



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
INSTITUTO DE INGENIERÍA MATEMÁTICA Y COMPUTACIONAL
IMT2220 – INTRODUCCIÓN A CIENCIA DE DATOS

Actualización Final del Repositorio

Tasación automática de propiedades

Grupo 23 — Javier Cuitiño, Simón Saravia, Alejandro Orellana

20 de noviembre de 2025

Resumen del proyecto

El proyecto se centra en la tasación automática de propiedades. En cuanto a los resultados, el modelo mostró un desempeño sólido en la estimación de precios de casas, alcanzando niveles de precisión entre 92 % y 99 % al compararlo con valores reales de prueba.

Sin embargo, el rendimiento fue significativamente inferior en el caso de departamentos. Esto se debe principalmente a que, durante el proceso de extracción de datos, no se incorporaron variables determinantes para este tipo de propiedad, como orientación, piso, entre otras características que influyen fuertemente en su valoración y que difieren de las variables clave para casas.

Como retroalimentación general, se concluye que es necesario ampliar y mejorar la recolección de datos de departamentos, junto con diseñar un enfoque de scraping específico para este segmento, de manera que se capturen las variables relevantes y se pueda mejorar el desempeño del modelo en futuras iteraciones.

Análisis de datos

La principal herramienta utilizada fue el web scraping, aplicado para extraer información desde portalinmobiliario.com. Adicionalmente, se implementó un proceso de geolocalización para obtener la latitud y longitud asociadas a cada dirección registrada.

Durante la etapa de preparación de los datos, se realizó una limpieza exhaustiva que incluyó la eliminación de outliers, valores nulos y registros con valores inconsistentes, como aquellos expresados en distintas unidades monetarias. Asimismo, no se consideraron variables como el título de la publicación ni la dirección completa, trabajando únicamente con las coordenadas geográficas para reducir el ruido en el modelo.

Todo este proceso buscó optimizar las condiciones de entrada para que la regresión lineal

operara con la menor cantidad de variabilidad irrelevante y pudiera entregar estimaciones más confiables.

Resumen de los resultados

Finalmente, el modelo alcanzó un R^2 de 0,9 tanto en entrenamiento como en prueba, lo que representa un desempeño altamente satisfactorio. En cuanto a la predicción de precios de casas dentro de la Región Metropolitana, los resultados fueron especialmente sólidos, llegando a ser comparables e incluso superiores en algunos casos a herramientas comerciales como Propiteq.

Respecto a la pregunta inicial, la respuesta es : sí es posible predecir tasaciones de propiedades de manera automática. En cuanto a la identificación de comparables, estos pueden obtenerse de forma efectiva mediante un enfoque de regresión KNN.

En conclusión, el programa desarrollado demostró ser un método confiable para estimar el valor de casas dentro de la Región Metropolitana, siempre que cuente con las variables adecuadas y un proceso de datos bien estructurado.

¿Qué podría salir mal?

Entre los posibles puntos débiles del modelo, uno de los principales riesgos es que en ciertos sectores los precios de proyectos nuevos y ventas “en verde” pueden distorsionar los valores reales de la zona, generando estimaciones menos precisas. Además, en comunas donde existe escasez de datos, por ejemplo, casos con menos de 100 viviendas disponibles el modelo puede producir predicciones poco representativas o alejadas del valor real.

La solución a estos problemas requiere una extracción de datos más extensa y profunda, que permita aumentar la cantidad y diversidad de registros. Asimismo, es fundamental actualizar los datos de manera periódica, de modo que el sistema refleje la evolución del mercado inmobiliario y no se vuelva una herramienta desactualizada u obsoleta.