, se cruza la capa geoespacial del Sistema Abierto de Información Geográfica (SIG) de la CDMX con su correspondiente base con las características específicas de los predios. Ambos tipos de bases se descargan desde el SIG.

8. Mapeo y superposición de las diferentes capas de datos geoespaciales.

- a) Se cruza la información del marco geoestadístico a nivel de AGEB y el índice de marginación del CONAPO.

- b) Se cruza de nuevo el marco geoestadístico a nivel de localidad con el índice de rezago del CONEVAL.
- c) Se cruza también la información del SIG de predios de CDMX, con la capa de Manzanas del marco geoestadístico.
9. **Generación de la malla en la ZMVM.** Se divide a la ZMVM en celdas de 50 ms x 50 ms y se genera para cada celda, su correspondiente imagen satelital de alta resolución. En otras palabras, se hacen cortes sobre las imágenes previamente almacenadas para generar imágenes a nivel de celda.
10. **Entrenamiento del modelo de Área Habitable Construida (*modelo-CHA*).**⁸ Con la capa de predios habitacionales y la capa geoestadística de manzanas urbanas sin población, se toma una muestra grande para definir aquellas celdas que contienen CHA y aquellas que no. Estas capas de entrenamiento por naturaleza son ruidosas, pues es posible que una celda dentro de una manzana-CHA, contenga una construcción No-CHA, como un parque, o cementerio. Sin embargo, en general la vasta mayoría de las imágenes sí corresponden a lo que se busca identificar.
11. **Evaluación del modelo-CHA.** Se generan varios diagnósticos del modelo para su evaluación. Se obtiene la curva ROC y la estadística AUC, así como también estadísticas de precisión y *recall*. Adicionalmente, se hace una evaluación cualitativa sobre el conjunto de prueba, buscando encontrar fortalezas y debilidades del modelo. Si el modelo demuestra buenos diagnósticos se pasa a la siguiente etapa.
12. **Predicción a nivel de celdas de CHA en CDMX.** Se aplica el modelo-CHA para todas las celdas de la malla en la CDMX.
13. **Entrenamiento del modelo de pobreza para celdas.** Análogamente, se definen dos capas geoespaciales de manzanas con población vulnerable y no-vulnerable a partir del índice de vulnerabilidad aquí construido. Se discriminan todas aquellas celdas donde el modelo-CHA predice baja probabilidad de CHA. Aquí también ambas capas son ruidosas. Por ejemplo, seguramente existen celdas dentro de una manzana no-vulnerable, que esté centrada en una casa pobre. De esta manera se filtra ruido para el entrenamiento, pues se genera una celda con la etiqueta equivocada. Sin embargo, estos casos son escasos en comparación con la gran muestra que contiene la etiqueta correcta.

⁸ La traducción al inglés es *Constructable Habitable Area*, que da pie a la abreviatura CHA.

14. **Evaluación del modelo de pobreza para celdas.** Se hace lo análogo realizado para el modelo-CHA en esta etapa.
15. **Predicción de pobreza para celdas en toda la malla.** Se predice probabilidad de vulnerabilidad en cada una de las celdas de la CDMX. Esta fase concluye la generación de las capas utilizadas para el mapa a nivel de celdas. El objetivo de los pasos siguientes es generar un segundo mapa similar a este pero a nivel de predios, por lo que casi todos los siguientes pasos son completamente análogos.
16. **Entrenamiento del modelo de pobreza para predios.** Se toman cada uno de los predios de la CDMX y se identifica la imagen o conjunto de imágenes que lo comprenden. Después, para cada foto se bloquean los píxeles que no pertenecen al predio, convirtiéndolos en píxeles negros. De esta manera el modelo se entrena con imágenes que contienen un predio a la vez, y no varios de ellos. Esto hace que la predicción se base únicamente en los píxeles que caen dentro del predio de interés.
17. **Evaluación del modelo de pobreza para predios.** Se repite el procedimiento para evaluación de modelos, ahora para predios.
18. **Predicción de pobreza a nivel de predio para toda la malla.** Se predice la probabilidad pobreza para cada uno de los predios. Análogamente, aquí también se aplica el bloqueo de píxeles con color negro en las áreas fuera del predio.
19. **Análisis de potenciales islas de pobreza.** Se identifican grupos de celdas cercanas entre sí, para conformar las islas de pobreza. Las celdas comprendidas dentro de una misma isla comparten ciertas características: alta probabilidad de pobreza, de CHA, y estar en una AGEb de baja marginación. En el siguiente capítulo se da la definición técnica de una isla de pobreza.
20. **Análisis de potenciales asentamientos irregulares.** Se genera una estadística para cada celda de la CDMX que indica si contiene predios registrados en el SIG. Esta estadística es útil para poder encontrar asentamientos irregulares: conjunto de celdas vecinas con alta probabilidad de CHA que no contengan predios en el SIG.
21. **Análisis de potenciales islas de pobreza a nivel predio.** Se identifican predios contenidos en manzanas con índice de no-vulnerables de acuerdo con nuestro índice calculado en el paso 6, y donde la predicción marca que son predios con alta probabilidad de pobreza.

22. **Cálculo del nivel de riesgo de inundación para islas de pobreza y asentamientos irregulares.** A partir del mapa de áreas inundables producido por la CONAGUA a nivel AGEB, se clasifican las islas y asentamientos con un nivel de riesgo de inundación con un periodo de retorno de cinco años.
23. **Generación de mapas y su visualización desde la plataforma de Incluia.** Se generan y cargan los mapas a nivel de celdas y de predio para poder visualizar la pobreza en CDMX en la plataforma de Incluia. Asimismo, se cargan las capas de islas de pobreza, asentamientos irregulares, e islas de pobreza a nivel de predio. Adicionalmente, se cargan otras capas geoespaciales: el mapa de riesgo de inundación, el mapa de marginación de CONAPO, y el de rezago social del CONEVAL.

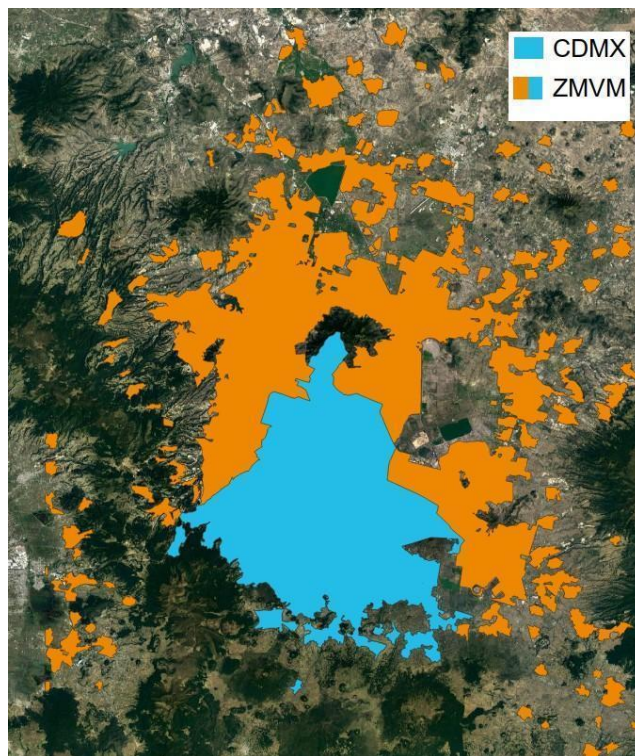


Figura 2.1: Vista aérea de la ZMVM y la CDMX

2.3 Diagnóstico de modelos

Aun cuando estos modelos tienen una larga lista de ventajas, es importante recalcar que no son perfectos; que existen distintas fuentes de error, que son inherentes al modelo probabilístico, y que hay que considerar al momento de aplicar estos modelos. Por

ejemplo, dado que el algoritmo se nutre de información que no podemos evitar que sea un poco ruidosa; o que, dentro de los millones de imágenes, seguramente existirán al menos un centenar de ellas con etiqueta equivocada. De esta manera el modelo aprende errores como si fueran aciertos. Además, dado que el modelo predice una probabilidad y la probabilidad tiene una relación directa con la incertidumbre, sabemos que cualquier modelo probabilístico va a errar eventualmente.

En esta breve sección mostramos algunos diagnósticos cuantitativos sobre el modelo-CHA y de pobreza, ambos a nivel de celdas. Los modelos se evaluaron con un conjunto de aproximadamente 50,000 celdas fuera de muestra con un número de etiquetas relativamente balanceado de pobreza (o CHA) y de no-pobreza (y no-CHA, respectivamente). El nombre de las estadísticas se da en inglés porque son nombres más estandarizados en la academia. La estadística de *precision* nos dice que de todas aquellas celdas que el modelo predice con presencia de CHA, verdaderamente lo tienen el 94.56 % de ellas. Mientras que de todas las celdas con presencia de CHA, la estadística de *recall*, nos dice que el 95.84 % sí son detectadas por el modelo. Finalmente, el modelo tiene una *accuracy* del 94.01 %, que nos da el porcentaje de veces que el modelo acierta la etiqueta de la imagen. Las tres estadísticas nos demuestran que podemos tener alta confianza en el modelo de CHA. Todas las estadísticas se pueden computar a partir de la matriz de confusión normalizada⁹ que se encuentra en la tabla 2.1.

	Negativo	Positivo
Falso	35.01 %	3.39 %
Verdadero	2.56 %	59.04 %

Tabla 2.1: Matriz de Confusión Normalizada de Modelo-CHA

Análogamente, para el modelo de pobreza, las estadísticas son: *precision* es igual a 73.3 %. Se tiene un *recall* igual a 89.82 %. La *accuracy*, el porcentaje de veces que el modelo acierta es igual a 78.2 %. Aun cuando el modelo de pobreza predice peor que el modelo-CHA pues es una tarea más difícil inclusive para los humanos— sigue siendo un modelo útil. El modelo tiene un altísimo valor de *recall*, lo cual es muy bueno. Por otro lado, hay mucho mayor número de falsos positivos en contraste con los falsos negativos. Es decir, el modelo cuando erra, tiende a ser porque predijo pobreza en una celda no-pobre, en comparación de predecir no-pobreza en una celda pobre. Las estadísticas pueden ser calculadas a través de la tabla 2.2.

	Negativo	Positivo
Falso	32.56 %	16.63 %

⁹ Normalizada indica que la tabla está dada con variables porcentuales, en vez del dato duro.

Verdadero	5.17 %	45.64 %
-----------	--------	---------

Tabla 2.2: Matriz de Confusión
Normalizada de Modelo de Pobreza

Finalmente, resta aclarar que los asentamientos irregulares y las islas de pobreza no tienen estadísticas del tipo porque se construyen a partir de la materia prima: celdas con un valor de probabilidad asignado por nuestros modelos de CHA y pobreza. Es decir, se construyen a partir de dos modelos ensamblados y son sólo éstos los que son evaluables de forma independiente cada uno. Asimismo, no se cuenta tampoco con una muestra de asentamientos irregulares de otra fuente con la cual sea posible contrastar polígonos generados; lo mismo sucede para las islas de pobreza.

3. Mapas

El presente capítulo tiene como propósito describir los mapas desplegados en la plataforma de *Incluida*, así como hablar brevemente de las funcionalidades que permite la plataforma. Esta plataforma—creada por Prosperia—permite la visualización y análisis de todos los resultados geoespaciales generados para este proyecto. En dicha plataforma, se generaron dos mapas de la CDMX: un mapa donde la observación primaria son las celdas micro-territoriales (sección 3.1), y un segundo mapa donde la unidad primaria son los predios (sección 3.2).

3.1 Mapa a nivel de celdas

A continuación, se describen las capas que componen el mapa a nivel de celdas.

1. **Malla general.** Esta capa es una malla compuesta por 454,166 celdas de 50 ms x 50 ms, que cubren todas las manzanas urbanas de CDMX registradas en el Marco Geoestadístico 2020 (ver Figura 3.1).¹⁰ Cada celda cuenta con las siguientes variables:
 - a) **Probabilidad de Área Habitacional Construida.** Probabilidad estimada a partir del modelo de CHA (ver Figura 3.2).
 - b) **Probabilidad de Pobreza.** Probabilidad estimada a partir del modelo de pobreza para celdas (Figura 3.3).
 - c) **Índice de Marginación 2022 categórico.** Cada celda contiene la categoría de marginación dada por la CONAPO a nivel de AGEB. Cada una la hereda del nivel de marginación de la AGEB que la contenga. Si una celda interseca con

¹⁰ Marco Geoestadístico del INEGI: www.inegi.org.mx/temas/mg/

dos o más AGEBs, se asigna aquella AGEB con mayor porcentaje de área de intersección (ver Figura 3.4).

- d) **Índice de Rezago Social 2020 categórico.** Análogamente, se hereda a nivel de localidades, el índice categórico calculado por el CONEVAL de rezago social para cada una de las celdas (ver Figura 3.5).
- e) **Si la celda contiene predios del catastro del SIG de la CDMX.** Columna binaria donde la categoría Sí indica que existe algún predio registrado en el SIG que intersecta con esa celda. En caso contrario se asigna la categoría de No. La columna es útil para identificar zonas donde el SIG no tiene información o donde existen asentamientos irregulares (ver Figura 3.6).
- f) **Islas de Pobreza.** Esta capa permite la detección de zonas con vulnerabilidad socio- económica dentro de una AGEB clasificada con baja o muy baja marginación según CONAPO. Es decir, se detectan zonas vulnerables en AGEBs no-vulnerables. A nivel técnico, una isla es un conjunto de celdas cercanas (un *cluster*) en las que todas las celdas satisfacen estas condiciones:
 - i) Celdas con índice de marginación bajo o muy bajo.
 - ii) La probabilidad de CHA es ≥ 0.50 .
 - iii) La probabilidad de pobreza es ≥ 0.66

Las islas se componen de celdas vecinas entre sí. Una celda es vecina a otra de forma lateral, es decir, si comparten un lado, o en forma diagonal, si comparten un vértice. Adicionalmente las islas satisfacen:

- i) El tamaño de la isla (número de celdas) es ≥ 4 .
- ii) El promedio de la probabilidad de pobreza de las celdas es ≥ 0.80 .

Las variables con las que cuenta cada isla son:

- a) **Probabilidad de pobreza.** Promedio de la probabilidad estimada de pobreza de las celdas que conforman la isla.
- b) **Tamaño de la isla.** Número de celdas que conforman la isla.
- c) **Riesgo de inundación.** Nivel de riesgo de inundación con un periodo de retorno de cinco años de la AGEB que contenga a la isla. En caso de que más de dos AGEBs la contengan, se despliega el nivel de riesgo de la AGEB que cubra la mayor área de la isla. Los niveles están definidos por cinco categorías. Cada categoría tiene un número entero del 1 al 5 y a lado, entre

paréntesis, el rango de la probabilidad de que se inunde el AGEB en ese periodo de retorno. Las categorías son: 1 (0 - 25), 2 (26 - 49), 3 (50 - 72), 4 (73 - 99), y 5 (100).

- d) **Tipo de riesgo.** Aun cuando en la mayoría de los casos una única AGEB contiene enteramente a la isla, para distinguir el caso donde no, se define esta variable. Si toda la isla está contenida dentro de un único AGEB, entonces el tipo de riesgo es homogéneo. En caso contrario, cuando está en más de una AGEB, el riesgo de inundación es de tipo heterogéneo.

La capa se compone de 39 islas. Algunos casos interesantes se analizan en el siguiente capítulo.

2. **Asentamientos irregulares.** A partir de la información de predios contenida en el catastro del Sistema de Información (SIG) de la CDMX, se detectan aquellas áreas con alta probabilidad de construcción habitable y sin embargo no hay registro de sus respectivos predios dentro de este catastro. A esto se le llama aquí un asentamiento irregular. Similarmente, el análisis se hace a nivel de *cluster* (grupo de celdas vecinas) y las condiciones técnicas de las celdas requeridas para conformar un *cluster* de asentamiento irregular son:

- i) Celdas sin predios de la capa del Catastro del SIG.
- ii) La probabilidad de CHA es ≥ 0.50 . Las condiciones a nivel de *cluster* fueron:
 - i) El tamaño de *cluster* (número de celdas) es ≥ 4 .
 - ii) El promedio de probabilidad de pobreza de celdas en el *cluster* es ≥ 0.50 .

Las variables con las que cuenta cada asentamiento irregular son las mismas que para el caso de las islas de pobreza.

Se recomendaron 453 *clusters* con estas condiciones. Se analizan algunos ejemplos de estos en el siguiente capítulo.

3.2 Mapa a nivel de intra-manzanas o predios

1. **Mapas de predios por alcaldía.** Se generaron 16 mapas de predios para cada alcaldía (ver Figura 3.8). No se puede comprender en un único mapa todos los predios de la CDMX porque el archivo sería demasiado pesado y el tiempo requerido para su visualización dinámica es excesivo, así que los mapas están independientemente almacenados. Estos cuentan con las siguientes variables:
 - a) Probabilidad de Área Habitacional Construida. Probabilidad estimada a partir del modelo de CHA.
 - b) Probabilidad de Pobreza. Probabilidad estimada a partir del modelo de pobreza para predios (ver Figura 3.9).
2. **Mapa de Islas de Pobreza a nivel de predios de CDMX.** Para compensar que no se sube una sola capa de todos los predios de CDMX, se hizo una capa que contiene una selección de cerca de 100,000 predios que satisfacen las siguientes condiciones:
 - i) Probabilidad de pobreza ≥ 0.70 .
 - ii) Probabilidad de CHA ≥ 0.50 .
 - iii) Contenidos en una manzana con Índice de Vulnerabilidad calculado a partir de variables del censo ≤ 0.50 .

De esta manera, se intentan identificar aquellos predios que potencialmente estén rodeados de otros predios con menor vulnerabilidad, y por esta razón se denominan como islas de pobreza. Las variables que esta capa contiene son:

- a) **Probabilidad de Área Habitacional Construida.** Probabilidad estimada a partir del modelo de CHA.
- b) **Probabilidad de Pobreza.** Probabilidad estimada a partir del modelo de pobreza para predios.

Algunos ejemplos de fotos de islas de pobreza a nivel de predios se encuentran en el anexo del presente documento.

3.3 Navegación en la plataforma

En la plataforma de *Incluia* se puede navegar a cada una de las islas una a una. Para esto existe un botón que acerca la vista hacia la isla seleccionada por el usuario. Sin embargo, para facilitar o hacer el análisis más ordenado, se pueden aplicar una serie de filtros o reordenamientos sobre las variables. Por ejemplo, se puede reordenar por mayor probabilidad de pobreza, o de acuerdo con el tamaño. También se pueden hacer filtros, por ejemplo, las islas que estén en alto riesgo de inundación, y sólo se muestran éstas. Más aún, la plataforma permite combinar varios filtros, bajo la operación Y-lógica. Al hacer esto, todas las islas que no satisfacen cualquiera de las condiciones de los filtros desaparecen de la visualización. Por ejemplo, el usuario puede ajustar los parámetros para sólo mostrar las islas con probabilidad de pobreza mayor a 0.70 y de tamaño mayor o igual a 10. Filtros análogos se pueden aplicar a cada una de las capas desplegadas de acuerdo a las variables que contiene. Particularmente es muy útil mostrar la *Malla General* filtrando por celdas con alta probabilidad de CHA, por ejemplo ≥ 0.50 .

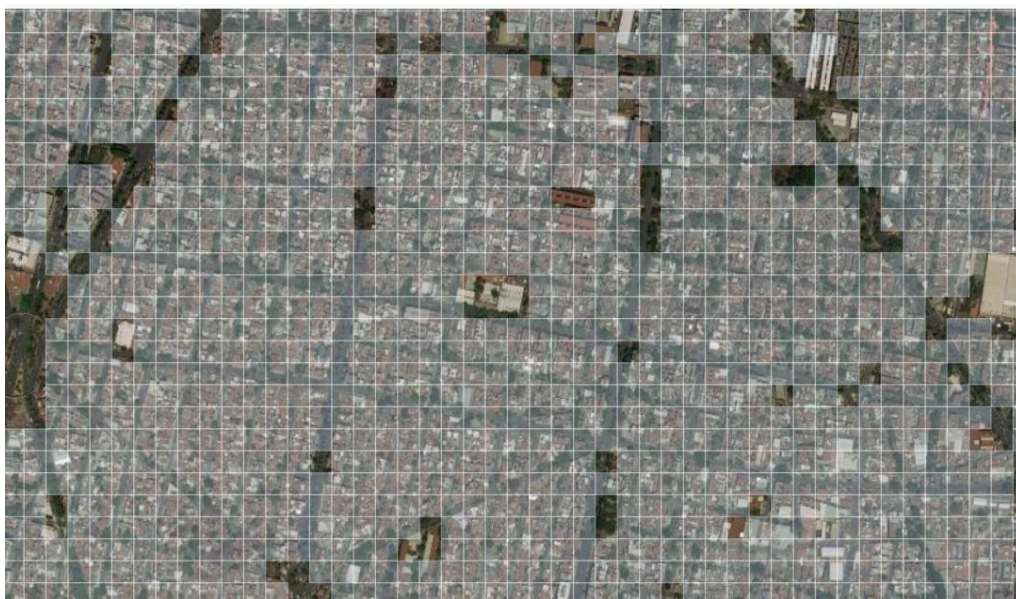


Figura 3.1: Acercamiento a muestra de celdas filtradas por CHA



Figura 3.2: Filtrado por alto nivel de CHA

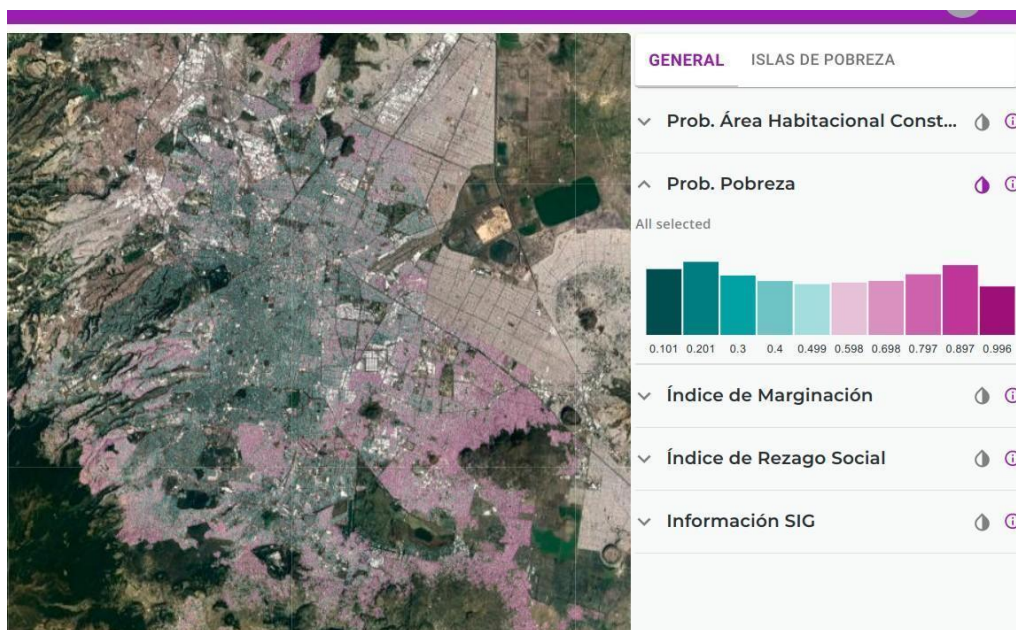


Figura 3.3: Celdas coloreadas por probabilidad de pobreza

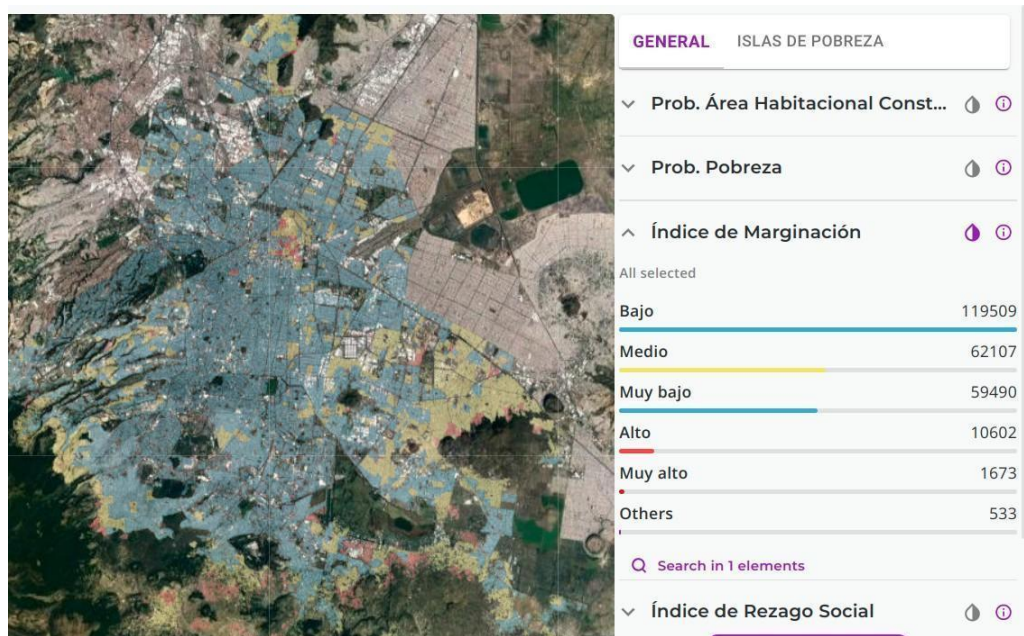


Figura 3.4: Celdas coloreadas por Índice de Marginación.

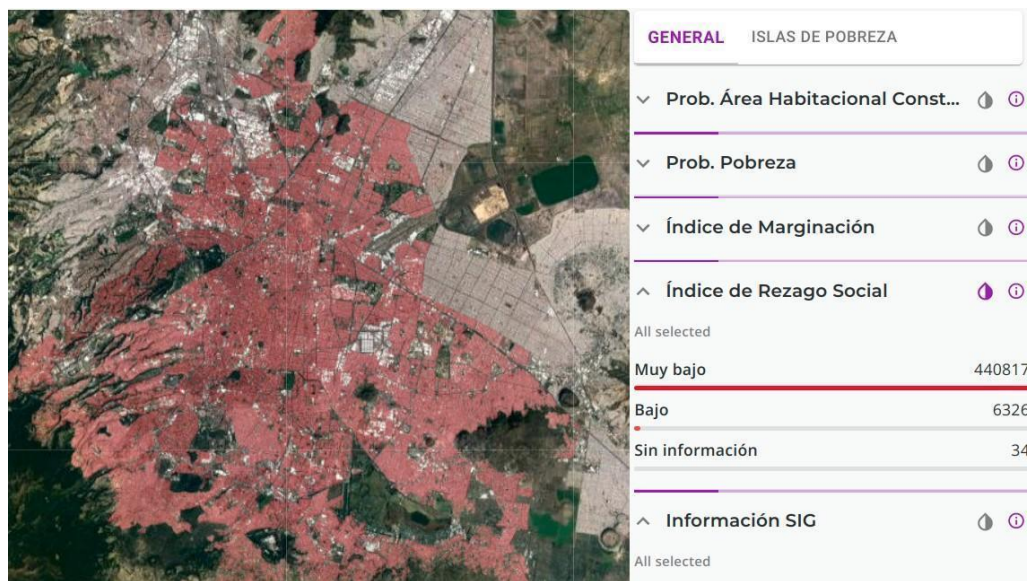


Figura 3.5: Celdas coloreadas por Índice de Rezago Social.

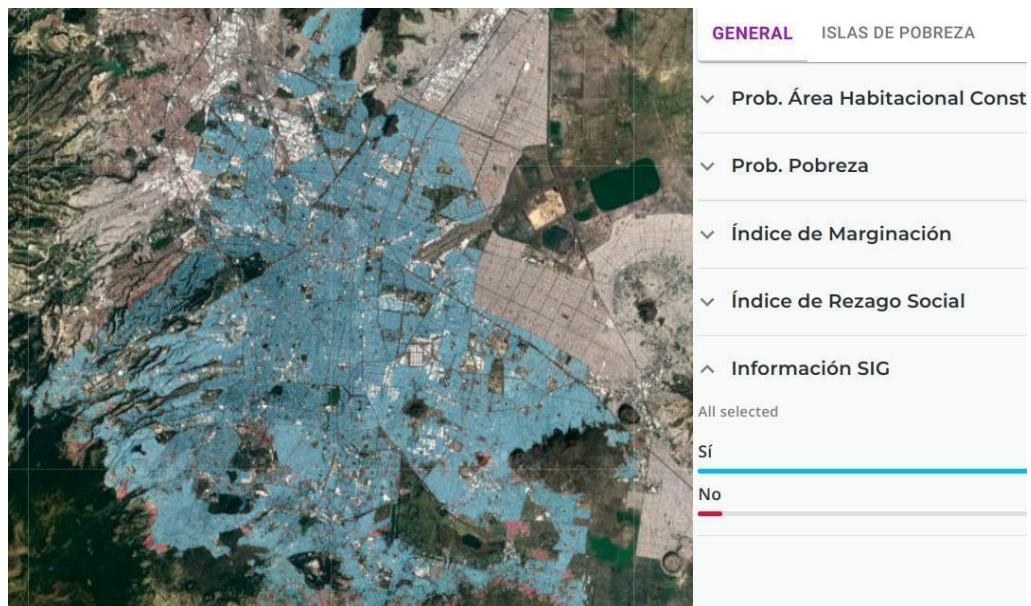


Figura 3.6: Celdas coloreadas por intersección con Catastro-SIG.

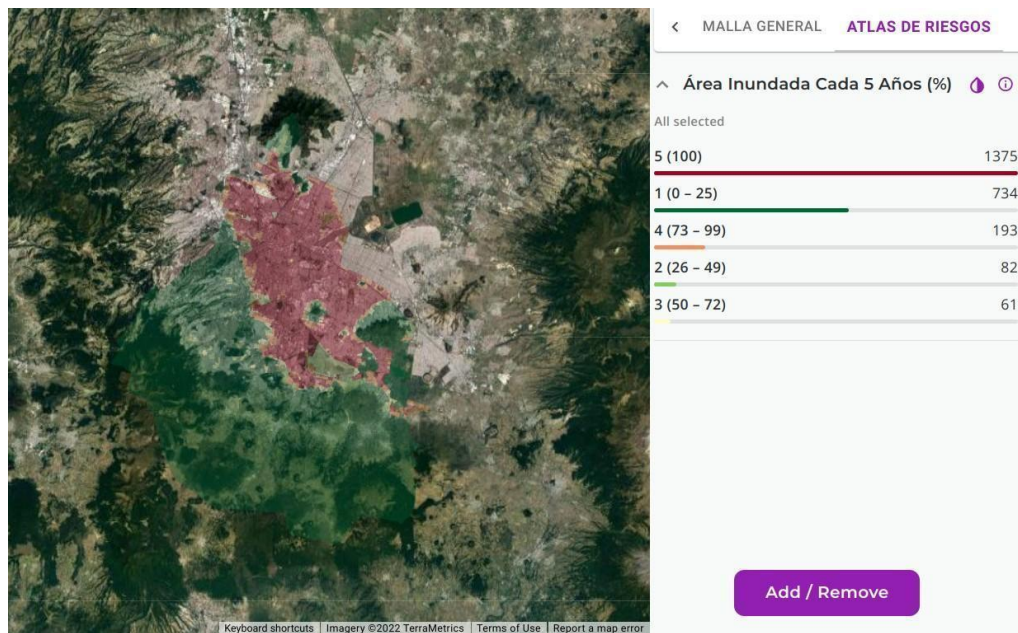


Figura 3.7: Celdas coloreadas por probabilidad de área inundada.

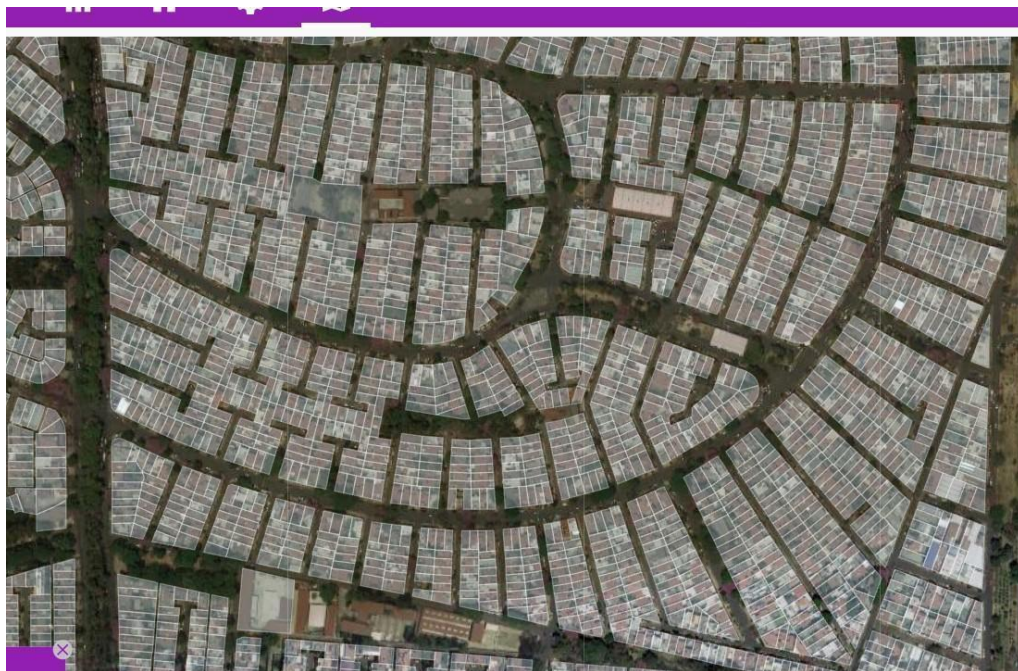


Figura 3.8: Acercamiento a muestra de predios.

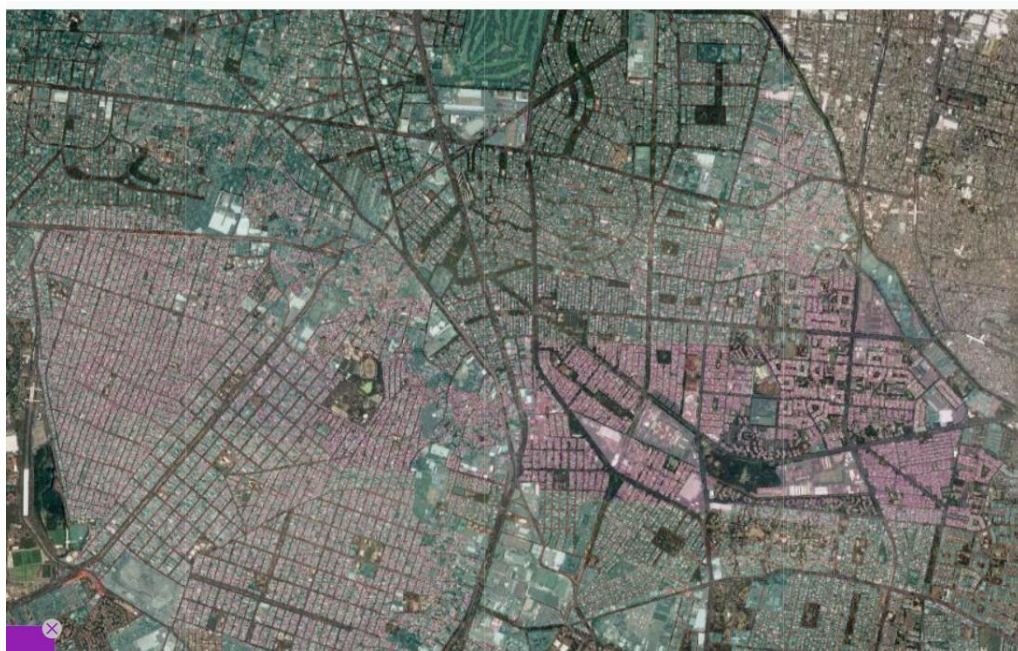


Figura 3.9: Predios en la alcaldía de Coyoacán coloreados por pobreza.

3.4 Casos de usos de política pública

En la plataforma se muestran tres casos de uso relevantes que pueden ser guías de gestión de política pública. El primero es la identificación de asentamientos irregulares; el segundo la identificación de islas de pobreza; y finalmente, cruzando información de cualquiera de estas dos capas junto con la capa de riesgos de inundación, es posible identificar dentro de esta selección de áreas, aquellas que están más endebles y que posiblemente requieran mayor atención si se toma como criterio un potencial riesgo de inundación.

3.4.1 Identificación de asentamientos irregulares

Los asentamientos irregulares son zonas con cierto grado de vulnerabilidad, comúnmente con una alta densidad de personas, y que, al no estar registradas en el sistema de información geográfica de la CDMX, es difícil focalizar la ayuda en ellos. En contraste con las islas de pobreza, estas zonas no necesariamente se encuentran en AGEBS con bajo nivel de marginación, sino que se pueden encontrar en cualquier parte de la ciudad. Sin embargo, como el SIG contiene información casi exhaustiva de la CDMX, usualmente estos asentamientos urbanos se acaban encontrando en lugares muy escondidos, de difícil acceso y detección terrestre. Por ejemplo, atrás de un bosque, o sobre un risco, o atrás de una zona residencial de alto nivel socioeconómico, etc. La ventaja de usar este tipo de modelos aéreos, que cubren a toda la CDMX de forma exhaustiva, es que nada se les esconde, mucho menos si los asentamientos son amplios. Además, como el modelo de CHA, tiene una alta precisión, esto se hereda para que la capa de asentamientos tenga un alto grado de precisión también. Por último, resaltar que esta capa también le puede ser útil a instituciones como el SIG. A continuación, mostramos algunos ejemplos de ellos:



Asentamiento irregular 1: 2D



Asentamiento irregular 1: 3D



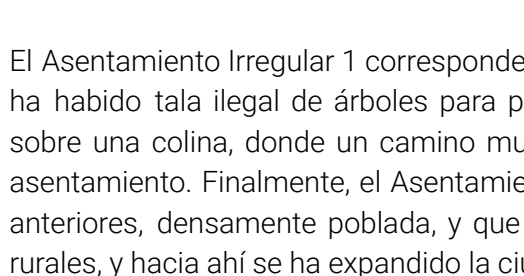
Asentamiento irregular 2: 2D



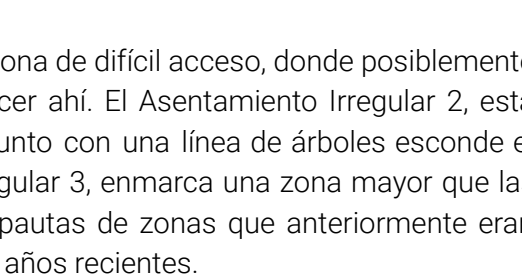
Asentamiento irregular 2: 3D



Asentamiento irregular 3: 2D



Asentamiento irregular 3: 3D



El Asentamiento Irregular 1 corresponde a una zona de difícil acceso, donde posiblemente ha habido tala ilegal de árboles para permanecer ahí. El Asentamiento Irregular 2, está sobre una colina, donde un camino muy fino junto con una línea de árboles esconde el asentamiento. Finalmente, el Asentamiento Irregular 3, enmarca una zona mayor que las anteriores, densamente poblada, y que da las pautas de zonas que anteriormente eran rurales, y hacia ahí se ha expandido la ciudad en años recientes.

3.4.1. Identificación de islas de pobreza

El análisis exhaustivo a un nivel de micro-área logra identificar cientos de áreas con alta vulnerabilidad en zonas de la CDMX asociadas a un alto nivel socioeconómico y por ello áreas propensas a ser invisibilizadas para la focalización de programas sociales. El valor de esta capa es que sería difícil su detección sólo a partir del uso de datos tradicionales, porque se requeriría una resolución de muy alto nivel y además con representatividad exhaustiva. Sin embargo, si no fuera por datos tradicionales, tampoco se tendrían las muestras necesarias de este tipo de áreas para que los modelos se entrenen. A continuación, mostramos algunos ejemplos de ellos:



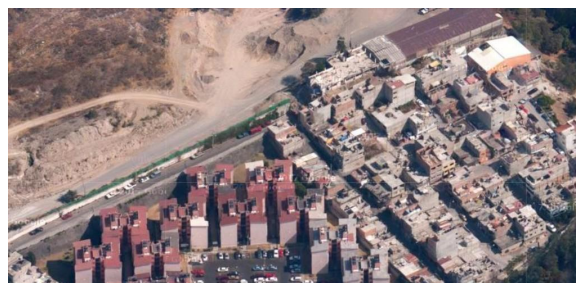
Isla de pobreza 1: 2D



Isla de pobreza 1: 3D



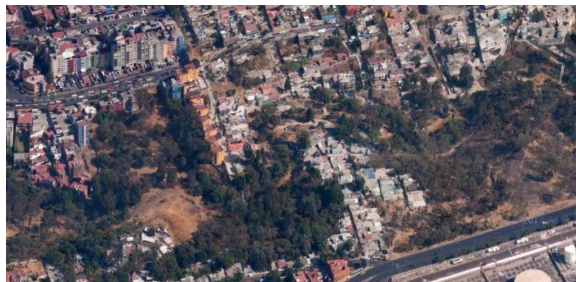
Isla de pobreza 2: 2D



Isla de pobreza 2: 3D



Isla de pobreza 3: 2D



Isla de pobreza 3: 3D

La Isla de Pobreza 1 está justo detrás de un condominio, o manzana de alto nivel socio-económico y por el otro lado una línea de árboles que hace difícil su detección. La Isla de Pobreza 2, similarmente, se sitúa enfrente de un complejo nuevo de edificios de alto nivel económico. A diferencia de la isla anterior, este *cluster* se detecta con facilidad a nivel terrestre, pero la AGEB y la zona en la que se sitúa dificulta su registro. La Isla de Pobreza 3 se ve como un área de difícil acceso, altamente poblada, y extremadamente vulnerable. De hecho, esta isla también es un asentamiento irregular. Es decir, hay *clusters* que son de ambos tipos: asentamientos irregulares e islas de pobreza.

3.4.2 Detección de áreas con intersección económica y vulnerabilidad a desastres naturales

Al hacer el cruce de la información de riesgo de inundación con las capas de asentamientos e islas de pobreza se generan nuevos niveles de información que dan a pie a nuevas formas de priorización, y/o focalización. Por ejemplo, en la zona de oriente, como en la alcaldía de Xochimilco—famosa por sus canales, que son lo que queda de lo que alguna vez fue un gran lago—tiene mucha más propensión a la inundación, que otras partes de la ciudad, y además, es una alcaldía con alto grado de pobreza. En contraste, están las AGEBS del poniente, que, al tener mayor elevación, ser una zona más montañosa, el agua baja a zonas de menor elevación y los riesgos de inundación para estos *clusters* son menores.

Enfatizamos que, para el presente proyecto, a manera de ejemplificación, se utilizó un único tipo de riesgo de desastre natural, pero en realidad se podrían usar otros, por ejemplo, riesgo de terremoto. Este fenómeno ocurre con alta frecuencia en la CDMX, y el nivel de riesgo varía entre AGEBS, al igual que el riesgo de inundación, por lo cual daría también otro tipo de esquema de focalización.



Asentamiento en riesgo bajo de inundación



Asentamiento en riesgo alto de inundación

4. Conclusiones

El uso de la inteligencia artificial ha revolucionado la focalización de políticas públicas. En el caso específico de la focalización geográfica, ésta permite integrar datos de alta resolución, junto con capas de información adicionales para realizar análisis geoespaciales de mayor complejidad.

En particular, los mapas de pobreza ahora pueden identificar no sólo concentraciones de población en pobreza extrema sino también áreas geográficas con patrones atípicos o criterios adicionales de vulnerabilidad contribuyendo a la generación de esquemas de focalización más precisos y políticas sociales más inclusivas.

Aprovechando estos avances tecnológicos, el presente proyecto generó un mapa de pobreza y asentamientos informales para la CDMX mediante una metodología de fuentes combinadas (tradicionales y no-tradicionales) y modelos de inteligencia artificial para posibilitar estimaciones y análisis con los datos más actualizados, a un alto nivel de resolución espacial y a gran escala.

Posteriormente, mediante el despliegue de dichos datos en una plataforma web interactiva que simula y genera recomendaciones para la planeación y optimización de políticas públicas, se analizaron tres casos de uso: i) detección de asentamientos irregulares, ii) detección de islas de pobreza e islas de pobreza a nivel predio, y iii) intersección entre vulnerabilidad socioeconómica y vulnerabilidad a desastres naturales, obteniéndose los resultados que se detallan a continuación.

Se lograron identificar cerca de 400 asentamientos irregulares, 150 islas de pobreza, y 100,000 islas de pobreza a nivel de predio. Al analizar distintos *clusters* o predios recomendados por este sistema, se identificaron zonas muy interesantes: *clusters* e islas de pobreza atrapadas en zonas de poca vulnerabilidad socioeconómica, así como asentamientos irregulares, que son fácilmente invisibilizadas al ser colindantes de edificios de alto nivel socioeconómico. Estas capas, aunque son imperfectas, son una gran forma para explorar los resultados de las predicciones de forma mucho más sistemática, dentro de un mar de información. El usuario puede analizar minuciosamente las potenciales zonas recomendadas y juzgar cada una de ellas. Aun cuando en el presente documento sólo se muestran ejemplos donde las zonas de recomendación detectadas sí son zonas, o bien con vulnerabilidad socioeconómica, o con asentamientos irregulares, también se hallaron casos de falsos-positivos donde la zona recomendada no parecía cumplir con las características requeridas.

Una desventaja es que debido a que existe una gran heterogeneidad en cada manzana de la CDMX, son precisamente los hogares escondidos en zonas que no corresponden a su

propio nivel socioeconómico y que aquí se buscan detectar, que contaminan de cierta manera las etiquetas generadas de forma sistemática para entrenar a los modelos de inteligencia artificial. En una futura línea de trabajo, se buscará etiquetar las imágenes de forma manual, donde humanos etiqueten miles de imágenes una por una, para evitar que ocurra este tipo de contaminación de etiquetas.

Sin embargo, a pesar de la contaminación de etiquetas mencionada, son tantas las etiquetas correctas, que los modelos actuales fueron bien entrenados y son muy precisos; arriba del 80% de los casos se acierta para celdas fuera de muestra para el modelo de pobreza y 91% para el modelo de CHA. A un nivel cualitativo, los *clusters* analizados hacen mucho sentido, corresponden a como un humano esperaría que se viera una isla de pobreza o un asentamiento irregular.

Una gran cualidad del presente proyecto es que los modelos fueron entrenados con datos a nivel de manzana del reciente Censo 2020. Esto implica que la muestra de manzanas con la que se cuenta, además de que es muy vasta, tiene información actualizada. Como las imágenes también son recientes, entonces la información del Censo 2020 es consistente con lo que se ve en la imagen y cuadra con la realidad actual.

Si bien la metodología utilizada es susceptible de mejoras, los resultados del proyecto tienen un gran potencial para ser un insumo estratégico en la toma de decisiones en materia de combate a la pobreza. Por ejemplo, al conocer las islas de pobreza hasta ahora invisibles los tomadores de decisiones contarán con un universo de beneficiarios potenciales a los que de otra manera no ubicarían ni podrían impactar positivamente.

Además de los casos de uso que se desarrollaron en este proyecto, futuras líneas de trabajo para el mapa de pobreza de la CDMX podrían estar enfocadas a la integración de capas de información adicionales (padrón de beneficiarios de programas sociales, infraestructura social, mapas de riesgos, etc.) para informar decisiones de política pública tales como:

1. **Priorización para la actualización de datos del registro social.** Las campañas para la búsqueda activa de población vulnerable presentan retos tanto en términos de inversión de recursos humanos y materiales como de tiempo. Por ello, es fundamental priorizar las áreas geográficas en la que se realizará dicho empadronamiento para garantizar campañas en campo precisas y costo-eficientes y la cobertura de la población más vulnerable (reducción de error de inclusión).
2. **Planeación y optimización de programas e infraestructura social.** Identificar y dimensionar la demanda social insatisfecha de programas e infraestructura social

con el objetivo de eficientar la gestión y distribución geográfica de recursos destinados a su expansión o consolidación.

3. Evaluación de la intersección entre la vulnerabilidad socioeconómica y vulnerabilidad a desastres naturales. La incorporación de información sobre el riesgo a desastres naturales y cambio climático es un elemento clave en el fortalecimiento de políticas sociales preventivas y más responsivas.

Cabe señalar que este tipo de sistema de apoyo a la toma de decisiones pueden ser de gran utilidad, no solo para gobiernos sino también para la sociedad civil (como organizaciones sin fines de lucro, organizaciones privadas que brindan asistencia social, etc.). Por ejemplo, "Un Techo para mi País México", una organización sin fines de lucro que busca superar la pobreza en los barrios marginales mediante un proceso de fortalecimiento comunitario podría identificar con menores costos de operación y en menor tiempo las zonas susceptibles de intervención.

Otro punto para destacar es que la metodología utilizada muestra un trayecto del *Big Data* al *Small Data*. Si bien es a través del procesamiento de aproximadamente 8,000 imágenes satelitales que identifican asentamientos irregulares, esto es un insumo que se suma a la experiencia y conocimiento de los funcionarios de las instituciones sociales. Por ejemplo, una vez detectada un área con asentamientos irregulares, es turno de los trabajadores sociales de realizar visitas para evaluar su estatus y definir pasos a seguir como podría ser su ingreso a un programa de regularización. Es el complemento de la tecnología y de quienes conforman a las instituciones sociales lo que posibilita hacer visibles a los invisibles.

Referencias bibliográficas

[2] Alejandro Noriega-Campero, Bernardo Garcia-Bulle, Luis Fernando Cantu, Michiel A Bakker, Luis Tejerina, and Alex Pentland. Algorithmic targeting of social policies: fairness, accuracy, and distributed governance. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 241–251, 2020.

[3] Rachel Slater, John Farrington, M Vigneri, M Samson, and S Akter. Targeting of social transfers: A review for dfid. *London*: ODI, 2009.

[6] Reuters Staff. Brazil postpones 2020 census to 2021 due to coronavirus out- break. Disponible en: <https://www.reuters.com/article/us-health-coronavirus-brazil-census/brazil-postpones-2020-census-to-2021-due-to-coronavirus-outbreak-idUSKBN2144AP>.

[4] Tara Bedi, Aline Coudouel, and Kenneth Simler. *More than a pretty picture: using poverty maps to design better policies and interventions*. World Bank Publications, 2007.

[5] Varios Autores. Confeccionar un mapa más preciso de la pobreza. <https://blogs.worldbank.org/es/opendata/confeccionar-un-mapa-mas-preciso-de-la-pobreza>, 2019.

[1] World Bank. *The state of social safety nets 2018*. The World Bank, 2018.

Anexo

Ejemplos de islas de pobreza identificadas a nivel de predio



Isla 1: 2D



Isla 1: 3D



Isla 2: 2D



Isla 2: 3D



Isla 3: 2D



Isla 3: 3D



Isla 4: 2D



Isla 4: 3D

