

Guía para la recolección de datos de entrenamiento

Modelo predictivo de valor de propiedades (en CLP)

Introducción

Los modelos predictivos basados en redes neuronales profundas son sistemas de inteligencia artificial que aprenden patrones complejos a partir de grandes cantidades de datos. En el contexto inmobiliario, estos modelos analizan información histórica de propiedades —como ubicación, tamaño, tipo y valor— para encontrar relaciones que permitan estimar el precio de venta o arriendo de nuevas propiedades. Cuanto más organizado y preciso sea el conjunto de datos de entrenamiento, mayor será la capacidad del modelo para generalizar y hacer predicciones confiables.

Esta guía explica cómo recolectar y estructurar datos para alimentar un modelo de machine learning que predice el valor de venta o arriendo de propiedades. Está pensada para que cualquier persona pueda realizar la recolección de manera clara y consistente.

1. Identifica las fuentes de datos

- Portales inmobiliarios (ej: Portal Inmobiliario, TocToc, Yapo).
- Bases de datos internas o registros históricos.
- Publicaciones de corredores o información pública de catastros municipales.

2. Estructura de la tabla de datos

Crea una tabla (en Excel o CSV) con las siguientes columnas:

operación	región	comuna	latitud	longitude	habitaciones	baños	m2_construidos	m2_terreno	tipo_propiedad	valor_CLP	fecha_valo
-----------	--------	--------	---------	-----------	--------------	-------	----------------	------------	----------------	-----------	------------

- **operación:** SALE/RENTAL
- **región:** Nombre de la región
- **comuna:** Nombre de la comuna
- **latitud:** Ej: -33.4489
- **longitude:** Ej: -70.6693
- **habitaciones:** N° de dormitorios
- **baños:** N° de baños
- **m2_construidos:** Superficie construida
- **m2_terreno:** Superficie de terreno
- **tipo_propiedad:** Casa/Depto/Oficina/etc.
- **valor_CLP:** Valor en CLP
- **fecha_valo:** Fecha en formato AAAA-MM-DD

Notas importantes:

- Latitud y longitud son opcionales: Si no conoces las coordenadas exactas, puedes dejar estos campos en blanco o NULL. El modelo usará región y comuna igualmente.
- fecha_valo: Indica cuándo se registró ese valor (ej: fecha de publicación o de venta real). Esto es útil para ajustar el modelo a cambios de mercado en el tiempo.

3. Ejemplos de filas

operación	región	comuna	latitud	longitude	habitaciones	baños	m2_construidos	m2_terreno	tipo_propiedad	valor_CLP
SALE	Metropolitana	Santiago	-33.4489	-70.6693	3	2	85	120	Departamento	150000000
RENTAL	Valparaíso	Viña del Mar			2	1	60		Departamento	650000
SALE	Biobío	Concepción	-36.8270	-73.0503	4	3	140	300	Casa	210000000

Observa que en la segunda fila latitud y longitud están vacías, pero igual es válida porque tiene región y comuna.

4. Recolección y registro de datos

- Completa una fila por cada propiedad histórica.
- Si no tienes un dato, déjalo en blanco o usa NULL.
- Revisa que los números sean coherentes (ej: una casa no puede tener 0 m² construidos).

5. Limpieza y validación

- Elimina duplicados.
- Corrige errores obvios.

- Valida con otras fuentes cuando sea posible.
-

6. Guardado del archivo

- Guarda la tabla en CSV (.csv) o Excel (.xlsx).
 - Usa un nombre descriptivo, por ejemplo:
 - datos_propiedades_venta_2025.csv
 - arriendos_santiago_2025-08.xlsx
-

Siguiendo estas recomendaciones, podrás crear una tabla de datos lista para entrenar un modelo predictivo de valor de propiedades en la plataforma.