



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

**FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES**

**DOBLE GRADO EN GESTIÓN Y ADMINISTRACIÓN PÚBLICA - ECONOMÍA**

**TRABAJO DE FIN DE GRADO**

**TÍTULO: Eco-RETINA: Inteligencia Artificial Verde para la Evaluación Inmobiliaria  
en Mercados Altamente Regulados. Evidencia desde Singapur.**

**AUTOR: SEBASTIÁN SIFONTES VALENTÍN**

**TUTORES: TEODOSIO PEREZ AMARAL**

**VÍCTOR MANUEL GONZALO ANGULO**

**CURSO ACADEMICO: 2024/2025**

**CONVOCATORIA: JUNIO**

## ÍNDICE:

1.	RESUMEN: .....	3
1.1.	ABSTRACT: .....	3
2.	INTRODUCCIÓN: .....	4
3.	GREEN AI Y MODELOS PREDICTIVOS EN ECONOMÍA: ECO-RETINA. ....	5
3.1.	Fundamentos de Green AI: .....	5
3.2.	El procedimiento de Eco-RETINA: .....	6
3.3.	Eco-RETINA y Green AI: .....	8
4.	MARCO TEÓRICO: .....	9
4.1.	Fundamentos del mercado inmobiliario de Singapur: .....	9
4.2.	Teoría de precios hedónicos y valoración de activos (Singapur): .....	12
5.	MODELO Y TÉCNICAS ECONOMETRICAS: .....	13
5.1.	Técnicas econométricas: Precios de reventa de Viviendas en Singapur. ....	13
5.2.	Modelos de Machine Learning y el mercado inmobiliario de Singapur: .....	15
5.3.	Modelo resultante de Eco-Retina: .....	17
6.	DATOS Y PREPROCESAMIENTO: .....	18
6.1.	Fuentes de Datos: <i>Datos de Singapur (HDB, variables contractuales)</i> . ....	18
7.	RESULTADOS: .....	21
7.1.	Precios de Reventa de Viviendas Públicas en Singapur .....	22
7.2.	Resultados adicionales: Eco-Retina y comparativa con métodos tradicionales: .....	26
7.3.	Resultados adicionales: Eco-Retina comparado con Machine Learning models: .....	30
8.	CONCLUSIONES: .....	37
9.	BIBLIOGRAFÍA: .....	39
10.	ANEXO-EL ESTIMADOR Y LA INTERFAZ DE ECORETINA .....	42
10.1.	¿Cómo fijar el entrenamiento de Eco-RETINA?: .....	42
10.3.	Construcción de la Primera Etapa IV-2SLS: .....	45
10.4.	Caracterización de los modelos de Machine Learning: .....	45
10.5.	Cálculo del máximo y el mínimo locales de los años residuales del contrato: .....	48

## 1. RESUMEN:

Este Trabajo analiza la dinámica del mercado de vivienda pública en Singapur, donde casi el 80% de los residentes viven en propiedades bajo contratos de 99 años. Utilizando Eco-RETINA -un algoritmo alineado con la IA Verde-, se examinan los determinantes de los precios de reventa destacando la relación no lineal entre los años residuales del contrato y los precios de reventa. Los resultados revelan un patrón cúbico de depreciación, con caídas aceleradas de precios al acercarse el vencimiento complementados por interacciones clave que marcan primas por bloque y altura de la vivienda, complementando a los métodos econométricos utilizados, así como a los modelos de machine learning de “caja negra”. Eco-RETINA logra precisión predictiva comparable a modelos de machine learning como LightGBM o Random Forest, aunque con una reducción de emisiones computacionales en más de un 70% en comparación. Este enfoque no solo mejora la comprensión de mercados regulados, sino que sugiere caminos concretos para un desarrollo económico ambientalmente sostenible, aplicable a políticas públicas en contextos urbanos densos.

Palabras clave: Vivienda Pública, Eco-RETINA, IA Verde, años residuales, Singapur.

### 1.1. ABSTRACT:

This paper analyses the dynamics of the public housing market in Singapore, where almost 80% of residents live in properties under 99-year contracts. Using Eco-RETINA - an algorithm aligned with Green AI - it examines the determinants of resale prices by highlighting the non-linear relationship between residual contract years and resale prices. The results reveal a cubic pattern of depreciation, with accelerated price drops as maturity approaches complemented by key interactions marking block and house height premiums, complementing the econometric methods used, as well as ‘black box’ machine learning models. Eco-RETINA achieves predictive accuracy comparable to machine learning models such as LightGBM or Random Forest, albeit with a reduction of computational emissions by more than 70% in comparison. This approach not only improves the understanding of regulated markets but also suggests concrete pathways for environmentally sustainable economic development, applicable to public policies in dense urban contexts.

Keywords: Public housing, Eco-RETINA, Green AI, residual years, Singapore.

## 2. INTRODUCCIÓN:

El mercado inmobiliario es uno de los pilares de la estabilidad económica, así como del bienestar social, requiriendo modelos predictivos sólidos para orientar políticas públicas y decisiones de inversión. En entornos altamente regulados como en Singapur se presenta crucial comprender las dinámicas de precios con el objetivo de evitar crisis que afecten la posibilidad de los ciudadanos de un país acceder a una vivienda y manejar con equilibrios estables las especulaciones de mercado.

Sin embargo, el creciente uso de análisis predictivos tanto en economía como en otras disciplinas ha generado un dilema entre precisión y sostenibilidad generado por las externalidades negativas provocadas por su actividad de explotación, tal y como ha sucedido a lo largo de la historia con el crecimiento de los sectores de producción tradicionales. Este dilema trata debates más amplios en Inteligencia Artificial (IA), donde la búsqueda de exactitud suele ignorar la eficiencia energética, una brecha que marcos recientes como la IA Verde (Green AI) buscan resolver.

El mercado inmobiliario de Singapur ofrece un contexto excepcional para estudiar la formación de precios bajo alta intervención estatal. Cerca del 80% de la población habita viviendas públicas del Housing Development Board (HDB), adquiridas bajo contratos de “compra” de 99 años que, al expirar, devuelven la propiedad al Estado.

Es un modelo diseñado para combatir la especulación y garantizar el acceso a la vivienda con una variable particular en la formación de los precios: los años residuales del contrato (`lease_years`). A esto se suma un desafío demográfico: Singapur ha pasado de tener una densidad poblacional de unas 2.831 personas/km<sup>2</sup> en 1950 a 8.207 en 2024 (SingStats, 2024), enfrentando una escasez creciente de suelo edificable, intensificando la presión sobre los precios de reventa, cuestionando la forma de modelar con precisión e interpretabilidad la relación entre años residuales del contrato y los precios de reventa.

En este sentido la literatura actual sobre valoración inmobiliaria en mercados regulados combina métodos econométricos tradicionales basado en la teoría de los precios hedónicos y la valoración de activos, buscando capturar la aceleración de la depreciación conforme el contrato se acerca a su vencimiento. No obstante, la exploración de interacciones entre variables clave como la ubicación geográfica y los años residuales es incipiente, y la selección de predictores son manuales o ad hoc en un contexto donde no

se evalúan los costos ambientales asociados a los métodos computacionales utilizados generando un vacío metodológico donde el enfoque de Eco-RETINA resulta oportuno para compensarlo.

Por otro lado, las técnicas de machine learning como LightGBM y Random Forests según Ke et al., 2017 o Phang y Wong, 2016, han demostrado superioridad para modelar relaciones complejas en los mercados inmobiliarios, aunque su naturaleza intrínseca de “caja negra” dificulta la interpretación económica, además de que su entrenamiento genera huellas de carbono significativas ampliamente detalladas por Strubell et al., 2019, derivando en que nos preguntemos lo siguiente: ¿cómo lograr modelos transparentes e interpretables sin comprometer su capacidad predictiva o la sostenibilidad?.

Este trabajo busca responder a esta pregunta mediante Eco-RETINA, un algoritmo alineado con el enfoque Green AI que combina la transformación automática de variables con la selección parsimoniosa de modelos para ofrecer una alternativa sostenible o complementaria a la econometría y al ML intensivo en recursos. Con ello, la aportación de este TFG pasa primero por demostrar cómo Eco-RETINA reconcilia la interpretabilidad de los modelos econométricos con la flexibilidad del Machine Learning. Luego se discuten los hallazgos y aportaciones sobre el mercado inmobiliario singapurense, destacando la relación cúbica entre los años residuales del contrario y los precios de reventa, idea recientemente adoptada en 2023 por Bingqiao et al., 2023.

Finalmente, se establece un marco replicable para investigaciones energéticamente eficientes en economía basada en datos gracias a la comparación del rendimiento de Eco-RETINA con marcos de trabajo del Machine Learning, subrayando la viabilidad de los enfoques de Green AI para abordar desafíos socioeconómicos complejos y motivando a que se aborden futuros estudios en mercados regulados de alto impacto utilizando Eco-RETINA.

### **3. GREEN AI Y MODELOS PREDICTIVOS EN ECONOMÍA: ECO-RETINA.**

#### **3.1. Fundamentos de Green AI:**

El objetivo de este apartado es la de contextualizar el concepto de Green AI para fundamentar todas las comparaciones que se realizarán con Eco-RETINA en este TFG. En concreto, Schwartz et al., 2020, explican cómo la Inteligencia Artificial Verde (Green

AI) surge debido a la necesidad de mitigar el impacto ambiental de los sistemas de IA, basándose tanto en criterios de sostenibilidad como de eficiencia computacional, englobando múltiples marcos de desarrollo.

Esto implica que ya no basta con entrenar el “mejor” modelo posible, sino además debe hacerse empleando los menores recursos posibles, sin sacrificar en exceso la calidad de la solución, contrastando con el enfoque de Red AI, que se centra exclusivamente en lograr el máximo rendimiento sin priorizar el coste energético o las emisiones asociadas.

Dentro de los marcos metodológicos de Green AI, Bolón-Canedo y Alonso-Betanzos, 2019, destacan la selección de características como estrategia clave para que un modelo requiera menos operaciones durante el entrenamiento y la inferencia, al mismo tiempo en el que previene el sobreajuste al demandar menos variables, minimizando la demanda computacional de dichos modelos ahorrando energía haciendo del enfoque Green AI un principio transversal para desarrollar y desplegar soluciones útiles al menor coste energético posible, alineándose con los objetivos globales de reducción de emisiones, así como de eficiencia de recursos.

### **3.2. El procedimiento de Eco-RETINA:**

Por su parte, Eco-RETINA se conceptualiza como un algoritmo de modelado automático que surgió como complemento al algoritmo de Relevant Transformation of the Inputs Network Approach (RETINA) desarrollado por Pérez Amaral, Gallo y White en 2003, motivado por profundizar tanto en el enfoque de Green AI, como de extender y adaptar las ya potentes características del algoritmo predecesor, RETINA. A continuación, pasaremos a explicar brevemente lo detallado por el desarrollador Capilla, 2024, sobre el funcionamiento de Eco-RETINA:

Eco-RETINA se basa en la selección de modelos de regresión a través de un proceso de búsqueda selectiva e iterativa con el objetivo de encontrar un modelo con buen rendimiento predictivo, al tiempo que minimiza su complejidad para arrojar resultados interpretables con una reducción significativa de su consumo energético. Es posible en primer lugar gracias a la transformación de variables que el usuario le introduzca, generando un subconjunto de transformaciones, por ejemplo,  $\zeta(X) = \{W_1, \dots, W_m\}$ , donde  $\{W_1, \dots, W_m\}$  son todas aquellas variables transformadas que se consideran óptimas para construir el modelo.

Esta transformación de variables incluye potencias prefijadas por el usuario, que pueden ser desde cuadradas e inversas hasta cúbicas o de raíz cuadrada, y para las variables categóricas sólo se consideran interacciones entre la transformación de esas categóricas a variables ficticias (en adelante, ficticias) y las variables continuas, permitiendo capturar relaciones no lineales sin agotar el espacio de transformaciones posibles. Una vez definido este subconjunto:

- Evita el sobreajuste de este subconjunto prefijado dividiendo los datos originales en hasta tres submuestras o pliegues, siguiendo la lógica de validación cruzada para entrenar y evaluar de Stone (1974) y Geisser (1975) para los primeros dos pliegues, incluyendo de forma opcional la tercera división para validar la elección final.
- Sigue con ordenar el subconjunto  $\{W_1, \dots, W_m\}$  según su correlación absoluta con la variable objetivo/explicada (Y) al agregar iterativamente hasta la variable  $W_m$  más correlacionada siempre que su colinealidad con las ya incluidas sea menor que un umbral  $\lambda$  predefinido también por el usuario entre 0 y 1 con un máximo de 50 valores (lo que entendemos como rejillas) gracias a las operaciones vectorizadas de la librería Numpy. Como resultado, la lista del subconjunto queda ordenada como  $W_1, W_2, \dots, W_m$ , donde  $W_1$  es el elemento más correlacionado con Y.
- No obstante, antes de incluir una nueva variable, llámese  $W_j$ , hasta  $W_m$  (la última variable del subconjunto), se evalúa cuánta información de esa nueva variable  $W_j$  ya está explicada mediante la regresión de  $W_j$  sobre las ya seleccionadas, al calcular el coeficiente de determinación  $R^2$ .
- Por ejemplo, si intentas “predecir” la nueva variable usando las que ya tienes, ¿cuánta varianza de la nueva explica ese minimodelo?, filtrando aquellas variables que ya están aportando la misma información ya recogida en la lista, así no duplicas información y reduces el gasto energético. Este ejemplo se define por parte de Capilla, 2024 como todo  $R^2 \leq \lambda$ .
- Estima el modelo en la primera submuestra, para luego calcular el rendimiento del modelo mediante el error cuadrático medio (MSE), el error absoluto medio porcentual (MAPE) o criterios de información AIC o BIC según el usuario escoja, que se aplicarán sobre la segunda submuestra y así quedarse con el subconjunto capaz de minimizar el criterio seleccionado para ese segundo pliegue de testeo.

- La tercera submuestra opcional funciona a modo de parsimonia del modelo al evaluar cada submodelo en la partición 3 y así escoger el más parsimonioso cuyo rendimiento no empeore de manera significativa.
- Repite los pasos anteriores permutando los tres pliegos, quedándose con el modelo de mejor desempeño promedio.

Cabe destacar como valor añadido del algoritmo que este integra el paquete CodeCarbon de Python para registrar el consumo de CPU/GPU, así como las emisiones de  $CO_2$ , asociándolas a las métricas de rendimiento del modelo para así poder priorizar las especificaciones de baja huella ambiental. Así mismo, el usuario tiene la posibilidad de fijar según sus necesidades concretas:

- Modelos de clasificación logit, probit y/o MLP (Modelo Lineal de Probabilidad).
- Covarianza robusta tipo HC0-HC3 y ejecución paralela de iteraciones.
- Búsqueda automática del umbral  $\lambda$  en función del tamaño máximo deseado del modelo (de variables que el algoritmo seleccione), controlando la parsimonia sin sacrificar precisión.

Este trabajo busca exponer cómo el algoritmo es un marco de trabajo que aúna la adaptabilidad de las redes neuronales, la solidez e interpretabilidad de los modelos lineales al mantener linealidad en parámetros, y la filosofía de Green AI para generar modelos precisos con responsabilidad ecológica.

### **3.3. Eco-RETINA y Green AI:**

La IA ya no puede evaluarse únicamente por su precisión. En un contexto global donde el consumo energético de modelos predictivos se vuelve crítico, la necesidad de enfoques “verdes” cobra relevancia. En este marco, Eco-RETINA representa una propuesta híbrida que reconcilia eficiencia, interpretabilidad y responsabilidad ambiental.

Según Raman et al., 2024, a pesar de los llamamientos a la transparencia metodológica en la literatura en aprendizaje automático y econometría, apenas se estandariza la medición de emisiones o se reportan huellas de carbono asociadas al entrenamiento, perpetuando la “caja negra” en torno al costo ambiental de los modelos.

Eco-RETINA adopta por un lado la búsqueda selectiva de predictores (reduciendo la dimensionalidad), para garantizar que cada variable incorporada aporte información novedosa, mientras que por otro lado integra de forma nativa la medición de consumo



energético y las emisiones de  $CO_2$  mediante CodeCarbon, cumpliendo con la demanda de transparencia y estandarización para proponer un algoritmo capaz de evadir la categorización de “caja negra”, haciéndolo compartir el mismo principio rector de la Green AI y sus diversos enfoques.

## **4. MARCO TEÓRICO:**

En este apartado se ofrece una perspectiva teórica del caso práctico y de los métodos predictivos seleccionados para explicar los precios de reventa de vivienda pública en Singapur, justificando a su vez su elección como caso práctico de interés para probar Eco-RETINA y su comparación con los resultados de los métodos predictivos. Luego, se define la teoría de precios hedónicos y la teoría del valor presente ajustada por riesgos como bases para luego construir los modelos econométricos de panel.

### **4.1. Fundamentos del mercado inmobiliario de Singapur:**

El contexto analizado del mercado inmobiliario de Singapur se considera “sui generis” debido a la creciente demanda de espacios de vivienda para una densidad de suelo cada vez más limitada. Por ello, el gobierno de Singapur ha aplicado desde inicios del siglo XX una política concreta para intentar frenar la especulación en precios y una consecuente crisis de oferta de vivienda sobre los precios.

Consistió en primera parte, en financiar con fondos públicos la construcción y desarrollo de las viviendas mediante el Housing Development Board (HDB), derivando en que para el año 2024, casi un 80% de la población local residiese en viviendas públicas gestionadas por el HDB (Vivienda Pública en Singapur, ICEX, 2024), y que el 92,2% sea propietaria de su vivienda.

**Figura 1: Evolución de la situación de mercado de propiedad inmobiliario en Singapur**

	2024	2023	2022	2021
Número total de viviendas HDB	1.132.100	1.108.100	1.090.100	1.089.100
Población local en viviendas HDB	77,4 %	77,8 %	77,9 %	78,3 %
Viviendas HDB en propiedad	92,2 %	91,3 %	90,9 %	90,9 %
Nuevas viviendas HDB construidas	19.600	22.800	23.200	17.100
Índice de precios de reventa <sup>1</sup>	197,9	180,4	171,9	155,7
Media de habitantes por vivienda HDB	3,01	3,04	3,02	3,09
Población local <sup>2</sup> total	4.180.868	4.149.253	4.073.239	3.986.842
Población total	6.036.860	5.917.648	5.637.022	5.453.566

Fuente: Housing and Development Board, Building and Construction Authority, Singapore Statistics.

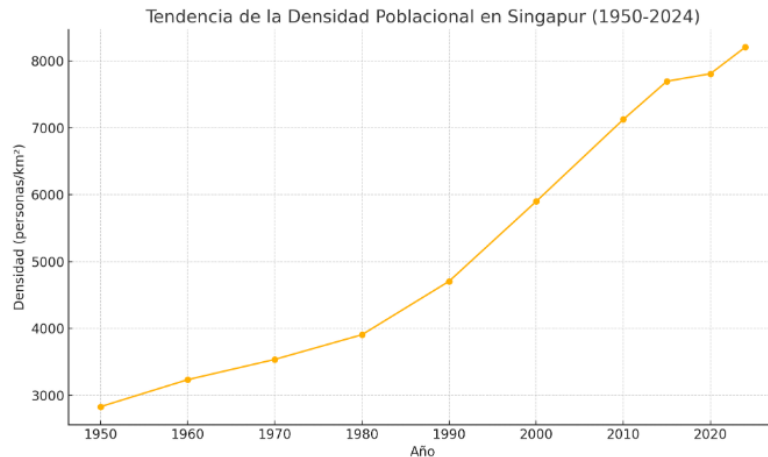
En segundo lugar, estas viviendas son traspasadas a los ciudadanos a precios accesibles (subsidiados) bajo contratos de arrendamiento de 99 años en los que el gobierno retiene la titularidad del suelo, entendiéndose que los compradores adquieren un derecho de uso temporal donde los residentes pueden vivir, heredar o vender la vivienda bajo ciertas normas del gobierno que raramente renueva el contrato una vez vence.

Los requisitos para vender pasan por esperar al menos 5 años desde la adquisición aparte de cumplir con ciertas restricciones, por ejemplo, el comprador en ciertos casos debe ser de un grupo étnico que mantenga el equilibrio racial del edificio. El alquiler de habitaciones es posible, siempre y cuando el titular temporal de la vivienda siga viviendo allí.

La aplicación de Eco-RETINA sobre este caso práctico cobra sentido al tratarse de un mercado donde los precios de reventa de los inmuebles se explican por razones poco comunes en comparación con otros mercados inmobiliarios a nivel mundial, como por ejemplo, los años residuales del contrato. Es complicado encontrar otro mercado con este alcance en su programa de vivienda pública con casi un siglo de implementación y funcionamiento, siendo aparte el aumento de la presión demográfica un problema creciente para Singapur:

Finalmente, para reflejar la presión demográfica de Singapur a modo de contextualización, se ilustra la siguiente tendencia desde 1950 hasta el año 2024:

**Figura 2: Evolución 1950-2024 de la densidad poblacional singaporense**



**Fuente: Elaboración propia con datos del Singapore Statistics.**

Dando valor añadido a los resultados obtenidos de cara a las posibles aplicaciones prácticas de políticas públicas de vivienda HDB.

Así mismo, se busca exponer la aplicabilidad del algoritmo sobre un análisis de datos de panel ya que actualmente no está preparada nativamente para paneles, sigue siendo aprovechable su capacidad para encontrar transformaciones de variables relevantes con aportaciones novedosas.

Además, la literatura existente sobre el mercado de Singapur como Ong et al., 2003 o Tan, 1990, suelen modelar la variable de “años residuales del contrato” de forma lineal o por semilogaritmos, rara vez explorando interacciones no lineales, o trabajos como Li et al., 2012 o Hoo, 2005, incorporan técnicas de efectos fijos o dinámicos, sin aplicar estrategias de selección modular de predictores o midiendo el coste energético del modelado.

En trabajos más recientes, como Bingqiao et al., 2023, reconocen partiendo de la misma lógica econométrica aquí propuesta en cuanto al análisis de panel, centran sus resultados obtenidos sobre la transformación no lineal de los años residuales del contrato de vivienda, aunque sin tomar otras transformaciones o interacciones entre variables en consideración.

En consecuencia se estima que Eco-RETINA con su transformación de variables descubra relaciones no evidenciadas hasta ahora, así como fijar un marco de trabajo para futuros estudios donde no se hayan aplicado hasta ahora enfoques de Green AI.

## 4.2. Teoría de precios hedónicos y valoración de activos (Singapur):

En este mercado de alojamientos hay variable crítica en la formación de precios: los años residuales del contrato, es decir, el tiempo restante hasta su vencimiento. A diferencia de mercados tradicionales, donde la propiedad es perpetua, en Singapur el valor de una vivienda no solo depende de atributos físicos típicos como el área construida, sino también de la seguridad temporal que ofrece el plazo remanente del contrato. Es, por tanto, un mercado altamente regulado cuyos precios responderían a: i) la teoría de precios hedónicos (Rosen, 1974), y ii) la valoración de activos con restricciones temporales (Geltner et al., 2018).

En concreto, según Rosen (1974), el precio de una vivienda ( $P_i$ ) es la suma de los valores marginales de sus atributos observables:

$$P_i = \sum_{j=1}^k \beta_j X_{ij} + \epsilon_i \quad (1)$$

En mercados con significativa homogeneidad arquitectónica, como el de la HDB, el área construida ( $X_i$ ) emerge como el principal determinante hedónico, donde estudios en contextos regulatorios similares, como Malpezzi (2002), documentan elasticidades a la superficie o elasticidades-precio del área construida superiores a 0,7 en algunos casos, sugiriendo que incrementos marginales en el espacio habitable son altamente valorados en entornos de escasez. Esta elasticidad de la superficie conlleva que, *ceteris paribus*, un aumento del 1% en el área construida debería elevar el precio de reventa en 0,7%.

Por otra parte, los contratos de arrendamiento de 99 años introducen un componente dinámico en la valoración, relación formalizada por el modelo de valor presente ajustado por riesgo (Geltner et al., 2018):

$$P = \sum_{t=1}^L \frac{E[R_t]}{(1+r)^t} + \frac{E[V_L]}{(1+r)^L} \quad (2)$$

donde  $L$  representa los años residuales del contrato,  $E[R_t]$  los flujos de alquiler implícitos, y  $E[V_L]$  el valor residual esperado, donde para Singapur en concreto, la no renovabilidad de los contratos genera una relación no lineal entre los años residuales

del contrato ( $L$ ) y el precio ( $P$ ): a medida que  $L$  disminuye, la incertidumbre sobre la revalorización se amplifica, depreciando el precio de forma acelerada.

## 5. MODELO Y TÉCNICAS ECONOMETRICAS:

### 5.1. Técnicas econométricas: Precios de reventa de Viviendas en Singapur.

En esta sección se aplica el marco teórico anterior en modelos econométricos concretos, discutiéndose las propiedades estocásticas de las ecuaciones, se justifican las técnicas de estimación seleccionadas, cerrando con explicar cómo los datos se integran en el diseño metodológico.

Desarrollando las ecuaciones (1) y (2) dentro del contexto del estudio, se opta en este caso por el modelo de efectos fijos (FE) sobre el de efectos aleatorios (RE) al prever una correlación entre heterogeneidad no observada y los regresores del análisis que invalidarían los resultados de los RE.

Más concretamente, las características invariables en  $t$  de cada bloque de viviendas como la calidad de los acabados, la tipología de la ubicación, o la cercanía al transporte público influyen a la vez en el precio y en variables como el área en  $m^2$  o los años residuales, por lo que para este mercado, así como los datos disponibles en concreto, se optó por especificar FE bajo la siguiente fórmula:

$$\ln(P_{it}) = \alpha_i + \beta_1 \ln(S_{it}) + \beta_2 L_{it} + \beta_3 L_{it}^2 + \beta_4 L_{it}^3 + \gamma X_{it} + \delta_t + \epsilon_{it} \quad (3)$$

Donde:

- $\ln(P_{it})$ : Logaritmo neperiano del precio de reventa de la vivienda  $i$  en el mes  $t$ .
- $\ln(S_{it})$ : Logaritmo neperiano del área construida (*floor\_area\_sqm*).
- $L_{it}, L_{it}^2, L_{it}^3$ : Años residuales de arrendamiento y sus términos polinómicos.
- $X_{it}$ : Variables de control (ej: flat type, ubicación geográfica en efectos fijos por bloque).
- $\delta_t$ : Efectos temporales mensuales para capturar tendencias macroeconómicas.
- $\alpha_i$ : Efectos fijos por bloque de viviendas para controlar heterogeneidad no observada (ej: proximidad al transporte público).
- $\epsilon_{it}$ : Término de error con  $E[\epsilon_{it} | X_{it}] = 0$ .

Es preciso comentar que este modelo enfrenta y aborda tres fuentes potenciales de sesgo en el análisis de precios de reventa de HDB: Primero, la dependencia espacial, ya que las viviendas situadas en bloques (de viviendas) adyacentes suelen compartir características, muchas no observadas, que introducen correlación en los errores. Para mitigarlo, se aplicaron errores estándar clusterizados por bloque y periodo que según Cameron y Miller, 2015, brinda robustez tanto ante heterogeneidad de varianza como a correlación dentro de cada clúster o grupo de viviendas.

Segundo, la heteroscedasticidad, aunque la transformación logarítmica de las variables suele estabilizar la varianza, la dispersión de precios puede seguir siendo función del tamaño de la vivienda, en este caso  $\ln(S_{it})$ . El uso de covarianza robusta corrige cualquier posible variación no uniforme de los residuos. Finalmente se incluye un modelo complementario que busca capturar inercia temporal y validar la robustez de las relaciones observadas en el modelo estático mediante las primeras diferencias.

Para ello partimos de la ecuación FE en niveles (3) para pasarla a su versión en primeras diferencias (FD) y así eliminar los efectos fijos no observados  $\alpha_i$ , ofreciendo un modelo alternativo al de los efectos fijos o aleatorios y capturando la evolución mensual de los precios:

$$\Delta \ln P_{it} = \beta_1 \Delta \ln S_{it} + \beta_2 \Delta L_{it} + \beta_3 \Delta L_{it}^2 + \beta_4 \Delta L_{it}^3 + \Delta \varepsilon_{it} \quad (4)$$

Los rezagos son de carácter mensual y capturan la persistencia entre los precios mes a mes consecutivamente. Como resultado el nuevo término de error  $\Delta \varepsilon_{it}$  esté correlacionado con esta expresión, por lo que la tratamos como endógena. Este enfoque permite valorar cómo influyen los cambios mensuales sobre la variación del precio.

Por otra parte, se exploró la posibilidad de incluir un modelo con término autorregresivo, no resultó estadísticamente significativo, optando por desestimarlos de los resultados. Esta decisión se basa en que dicho término no resultó eficaz en capturar la reversión a la media al ser un fenómeno que no puede modelarse adecuadamente con una sola ecuación autorregresiva sin información adicional sobre el equilibrio estructural del sistema, además de que su inclusión no mejora sustancialmente la precisión del modelo. Este procedimiento es consultable en el apartado 10.2. del Anexo.

Así, se optó por una especificación más sencilla que permite valorar directamente el efecto de los cambios mensuales en los años residuales del contrato y el área construida sobre la variación mensual del precio de reventa.

Por ello el objetivo con los resultados econométricos, aparte de obtener conclusiones valiosas sobre las dinámicas de mercado per se, es evaluar el potencial de modelos como Eco-RETINA para complementar o incluso desafiar metodologías estructuradas de panel, especialmente cuando se carece de información longitudinal explícita.

## **5.2. Modelos de Machine Learning y el mercado inmobiliario de Singapur:**

Durante décadas, los economistas hemos confiado en modelos lineales y técnicas tradicionales, como los análisis de panel o los modelos de efectos fijos, enfoques que aunque robustos y eficientes, suelen chocar con limitaciones cuando se enfrentan a relaciones complejas o no lineales, en parte por su menor capacidad de adaptarse a los datos disponibles que en la práctica es frecuente, e indirectamente se busca reflejar en el caso de estudio TFG sobre los mercados inmobiliarios altamente regulados como el de Singapur.

Aquí, métodos más flexibles relacionados con el Machine Learning (ML) como las redes neuronales o el gradient boosting (LightGBM), han ganado terreno por su capacidad para capturar estas dinámicas no lineales en contextos de alta dimensionalidad o grandes volúmenes de datos, formando una nueva vertiente de análisis en la disciplina económica a la hora de explicar o predecir los fenómenos de interés.

Es si bien un cambio que se ha traducido en un gran progreso, hoy en día y en el futuro cada vez más presente va a estar el nuevo gran desafío no ajeno a otras disciplinas de estudio, el equilibrar la precisión con la sostenibilidad ambiental, ya que acarrear un costo oculto y complicado de dimensionar.

El entrenamiento de modelos complejos, como redes neuronales profundas, genera una huella de carbono significativa. Strubell et al. (2019) estiman que entrenar un modelo de búsqueda de arquitectura neuronal (NAS) emita hasta 284 toneladas métricas de CO<sub>2</sub>, equivalente a las emisiones de 5 automóviles durante su ciclo de vida

útil, resaltando la urgencia de adoptar enfoques como la Green AI que prioricen la eficiencia energética sin sacrificar rendimiento.

En el caso de este TFG, utilizaremos los modelos de ML para compararlos estrictamente con el rendimiento ofrecido por Eco-RETINA tanto en precisión medida desde términos como MSE, MAPE o  $R^2$ , la interpretabilidad de los resultados obtenidos sacando las variables más relevantes según cada modelo, para finalizar con la comparativa del nivel de emisiones, poniendo a prueba desde todos los puntos posibles al algoritmo si se complementa con las técnicas econométricas del apartado anterior. Desde luego, todos partiendo del mismo conjunto de datos de precios de reventa de viviendas públicas en Singapur.

### **5.2.1. Técnicas de Machine Learning aplicadas:**

En cuanto a los modelos de machine learning empleados en este estudio, LightGBM, Random Forest (no paramétricos) y ElasticNet (semiparamétrico), representan enfoques capaces de capturar relaciones complejas en los datos, superando varias limitaciones de los modelos econométricos tradicionales dentro de un contexto de alta dimensionalidad.

A continuación, se contextualiza en qué consiste cada uno de estos modelos, así como una breve justificación de su elección, sin entrar en los detalles técnicos de caracterización, que sin embargo, son consultables en el anexo; particularmente:

- **LightGBM (Gradient Boosting Machine):** Construido sobre el algoritmo de Gradient Boosting Machines de Friedman, 2001, LightGBM está diseñada para maximizar la eficiencia y la precisión en conjuntos de datos de gran tamaño y alta dimensionalidad (Ke et al., 2017). A diferencia de su antecesor, XGBoost, LightGBM optimiza el crecimiento de árboles al construir secuencialmente árboles de decisión poco profundos, orientando cada nuevo árbol a corregir los errores residuales del ensamblado anterior.

Dentro de los árboles aplica técnicas como el agrupamiento de características para acelerar el entrenamiento, y resulta idóneo para capturar las relaciones no lineales entre las variables de interés del caso práctico, sin necesidad de especificar a priori la forma funcional de dichas interacciones, alcanzando altos niveles predictivos con procedimientos optimizados.



- **Random Forest:** Según Breiman, 2001, este modelo ensambla decenas o cientos de árboles de decisión entrenados sobre muestras aleatorias del conjunto de datos y subconjunto de variables. Al promediar sus predicciones, reduce el sobreajuste de cada árbol individual mientras ofrece robustez ante ruido y outliers.

La heterogeneidad espacial del mercado inmobiliario de Singapur (por ejemplo, diferencias entre distritos más nuevos que otros), y la presencia de valores atípicos hace de Random Forest un modelo candidato a modelar sin sesgos estos efectos, al tiempo que proporciona una medida intuitiva de la importancia de cada predictor en la estimación de precios.

- **ElasticNet (Regresión penalizada):** Zou & Hastie, 2005, lo definen como un modelo lineal penalizado el cual fusiona la norma L1 (Lasso) y L2 (Ridge) para estimar modelos lineales en contextos de alta dimensionalidad y correlación entre predictores.

Este modelo ofrecerá en un único paso, un conjunto de regresores lineales relevantes que facilitarán la interpretación económica de los coeficientes asociados al modelo como tipo de vivienda o efectos temporales, actuando como un contrapunto claro a los modelos más flexibles de árboles, aunque no captura ciertas interacciones clave no lineales por ejemplo, haciendo que subestime cierta complejidad cuando se aplica.

### 5.3. Modelo resultante de Eco-Retina:

Es en ese punto donde por su parte, Eco-RETINA genera un modelo de la forma:

$$\ln(P_i) = \beta_0 + \sum_{k=1}^K \beta_k W_j(X_i) + \epsilon_i \quad (6)$$

Donde  $W_j$  denota el valor de la j-ésima transformación no lineales del subconjunto  $\zeta(X)$  óptimas seleccionadas de las variables originales  $X_i$  para la unidad i, por ejemplo, el área en m<sup>2</sup>, los años residuales, o interacciones entre estas variables.

Se busca que siga la misma lógica marcada por el marco teórico en cuanto a los precios hedónicos y el modelo de valor presente ajustado por riesgo, manteniendo la interpretabilidad de los resultados obtenidos, así como obtener las métricas de precisión necesarias para realizar la comparación con los modelos de machine learning.

## 6. DATOS Y PREPROCESAMIENTO:

En este apartado, se describe a detalle el preprocesamiento aplicado a los datos originales para poder luego aplicarlos en los análisis posteriores.

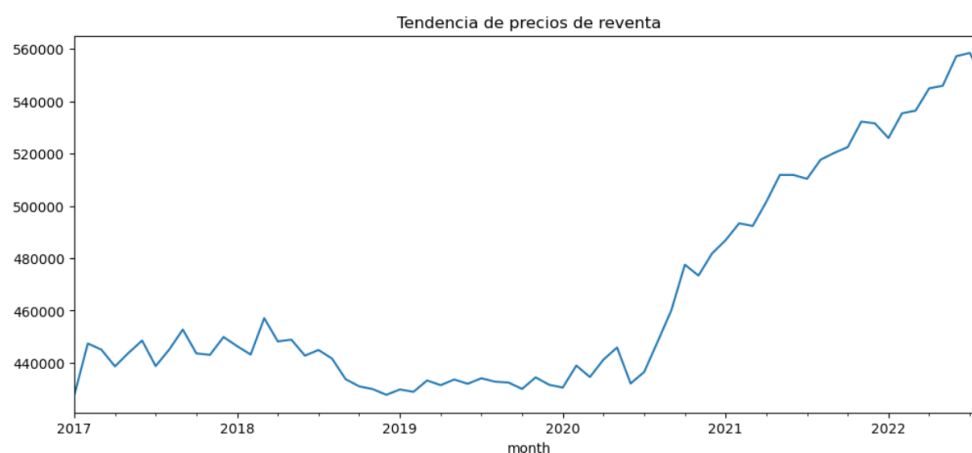
### 6.1. Fuentes de Datos: *Datos de Singapur (HDB, variables contractuales)*.

El estudio se basa en un conjunto de datos obtenido de registros administrativos del Gobierno de Singapur y disponible públicamente en la plataforma Kaggle bajo el repositorio "*singapore-housing-price-20172022*", incluyendo 134.168 transacciones de reventa de viviendas públicas (HDB) registradas entre enero de 2017 y diciembre de 2022.

Se abarcan un total de 26 distritos (*towns*) como Ang Mo Kio o Queenstown, donde las variables clave incluyen el precio de reventa (en SGD), el área construida ( $m^2$ ) y los años residuales del contrato de arrendamiento, este último convertido de formato textual ("*XX años YY meses*") a escala decimal mediante una transformación simple.

Los precios promediaron 470.666 SGD ( $\sigma=163.050$ ), con un crecimiento acumulado del 22.3% entre 2017 y 2022 interrumpido por una caída del 5,1% durante el segundo trimestre de 2020 por las restricciones pandémicas. El área construida mostró estabilidad ( $x=97,81 m^2$ ,  $\sigma=24,13$ ), mientras que los contratos residuales promediaron 74,82 años ( $\sigma=13,51$ ), con una disminución anual del 0,8%.

**Figura 3: Evolución de precios de reventa de viviendas públicas en Singapur**



Fuente: Elaboración propia con datos del Singapore Statistics.

Las correlaciones preliminares revelaron una relación entre el área y el precio ( $r=0,79$ ), una moderada asociación entre años residuales y precio ( $r=0,34$ ), así como la antigüedad del contrato correlacionando positivamente con los precios ( $r=0,62$ ).

Esta última se debe a la localización de las viviendas. Las que son más antiguas suelen ubicarse en zonas céntricas, como Queenstown o Toa Payoh, normalmente con más demanda, mientras que las viviendas nuevas suelen estar en áreas periféricas como Punggol o Sengkang, explicando el porqué de esta correlación.

Con esta información, se construye a partir de los datos un panel con la dimensión temporal mensual junto a los bloques de vivienda como variable espacial en aras de tener un sujeto comparable por meses en el panel. Se aplica una exclusión de bloques con menos de 24 meses con transacciones registradas fundamentalmente debido a una de las condiciones fijadas por el gobierno en el mercado relacionada con la tenencia mínima de 5 años en propiedad como vivienda habitual.

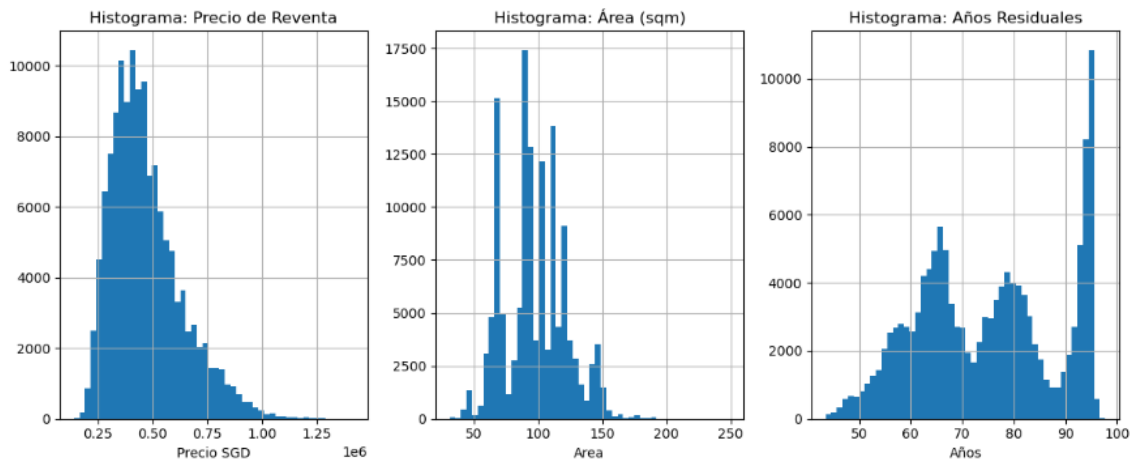
Como resultado, las transacciones de reventa tenidas en cuenta en el análisis final aseguran que los efectos temporales puedan estimarse sin sesgos por datos intermitentes de un bloque con actividad intermitente (por ejemplo, que haya tenido transacciones concentradas en un único año), o el capturar las dinámicas de mercado más activas por bloques de viviendas.

Esta base de datos analiza muestras en un periodo de 5 años, pasando de las 134.168 transacciones en 2.611 bloques originales a una base de datos filtrada de 71.276 observaciones en 1.257 bloques, con una media en los bloques incluidos de 27,3 meses de actividad, con un mínimo de 24 y un máximo de 68 para todo el periodo.

Siendo como resultado coherente con estudios similares de mercados inmobiliarios, donde aplican umbrales de entre 18-24 meses para paneles urbanos como Phang y Wong, 2016, o Deng et al., 2019.

En cuanto a las transformaciones de las variables, las variables de interés en la base de datos filtrada presentaban las siguientes distribuciones representadas por histogramas:

**Figura 4: Histogramas de las variables de interés**

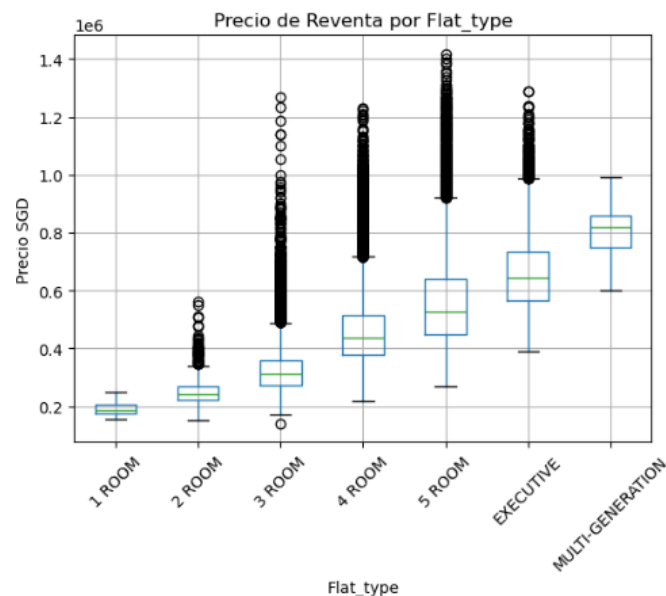


Fuente: Elaboración propia con datos del Singapore Statistics.

Justificando la decisión de explorar ciertas transformaciones no lineales para reducir asimetrías, optándose para los modelos manuales por transformaciones logarítmicas en precios de reventa y en el área en metros cuadrados de las viviendas, así como de las ya mencionadas relaciones cúbicas de los años residuales del contrato.

Además, mediante esta representación de precios de reventa por tipo de vivienda:

**Figura 5: Gráfica de cajas del precio de reventa (en SGD) por tipo de vivienda.**



Fuente: Elaboración propia con datos del Singapore Statistics.

Se decidió recortar las observaciones en el 1% superior de la distribución de precios, prescindiendo de un total de 712 transacciones atípicas que podrían distorsionar las elasticidades, mostrando las siguientes características:

**Tabla 1: Medias de las variables de interés según el número de habitaciones.**

<i>Flat Type</i>	<i>Precio Promedio</i> <i>(SGD)</i>	<i>Área Promedio</i> <i>(m<sup>2</sup>)</i>	<i>Años</i> <i>(Media)</i>	<i>Residuales</i>
<i>1 habitación</i>	193.112	31,0	54,2	
<i>2 habitaciones</i>	247.927	46,0	77,6	
<i>3 habitaciones</i>	327.471	68,2	65,4	
<i>4 habitaciones</i>	465.864	95,1	78,5	
<i>5 habitaciones</i>	563.589	117,8	77,9	
<i>Executive</i>	661.799	144,7	72,9	
<i>Multi-Generation</i>	809.355	161,1	66,9	

Mostrando como los precios y el área construida aumentan con el n° de habitaciones, mientras que los años residuales de arrendamiento son mayores en viviendas premium (Executive), sugiriendo una prima por seguridad contractual en segmentos de lujo.

Por su parte, se modificó la variable Storey\_range (número de planta de una vivienda) para facilitar la interpretabilidad. Se agruparon las plantas dentro de tres grandes grupos: low, para pisos desde la planta 0 hasta la 10, mid, para pisos desde la planta 11 hasta la 30, y finalmente high, para pisos que se encuentran a partir de la planta 30, creándose una nueva variable dentro de la base de datos llamada Storey\_group: low/mid/high según corresponda para cada planta.

Todo este análisis es replicable, la base de datos completa y los scripts de procesamiento están disponibles en [Kaggle](#) y un repositorio GitHub en el anexo.

## 7. RESULTADOS:

Este apartado destaca los resultados más relevantes obtenidos en primera instancia por los modelos econométricos, exponiendo únicamente sus resultados como se haría en cualquier investigación. Luego, se detalla lo obtenido con Eco-RETINA, explicando

punto por punto su comparación con los rendimientos de los modelos de Machine Learning, así como de las emisiones de  $CO_2$  por entrenamiento.

## 7.1. Precios de Reventa de Viviendas Públicas en Singapur

El estudio de los determinantes de los precios de la vivienda en mercados altamente regulados constituye un desafío teórico y empírico de relevancia global. En este apartado ahora se exponen los resultados más destacados obtenidos por los métodos econométricos de panel seleccionados.

Los coeficientes estimados, por efectos fijos de la **ecuación (3)** son:

$$\ln(P_{it}) = \alpha_i + 0,9834 \ln(S_{it}) + 0,1836 L_{it} - 0,0027 L_{it}^2 + 1,285 \times 10^{-5} L_{it}^3 + \delta_t + \epsilon_{it} \quad (7)$$

Para los cuales:

- $\alpha_i$ : Efectos fijos por bloque con un total de 1.257 ficticias. La media fue  $\approx 4,33$ .
- $\delta_t$ : Efectos mensuales, 68 ficticias.
- $\epsilon_{it}$ : Término de error con errores estándar clusterizados por bloque y mes.
- Recordando además que  $S_{it}$  se refiere al área en  $m^2$  de la vivienda, y  $L_{it}$  son los años residuales del contrato.

Por su parte, sustituyendo en la **ecuación (5)** para el modelo en primeras diferencias (FD):

$$\Delta \ln(P_{it}) = 0,97 \Delta \ln(S_{it}) + 0,21 \Delta L_{it} - 0,003 \Delta L_{it}^2 + 1,49 e^{-5} \Delta L_{it}^3 + u_{it} \quad (8)$$

Pasando ahora a los resultados globales junto con estadísticas de interés:

**Tabla 2: Resultados de los modelos econométricos de panel.**

<i>Variable</i>	<i>Pooled OLS</i>	<i>Efectos Fijos (FE)</i>	<i>Efectos Aleatorios (RE)</i>	<i>Primeras diferencias (FD)</i>
$\ln(S_{it})$	0,94*** (0,005)	0,98*** (0,004)	0,96*** (0,005)	0,9716*** (0,006)
<i>lease_years</i>	0,07*** (0,002)	0,18*** (0,007)	0,12*** (0,005)	0,21*** (0,008)
<i>lease_years</i> <sup>2</sup>	-0,001** (0,0003)	-0,003*** (0,0001)	-0,002*** (0,0002)	-0,003*** (0,0002)
<i>lease_years</i> <sup>3</sup>	0,00001 (0,000007)	$1,29 e^{-5}$ *** ( $4,3 e^{-7}$ )	$0,85 e^{-5}$ * (0,000006)	$1,49 e^{-5}$ *** ( $5,1 e^{-7}$ )
<i>R<sup>2</sup> Within</i>	0,482	0,532	0,490	-
<i>R<sup>2</sup></i>	0,482	0,532	0,9959	0,629
<i>Observaciones (n)</i>	71.276	71.276	71.276	60.931

Notas: Significancia  $p < 0,01$ \*\*\*,  $p < 0,05$ \*\*,  $p < 0,1$ .\*

Observamos al  $\ln(S_{it})$  como el determinante hedónico más “relevante” en todas las especificaciones estimadas. Bajo FE, el coeficiente estimado es 0,98 (s.d. 0,004), por lo que ceteris paribus, un aumento del 1% del área se asocia en promedio a un incremento del 0,98% en el precio de reventa, marcando una elasticidad precio de la superficie estadísticamente significativo al 1%.

Los resultados del Pooled OLS se marcan para exponer las consecuencias de estimar un panel sin tener en cuenta las entidades de individuo y tiempo.

Tal y como se anticipó, dichos resultados coinciden con hallazgos en mercados residenciales de alta densidad y regulación, indicando cómo el mercado valora casi proporcionalmente cada m<sup>2</sup> adicional (en promedio), reflejando la escasez de espacio señalada en el marco teórico del mercado singapurense (Malpezzi, 2002; Phang y Wong, 2016).

En cuanto a  $L_{it}$  (los años residuales del contrato), el comportamiento temporal de los precios conforme quedan más o menos años residuales del contrato muestra un patrón de depreciación de la vivienda cúbico:

$$\frac{\partial P_i}{\partial L_i} = \beta_1 + 2\beta_2 L_i + 3\beta_3 L_i^2 \quad (9)$$

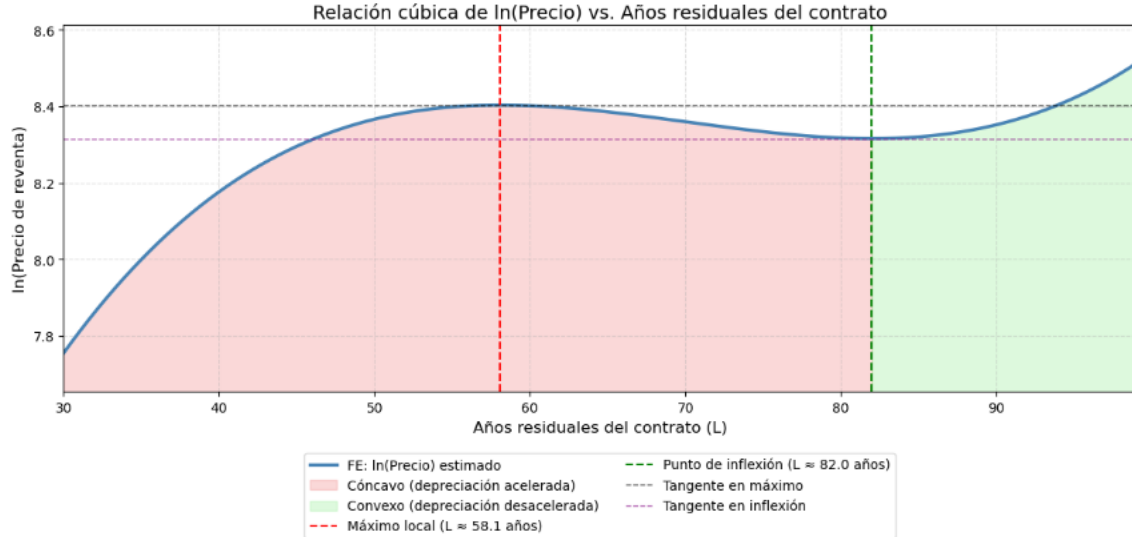
Li, Gao and Tan, 2023, coinciden con esta tendencia polinómica en los años residuales, exponiendo significancia estadística al 1% de  $L_{it}$ , reconociendo en su trabajo que esta variable es poco probable que siga una función lineal para los precios de reventa del HDB, coincidiendo con esta estimación en que es necesaria para comprender las dinámicas del mercado.

Profundizando en esta propuesta de Li, Gao and Tan, 2023, para los coeficientes ( $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ ) se aplica el marco teórico previo para derivar la siguiente ecuación de depreciación del precio respecto a los años residuales. Sustituyendo los coeficientes obtenidos por FE en la ecuación (8) se obtiene:

$$\frac{\partial \ln(P_{it})}{\partial L_{it}} = 0,1836 - 0,0054L_{it} + 0,00003855L_{it}^2 \quad (10)$$

Comportándose gráficamente de la siguiente manera:

**Figura 6: Evolución de los coeficientes modelo FE de los años residuales del contrato vs el precio de reventa de las viviendas HDB.**



Fuente: Elaboración propia con datos del Singapore statistics.

En esta gráfica se han marcado los puntos de inflexión encontrados partiendo desde la ecuación (9), marcando un máximo local de color rojo en  $\approx 58,1$  años residuales del contrato, así como un mínimo local (cambio de concavidad) de color verde sobre  $\approx 82$  años residuales.



Para Li, Gao and Tan el máximo local con sus resultados se marca sobre los 60 años, muy cercano al máximo local encontrado en este TFG; queda disponible en el anexo el cálculo de los puntos críticos anteriormente mencionados, que nos dejan con las siguientes ideas:

- **Tramo creciente ( $L < 58,1$  años):** En este intervalo, la derivada es positiva y se considera un máximo local, por lo que, cada año adicional de contrato en este tramo aumenta el precio marginalmente. Por ejemplo, si  $L = 30$  años, se sustituye en la fórmula:  $\frac{\partial P}{\partial L} = 0.1836 - 0.0054(30) + 0.00003855(30)^2 \approx 0,056$ , por lo que cada año adicional en este corto plazo aumenta el precio en  $\approx 5,6\%$ .
- **Tramo decreciente ( $58,1 \leq L < 99$  años):** Aquí, a partir de  $\approx 40,6$  años, cada año adicional reduce el precio de reventa con una disminución más pronunciada hasta  $L \approx 81,9$  años, siendo a partir de este punto cuando la caída del precio se desacelera.

El área sombreada en color verde claro muestra esa desaceleración del precio, mientras que el área en color rojo claro trata la bajada (o subida) pronunciada de los precios de reventa conforme menos (o más) años residuales de contrato queden hasta  $\approx 40,6$  años.

Cabe destacar que se observa un comportamiento no monótono en el precio estimado donde las unidades con  $\approx 58$  años residuales alcanzan un precio localmente superior al de aquellas con más años restantes, antes de que la curva vuelva a declinar. Este fenómeno encuentra fundamento en varios factores institucionales propios del mercado inmobiliario de Singapur.

En particular, programas como el Selective En Bloc Redevelopment Scheme (SERS) y el Lease Buyback Scheme (LBS) generan expectativas de revalorización o retorno económico favorable en bloques con leases intermedios sobre los 50 años de antigüedad. Muchas viviendas con cerca de 58 años remanentes fueron beneficiarias de reformas integrales bajo los programas Interim Upgrading Programme (IUP) y Main Upgrading Programme (MUP), lo que incrementa su valor relativo respecto a unidades más nuevas pero no renovadas.

Es preciso comentar cómo los resultados próximos a comentar por Eco-RETINA ofrecen una relación más suavizada y aproximadamente monótona o lineal al no imponer forma estructural concreta, sugiriendo cómo este efecto cúbico quizás refleje patrones locales más que relaciones universales.

En cuanto al modelo en FD, esta transformación modifica la interpretación de los coeficientes. Mientras que FE en niveles analiza cómo las características afectan al precio de reventa en términos absolutos, el modelo en diferencias captura cómo las variaciones mensuales en las características contractuales y estructurales afectan al precio en corto plazo.

Esta formulación permite identificar relaciones dinámicas sin necesidad de estimar explícitamente los efectos invariables en el tiempo, preservando la significancia estadística y el sentido económico de los regresores. Los coeficientes mantienen su interpretación en términos de elasticidades y efectos marginales, siendo coherentes con los resultados obtenidos en el modelo en niveles. Sin embargo, implica una pérdida de observaciones correspondientes a bloques con menos de dos transacciones consecutivas, traduciéndose en una reducción de aproximadamente 7.784 casos. En definitiva, FD se presenta como una herramienta complementaria al modelo estático para contrastar la estabilidad de las relaciones observadas en un marco dinámico.

## 7.2. Resultados adicionales: Eco-Retina y comparativa con métodos tradicionales:

Tras discutir los resultados obtenidos por el análisis en panel, pasamos a repetir el proceso con Eco-RETINA para 4 submuestras del 10%, 25%, 50% y 75%, sin dejar de lado los resultados para el 100% de la muestra, cerrando con métricas de precisión del algoritmo para las 4 submuestras realizadas.

Por otro lado, se incluye la variable ‘block’ como variable dicotómica o ficticia en Eco-RETINA, para así controlar la heterogeneidad no observada a nivel de bloque, y obtener resultados similares a unos FE. De esta manera, aprovechamos la capacidad del algoritmo para potenciar o complementar los resultados obtenidos anteriormente.

Partiendo de la **ecuación (6)**, se identifica la siguiente relación econométrica fijada por Eco-RETINA:

$$\ln(P_i) = \beta_0 + \beta_1(\text{OTHER} \cdot L_i) + \beta_2\sqrt{S_i} + \beta_3(\sqrt{S_i} \cdot L_i^2) + \beta_4(\text{Storey\_group:low} \cdot L_i) + \sum_j \beta_j(\text{Block}_j \cdot S_i) + \epsilon_i \quad (11)$$

Donde:

- $\ln(P_i)$ : Logaritmo del precio de reventa.

- $\sqrt{S_i}$ : Raíz cuadrada del área de las viviendas en m<sup>2</sup>.
- $L_i$ : Años residuales de contrato (lease\_years).
- $\text{Block}_j \cdot S_i$ : Interacciones específicas de bloques con el área construida.

Para las submuestras propuestas:

**Tabla 2: Resultados obtenidos por Eco-RETINA.**

Variable	10%	25%	50%	75%	100%
<b>Town: OTHER *</b>	0,0025*	0,0025*	0,0025*	0,0025*	0,0025*
$L_i$	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
$\sqrt{S_i}$	0,1950*	0,1951*	0,1949*	0,1951*	0,1951*
	(0,001)	(0,001)	(0,001)	(0,001)	(0,000)
$\sqrt{S_i} \times L_i^2$	0,0003*	0,0003*	0,0003*	0,0003*	0,0003*
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
<b>Storey_group: low</b>	-0,0015*	-0,0015*	-0,0015*	-0,0015*	-0,0015*
<b>* <math>L_i</math></b>	(0,002)	(0,002)	(0,001)	(0,000)	(0,000)
<b>-Block- * <math>S_i</math></b>					
<b>Coeficientes:</b>	-	-	0,0040*	0,0040*	0,0031*
<b>Block:</b>			<b>-6A-</b>	<b>-6A-</b>	<b>-2B-</b>
<b>s.d.</b>			(0,000)	(0,000)	(0,000)
<b>Coeficientes:</b>	0,0037*	0,0036*	0,0051*	0,0047*	0,0046*
<b>Block:</b>	<b>-2C-</b>	<b>-6A-</b>	<b>-138A-</b>	<b>-139B-</b>	<b>-138A-</b>
<b>s.d.</b>	(0,001)	(0,000)	(0,000)	(0,000)	(0,000)
<b>Coeficientes:</b>	0,0019*	0,0038*	0,0028*	0,0025*	0,0022*
<b>Block:</b>	<b>-104A-</b>	<b>-8B-</b>	<b>-275A-</b>	<b>-275A-</b>	<b>-273B-</b>
<b>s.d.</b>	(0,001)	(0,000)	(0,000)	(0,000)	(0,000)

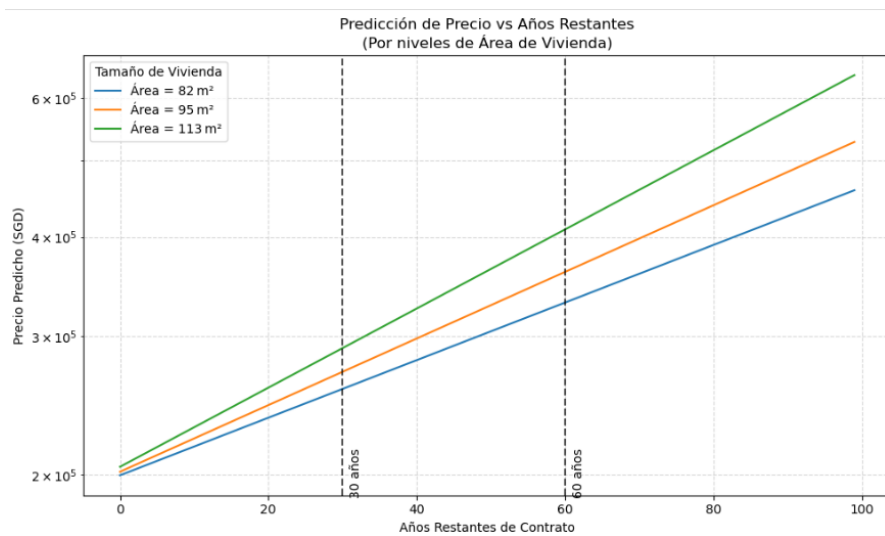
**Notas:** significancias con  $p < 0,01$  se denotan con \*. Para las ficticias de bloque, se indica el coeficiente en primer lugar; luego le sigue el número de bloque indicado entre guiones “-“, y debajo su desviación estándar. Errores: HCl (Corrección de White).

Los resultados denotan que el área en  $m^2$  de las viviendas en su forma  $\sqrt{S_i}$  son un determinante constante para explicar los precios de reventa en todas las submuestras. En concreto, un aumento en una unidad de la raíz cuadrada del área incrementa el precio de reventa en  $\approx 21,54\%$  *ceteris paribus*:

$$e^{0.1951} - 1 \approx 0.2154$$

Siguiendo esta fórmula, la forma en la que Eco-Retina explica la evolución de los años residuales del contrato, ahora teniendo en cuenta tres distintas áreas en  $m^2$  es la siguiente:

**Figura 7: Gráfica predicción Eco-RETINA de precio de reventa para 3 niveles de área por  $m^2$  distintas.**



Fuente: Elaboración propia.

Es decir, cuanto más  $m^2$  disponga una vivienda en Singapur, mayor será su precio de reventa. Las ampliaciones de espacio son altamente valoradas en este mercado en concreto para el periodo de años seleccionados (2017-2022) reflejado por la interacción  $\sqrt{S_i} \times L_i^2$ : el valor de las propiedades grandes crece con los años residuales del contrato, aunque el efecto es más leve  $\beta_3 = 0,0003$ .

Se intuye una prima por altura en el resultado aportado por Storey\_group: low \*  $L_i$ , ya que los pisos de baja altura (low), pierden  $\approx 0,15\%$  de valor por cada año adicional de contrato, intuitivamente se debe a la preferencia por vistas más privilegiadas de la ciudad, el evitar riesgos de inundaciones en climas tropicales, o bien la contaminación acústica en ciudades con alta densidad poblacional. Este es un efecto consistente con trabajos que analizan depreciación por localización vertical (Wong et al., 2020).

Sin embargo, con este modelo desarrollamos más esa teoría del valor de la vivienda al haberse obtenido diversos resultados relevantes y consistentes en todas las submuestras para las ficticias de bloque. La tabla de resultados muestra como bloques concretos presentan coeficientes positivos y significativos en su interacción con el área construida ( $S_i$ ):

- **Bloque 138A:** Cada  $m^2$  adicional en este bloque incrementa el precio en un 0,46% *ceteris paribus*.
- **Bloque 2B:** Contribuye un 0,31% por  $m^2$ .
- **Bloque 273B:** Aporta un 0,22% por  $m^2$ .

Esto implica que un apartamento/vivienda de  $100m^2$  en el bloque 138A tendría  $\approx 46\%$  de prima respecto a uno equivalente en  $m^2$  en un bloque no significativo. Son resultados que encapsulan atributos no observables como la cercanía a estaciones de metro o centros comerciales, viviendas recientes con estándares superiores o externalidades sociales positivas.

La teoría hedónica de precios recoge estos efectos, definiendo el valor de una propiedad como la descomposición en atributos observables y latentes (Rosen, 1974). En cuanto a su rendimiento por precisión en cada una de las submuestras entrenadas, se obtuvieron las siguientes métricas:

**Tabla 3: Métricas por Submuestra**

<i>Submuestra</i>	<i>MAPE (%)</i>	<i>MSE (<math>\times 10^9</math>)</i>	<i>MAE</i>
<b>10%, <math>n = 9,847</math></b>	16,86	19,47	100.153,91
<b>25%, <math>n = 24,619</math></b>	16,93	19,04	99.799,82
<b>50%, <math>n = 49,239</math></b>	16,86	18,42	98.456,38

<b>75%, <math>n = 73,859</math></b>	16,81	18,09	97.902,26
<b>100%, <math>n = 98,479</math></b>	16,78	18,01	97.593,91

Eco-RETINA muestra una estabilidad notable tanto en la precisión como en el error, aun cuando la cantidad de datos varía de 4.923 a 98.479 observaciones, validando su aplicabilidad en entornos de datos parciales o completos, al mantener un rendimiento predictivo parsimonioso a lo largo de distintos volúmenes de datos.

### 7.3. Resultados adicionales: Eco-Retina comparado con Machine Learning models:

Una vez expuestos los resultados de Eco-RETINA, así como de sus métricas de precisión y error, se destaca el hecho de que a la hora de representar los resultados de los modelos de Machine Learning, la condición que sustentan de “cajas negras” no ofrece una comparación equiparable directa.

Modelos como Random Forest o LightGBM están basados en árboles que capturan interacciones complejas mediante importancia porcentual de las variables de tipo adimensional no comparables con los coeficientes obtenidos por ElasticNet y Eco-RETINA. En ese sentido, se ofrece en el anexo una breve discusión de la importancia de estas variables para los modelos de ML, así como los coeficientes de ElasticNet.

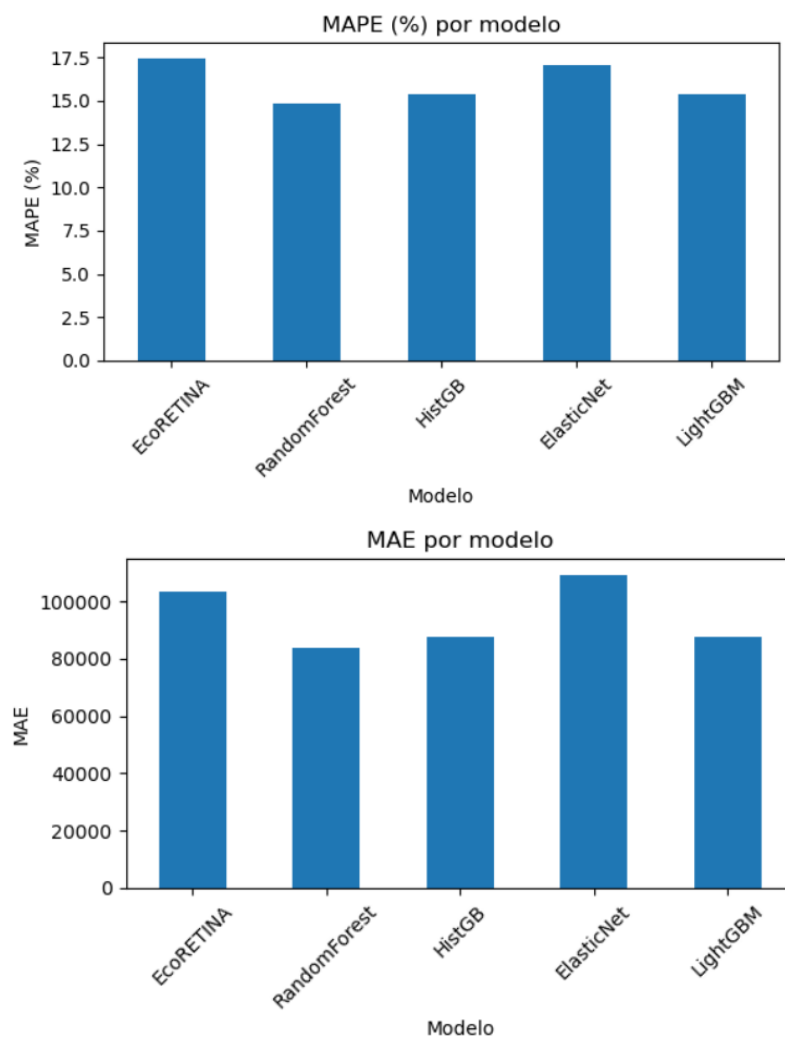
Volviendo sobre las métricas de precisión y error, se exponen los resultados obtenidos mediante la siguiente tabla:

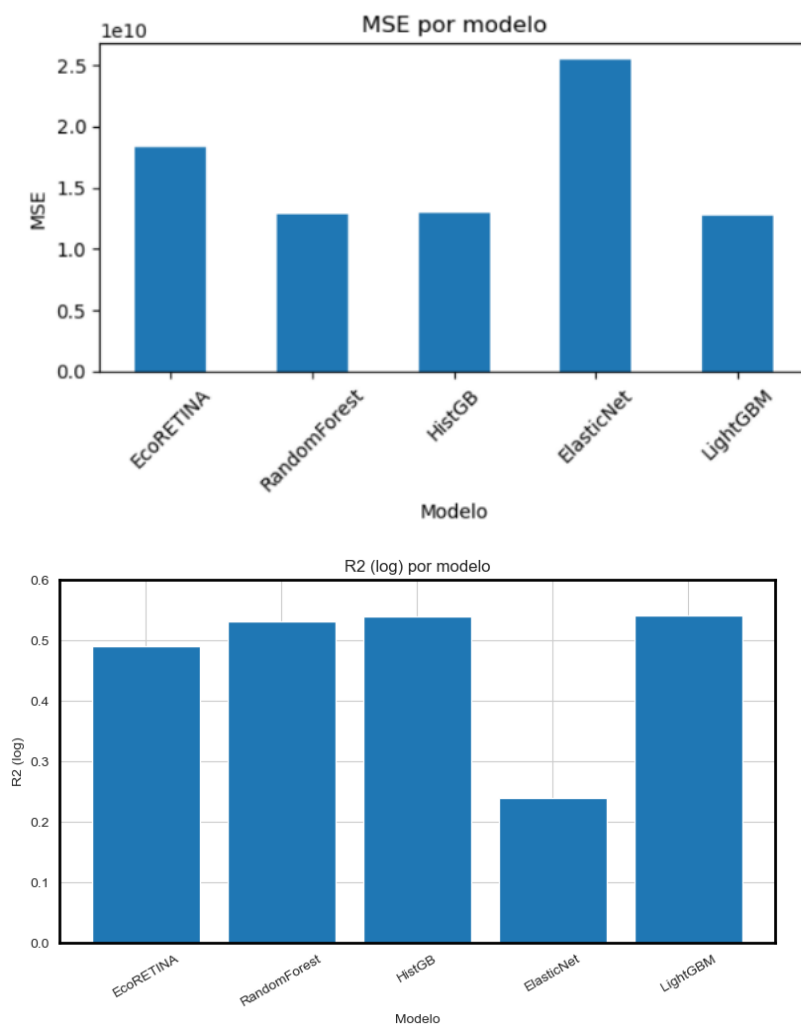
**Tabla 4: Métricas de desempeño de Eco-RETINA.**

<i>Modelo</i>	<i>Train</i>	<i>MSE</i>	<i>MAE</i>	<i>MAPE</i>	<i>Emissiones</i>	
	<i>Time(s)</i>	<i>(SGD<sup>2</sup>)</i>	<i>(SGD)</i>	<i>(%)</i>	<i>R<sup>2</sup></i>	<i>CO<sub>2</sub></i>
<b><i>Eco-RETINA</i></b>	7,27s	$1,8276 \times 10^{10}$	103 243	17,47	0,4913	0,00001
<b><i>RandomForest</i></b>	141,15 s	$1,2874 \times 10^{10}$	83 905	14,82	0,5309	0,0065
<b><i>ElasticNet</i></b>	8,10 s	$2,5489 \times 10^{10}$	109 337	17,04	0,2394	0,00002
<b><i>LightGBM</i></b>	3,15 s	$1,2762 \times 10^{10}$	87 605	15,42	0,5417	0,00075

Donde RandomForest fue el más lento, mientras ElasticNet y LightGBM resultan ser los más rápidos, aunque los modelos con menor error absoluto promedio reducen dicho error en unos  $\sim 20.000$  SGD frente a Eco-RETINA, que ofrece un rendimiento intermedio en comparación de estas métricas, a cambio de una interpretabilidad completa junto a una huella computacional mínima. Gráficamente:

**Figura 8: Comparación de métricas de error y precisión entre modelos de Machine Learning y Eco-RETINA.**



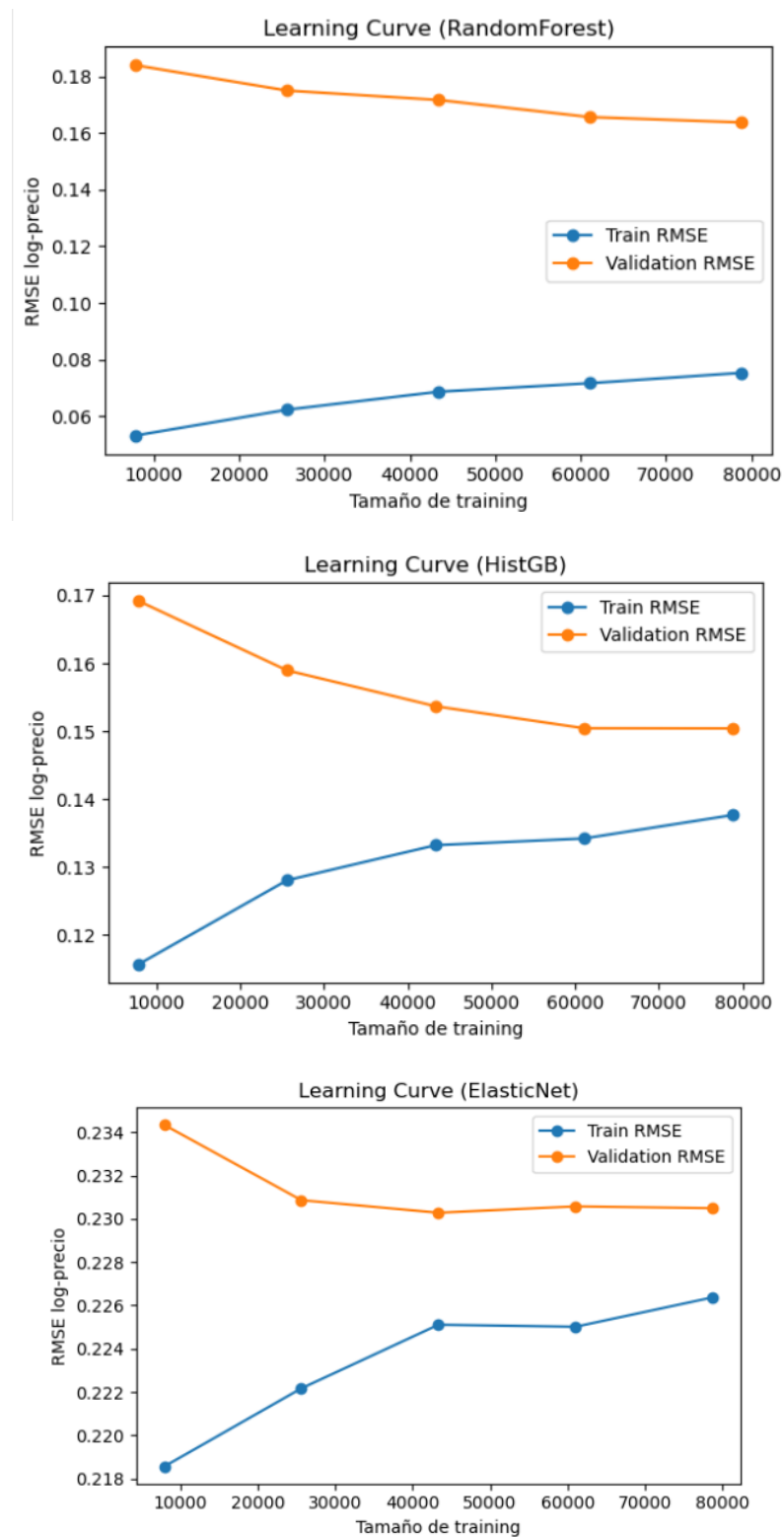


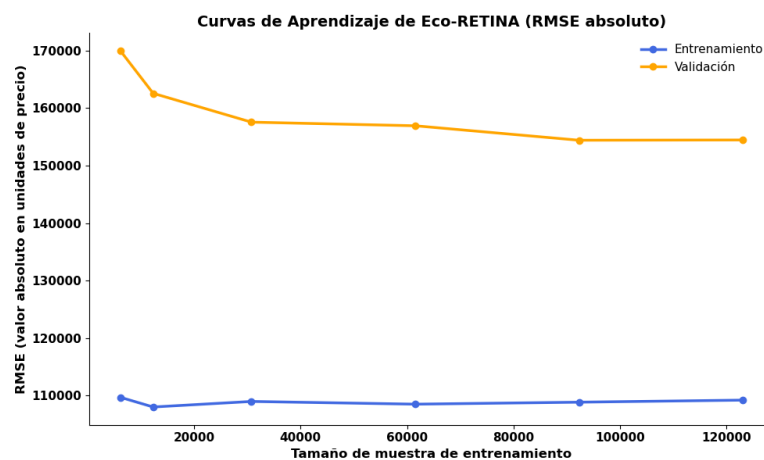
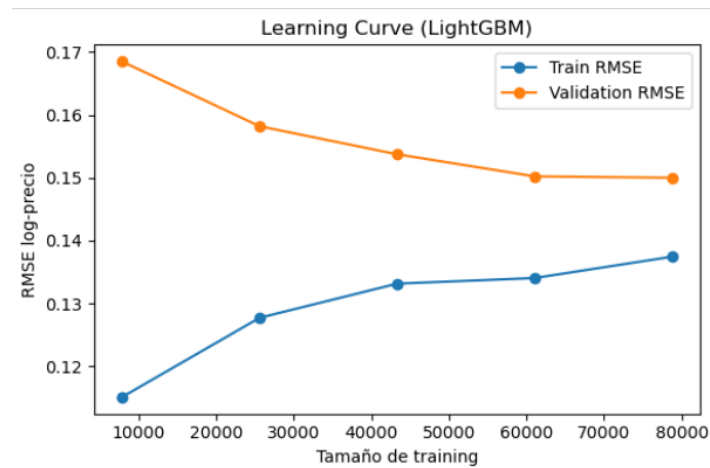
**Fuente: Elaboración propia.**

De cara a la precisión, se crearon curvas de aprendizaje para los modelos de ML con la finalidad de diagnosticar sesgo y varianza sobre fracciones crecientes del training set, del 10 al 100%, recopilándose el RMSE en las dos submuestras:



**Figura 9: Curvas de aprendizaje modelos ML.**





Fuente: Elaboración Propia.

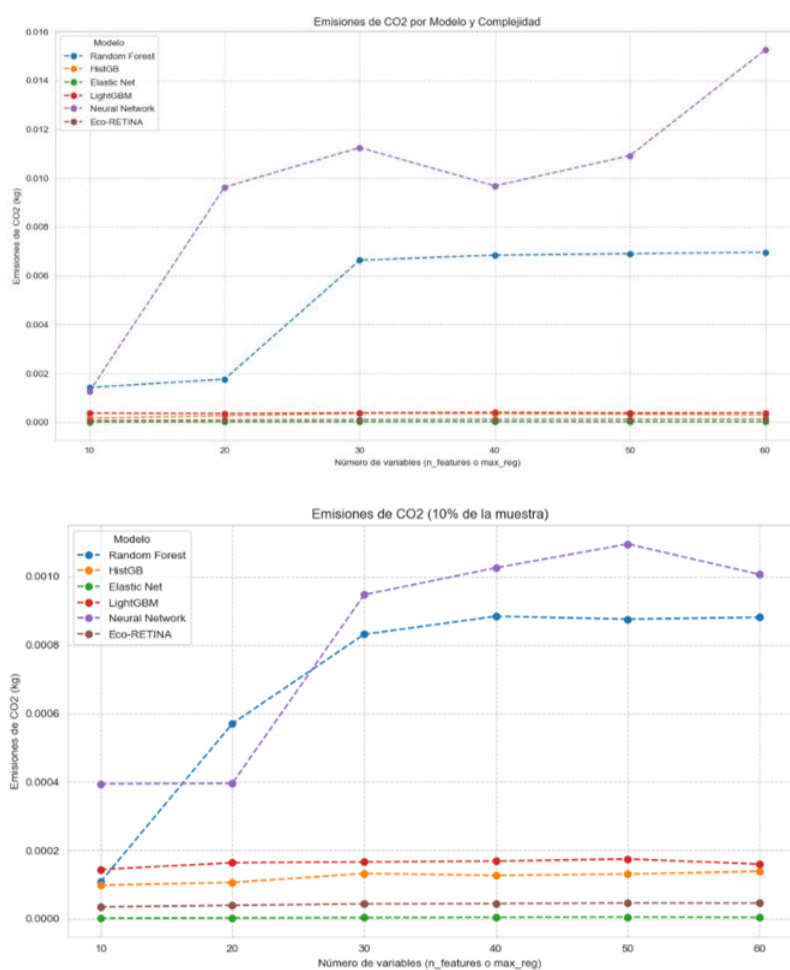
Explicando en parte por qué los boosting tres generalizan mejor y sostienen su ventaja de precisión, marcado por la cercanía entre curvas de entrenamiento y validación en el caso del HistGB y LightGBM, siendo el RandomForest el caso contrario por la distancia entre curvas. Eco-RETINA muestra mejor capacidad para generalizar conforme dispone de más información sin signos evidentes de sobreajuste, y destaca su comportamiento sobre el conjunto de entrenamiento al enseñar cómo limita la complejidad en muestras pequeñas. Sin embargo, la brecha entre los errores de entrenamiento y validación similar a la del RandomForest aunque se reduce parcialmente, indica que ambos modelos aún podrían beneficiarse de conjuntos de datos más amplios.

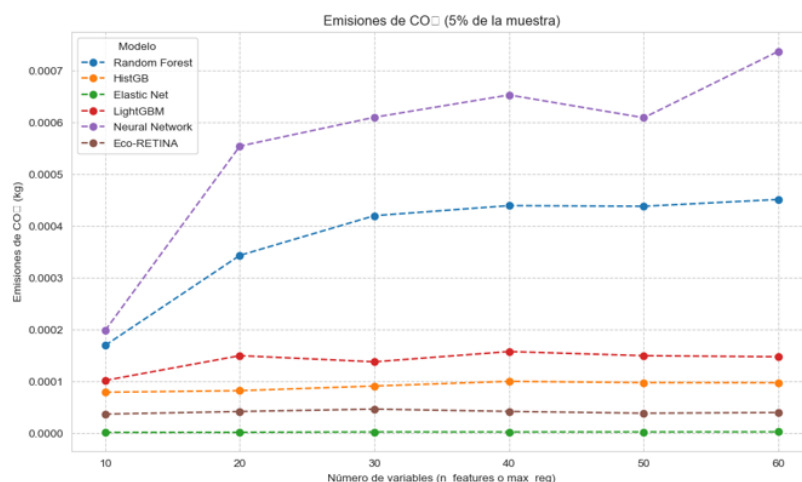
Finalmente, se comparan las emisiones de  $CO_2$  generadas por cada uno de estos algoritmos a medida que aumenta la complejidad del modelo (número de variables), así como el tamaño de la submuestra. El objetivo es no solo ilustrar la calidad

estadística de las predicciones como en los puntos anteriores, sino también el coste ambiental de entrenarlas para ilustrar los principios de la Green AI.

La figura a continuación representa, para un método dado, cómo evolucionan las emisiones de  $CO_2$  (en Kg de  $CO_2$ ) cuando se añaden variables al modelo, una práctica común en el entrenamiento de estos modelos. En concreto:

**Figura 10: Emisiones de  $CO_2$  por modelo y complejidad para los modelos de Machine Learning seleccionados y Eco-RETINA.**





**Fuente: Elaboración propia.**

Las curvas obtenidas permiten identificar la escalabilidad de cada algoritmo, ya que un crecimiento “pronunciado” conlleva un alto coste marginal por variable adicional, mientras que una curva horizontal indica una huella ambiental más ligera y estable. Eco-RETINA demuestra una huella de carbono baja y estable, emitiendo al orden de  $10^{-4}$   $kgCO_2$  (entre 0,00002 y 0,0001  $kg CO_2$ ) incluso al incrementar de 10 a 60 variables.

En comparación con las redes neuronales (neural network), esta alcanza hasta 0,015  $kgCO_2$  en la muestra completa, mientras que Random Forest crece de 0,00018 a 0,0065  $kgCO_2$ , y LightGBM ofreció una baja emisión relativa en este caso, entre unos 0,00008 – 0,0002  $kgCO_2$  según la complejidad, aunque sus modelos son menos interpretables que los de Eco-RETINA

ElasticNet ofreció emisiones similares o inferiores a las de Eco-RETINA por ser un modelo lineal penalizado sin etapas adicionales de validación cruzada o búsqueda de transformaciones. Si bien con esto se busca evidenciar como Eco-RETINA entrega un marco de trabajo mucho más flexible y robusto alineado con los principios de la Green AI de serie, comparable con el de regresiones lineales penalizadas ejecutadas también por defecto sin incluir todas las posibilidades del algoritmo expuesto por Capilla, 2024.

## 8. CONCLUSIONES:

Este trabajo ha explorado la relación entre características estructurales y contractuales de la vivienda pública en Singapur y su precio de reventa al incorporar enfoques econométricos, de aprendizaje automático y de inteligencia artificial verde. El desarrollo e implementación de Eco-RETINA permitió obtener un marco interpretativo alternativo mientras contribuía a una práctica computacional más sostenible.

En primer lugar, se ha demostrado la existencia de una relación polinómica entre los años residuales del contrato HDB y el precio de reventa de las viviendas públicas, revelando distintos tramos de depreciación del valor de reventa conforme el contrato se acerca a su vencimiento. Esta relación, en línea con estudios recientes (Li, Gao y Tan, 2023), permite establecer umbrales útiles para el diseño de políticas de renovación urbana o extensiones de contratos.

Además Eco-RETINA ha permitido observar cómo los pisos bajos pierden valor un 0,15% adicional por año residual de contrato, así como la existencia de externalidades positivas por bloque en el país con primas de hasta 0,46% por m<sup>2</sup> en zonas céntricas. Todo esto con unas métricas de error y de precisión comparables con las de técnicas avanzadas como LightGBM o Random Forest, pero con una huella de carbono significativamente menor, alineando al algoritmo con los objetivos de la Green AI.

A partir de estos resultados se derivan recomendaciones concretas de política pública como el evaluar la posibilidad de reformular o extender los contratos de 99 años en tramos específico de la vida útil del contrato, sobre todo donde se evidencia una pérdida abrupta de valor, así como desarrollar un sistema de monitoreo de valor latente por bloque, incorporando criterios urbanos y sociales que permitan una gestión territorial más equitativa. No obstante, este análisis se ha realizado sobre una base de datos concreta del HDB que no incorpora variables socioeconómicas ni factores externos como la movilidad urbana o calidad de los servicios, limitando su posibilidad de extrapolar directamente los resultados a otros países o contextos.

De cara a futuras líneas de investigación se propone incorporar variables geoespaciales y de entorno para explicar las primas latentes por bloque, valorando la posibilidad de incluir variables socioeconómicas como los niveles de inflación o de

distribución de riqueza. También se plantea aplicar Eco-RETINA a contextos urbanos europeos o latinoamericanos donde la vivienda pública juega un papel relevante y los datos disponibles son limitados.

En síntesis, este TFG profundiza en la comprensión del mercado inmobiliario singapurense reciente, así como establecer un marco alternativo de trabajo para investigaciones económicas que busquen tanto rigor analítico como responsabilidad ambiental, instando a adoptar enfoques de algoritmos como los de Eco-RETINA en el diseño de políticas públicas dentro del marco del análisis aplicado.

## 9. BIBLIOGRAFÍA:

- **Astobiza, A. M., Toboso, M., Aparicio, M., & Lopez, D.** (2021). AI ethics for sustainable development goals. *IEEE Technology and Society Magazine*, 40(1), 66–71. <https://doi.org/10.1109/MTS.2021.3056294>
- **Amaral, P., Gallo, G. and White, H.** (2021). *A flexible tool for model building: the relevant transformation of the inputs network approach*. Docta.ucm.es. <https://docta.ucm.es/entities/publication/fb440ce4-7e30-4436-96d8-fb216f1d9aeb>.
- **Barbierato, E., & Gatti, A.** (2024). Toward green AI: A methodological survey of scientific literature. *IEEE Access*, 12, 23989–24013. <https://doi.org/10.1109/ACCESS.2024.3360705>
- **Bolón-Canedo, V., & Alonso-Betanzos, A.** (2019). Ensembles for feature selection: A review and future trends. *Information Fusion*, 52, 1–12. <https://doi.org/10.1016/j.inffus.2018.11.008>
- **Bolón-Canedo, V., Morán-Fernández, L., Cancela, B., & Alonso-Betanzos, A.** (2024). A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing*, 599, 128096. <https://doi.org/10.1016/j.neucom.2024.128096>
- **Bouza, L., Lannelongue, A., & Dahdaleh, V. P.** (2023). How to estimate carbon footprint when training deep learning models? A guide and review. *arXiv preprint arXiv:2309.12345*.
- **Escarda, M., Eiras-Franco, C., Cancela, B., Guijarro-Berdiñas, B., & Alonso-Betanzos, A.** (2025). Performance and sustainability of BERT derivatives in dyadic data. *Expert Systems with Applications*, 262, 125647. <https://doi.org/10.1016/j.eswa.2024.125647>
- **Geltner, D., Miller, N. G., Clayton, J., & Eichholtz, P.** (2018). *Commercial real estate analysis and investment* (3rd ed.). On the Market.
- **Hausman, J. A.** (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251–1271.
- **ICEX España Exportación e Inversiones.** (2024). *Vivienda pública en Singapur*. <https://www.icex.es>

- **Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., & Creutzig, F.** (2021). Aligning artificial intelligence with climate change mitigation. HAL open science. <https://hal.archives-ouvertes.fr/hal-03368037>
- **Khandel, P., Yates, A., Varbanescu, A.-L., de Rijke, M., & Pimentel, A.** (2024). Distillation vs. sampling for efficient training of learning to rank models. In Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval (pp. 51–60). <https://doi.org/10.1145/3664190.3672527>
- **Li, B., Gao, F., & Tan, S.** (2023). Aging like fine wine: A Singapore public housing story. *International Real Estate Review*, 26(1), 95–126. [https://www.gssinst.org/irer/wp-content/uploads/2023/04/v26-no1-4\\_Singapore-Public-Housing.pdf](https://www.gssinst.org/irer/wp-content/uploads/2023/04/v26-no1-4_Singapore-Public-Housing.pdf)
- **Li, X., Ng, K., & Wong, J.** (2019). Panel data analysis of housing prices in Hong Kong. *International Journal of Housing Markets and Analysis*, 12(3), 456–478.
- **Malpezzi, S.** (2002). Hedonic pricing models: A selective and applied review. *Real Estate Economics*, 30(2), 199–236.
- **Nerlove, M.** (1958). Dynamic econometric models. *Econometrica*, 26(3), 315–335.
- **OECD.** (2022). Measuring the environmental impacts of AI compute and applications: The AI footprint. *Digital Economy Papers* (341).
- **Ooi, J., & Lee, T.** (2021). The impact of leasehold contracts on housing prices in Singapore. *Urban Studies*, 58(5), 987–1003.
- **Phang, S. Y.** (2020). Housing market dynamics under regulatory constraints: Evidence from state interventions. *Journal of Urban Economics*, 110, 1–15.
- **PropertyGuru Singapore.** (2023). HDB SERS (Selective En Bloc Redevelopment Scheme): What is it and does your HDB flat stand a chance? <https://www.propertyguru.com.sg/property-guides/selective-enbloc-redevelopment-scheme-guide-12617>
- **Raman, R., Pattnaik, D., & Lathabai, H. H.** (2024). Green and sustainable AI research: An integrated thematic and topic modeling analysis. *Journal of Big Data*, 11, 55. <https://doi.org/10.1186/s40537-024-00920-x>
- **Rosen, S.** (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.



- **Scells, H., Zhuang, S., & Zuccon, G.** (2022). Reduce, reuse, recycle: Energy-efficient information retrieval. In *\*Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval\**.  
<https://doi.org/10.1145/3477495.3531766>
- **Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O.** (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.
- **Singapore Department of Statistics.** (2024). Population and housing statistics.  
<https://www.singstat.gov.sg>
- **Strubell, E., Ganesh, A., & McCallum, A.** (2019). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics.
- **Verdecchia, R., Sallou, J., & Cruz, L.** (2023). A systematic review of Green AI. *WIREs Data Mining and Knowledge Discovery*, 13(1), e1507.  
<https://doi.org/10.1002/widm.1507>

## 10. ANEXO-EL ESTIMADOR Y LA INTERFAZ DE ECO-RETINA

Acceso al repositorio Github de Eco-RETINA: <https://github.com/jcapilla780/Eco-RETINA>

### 10.1. ¿Cómo fijar el entrenamiento de Eco-RETINA?:

**fit () Parameters:** Esta tabla, disponible en el repositorio Github, explica parámetro a parámetro las entradas que acepta para que el usuario pueda personalizar según sus necesidades las capacidades de Eco-RETINA; en síntesis:

**Tabla 5: Instrucciones para fijar la función de ejecución de Eco-RETINA**

Parameter	Type	Default	Description
<b>y</b>	NDArray	—	Target variable array (1D).
<b>X</b>	NDArray	—	Feature matrix.
<b>con_cols_indices</b>	List[int]	—	Indices of continuous columns in X.
<b>dummy_cols_indices</b>	List[int]	—	Indices of dummy (categorical one-hot) columns in X.
<b>col_names</b>	List[str] or None	—	Names of original input features (used for labeling).
<b>params</b>	List[float]	[-1.0, 0.0, 1.0]	Exponents to apply for generating transformed features.
<b>cross_dummy</b>	bool	False	Allow dummy-dummy feature interactions.
<b>max_R<sup>2</sup></b>	float	0.9	Maximum R <sup>2</sup> allowed for selecting a new feature.
<b>grid</b>	float	0.005	R <sup>2</sup> step used during feature selection.
<b>reg_type</b>	str	'linear'	Type of regression: 'linear', 'logit', or 'probit'.

<b>loss</b>	str	'mse'	Metric to optimize: 'mse', 'mae', 'mape', 'aic', or 'bic'.
<b>max_instances</b>	int	100000	Maximum number of rows to use from X and y.
<b>max_reg</b>	int	100	Maximum number of features allowed in the model.
<b>model_step</b>	int	1	Interval used to evaluate nested models.
<b>chunk_size</b>	int	500	Feature generation batch size (used for memory efficiency).
<b>seed</b>	int	8	Random seed for reproducibility.
<b>cov_type</b>	str	'nonrobust'	Covariance type for statsmodels fit (e.g., 'HC0', 'HC3').

**Fuente: Repositorio de Eco-RETINA en Github. Capilla, 2025.**

Basándonos en esta tabla, pudimos especificar Eco-RETINA para el análisis de los precios de reventa de viviendas públicas en Singapur mediante la siguiente especificación:

**Figura 11: especificación de los parámetros escogidos para entrenar Eco-RETINA.**

```
# =====
# 5. CONFIGURACIÓN DE ECO-RETINA |
# =====
params = {
    'y': y_train.values.astype('float32'),
    'X': X_train.astype('float32').values,
    'con_cols_indices': np.array(con_cols_indices, dtype='int32'),
    'dummy_cols_indices': np.array(dummy_cols_indices, dtype='int32'),
    'col_names': X.columns.tolist(),
    'params': [-1.0, -0.5, 0.0, 0.5, 1.0, 2.0, 3.0],
    'cross_dummy': False,
    'max_r2': 0.9,
    'grid': 0.0025,
    'reg_type': 'linear',
    'loss': 'aic',
    'max_reg': 15,
    'cov_type': 'HC1',
    'chunk_size': 20000, # Mayor chunk_size para mejor manejo de memoria
    'seed': 42
}
```

**Fuente: Elaboración propia.**

Se fijaron como parámetros para estimar el algoritmo las siguientes condiciones para priorizar en parte la eficiencia computacional sin sacrificar la optimalidad partiendo con la exploración de transformaciones funcionales de variables continuas mediante exponentes predefinidos  $\in \{-2, -0.5, -1.0, 0.0, 0.5, 1.0, 2.0\}$  generando relaciones no lineales sin asumir formas paramétricas previas, así como un umbral de  $R^2 \leq 0.9$  para excluir variables altamente correlacionadas y así evitar redundancia o sobreajuste. Luego, se fija el criterio de información de Akaike (AIC) para seleccionar el modelo final (aunque se probaron también los criterios de MCE, MAPE, y el criterio de información Bayesiano BIC), esta vez con el objeto de penalizar la adición de variables marginalmente relevante y alcanzar un punto estable entre bondad de ajuste y complejidad.

Finalmente, se ajustaron los errores estándar mediante la corrección de White para heteroscedasticidad (HC1) dada la varianza no constante presente en el modelo, así como también una elección de máximo 15 hiperparámetros para armar el modelo óptimo.

## 10.2. Especificación del Modelo AR (1):

Cabe destacar que, al introducir una especificación dinámica mediante un término AR (1), el modelo se reformula desde una estimación en niveles a una en primeras diferencias. Esto implica un cambio en la naturaleza del modelo: los coeficientes estimados ya no representan efectos en niveles absolutos, sino variaciones mes a mes. Por tanto, esta especificación no se presenta como una extensión directa del modelo estático, sino como un enfoque complementario que permite capturar la persistencia temporal en los precios de reventa. El término autorregresivo AR (1) se desarrolla de la siguiente manera:  $\Delta \ln(P_{i,t-1}) = \ln(P_{i,t-1}) - \ln(P_{i,t-2})$ . Esto es una estrategia válida ya que cada rezago está correlacionado con  $\Delta \ln(P_{i,t-1})$  por construcción, y bajo el supuesto de no autocorrelación de segundo orden en  $\varepsilon_{it}$ ,  $\ln(P_{i,t-2})$ , no comparte relación con  $\Delta \varepsilon_{it}$ , tal y como lo definen Arellano-Bond, 1991.

Partimos en consecuencia de la ecuación FE en niveles (3) para pasarla a su versión dinámica mediante primeras diferencias para así eliminar los efectos fijos no observados  $\alpha_i$  mediante el siguiente modelo IV-2SLS (Arellano-bond clásico):

$$\Delta \ln P_{it} = \beta_1 \Delta \ln S_{it} + \beta_2 \Delta L_{it} + \beta_3 \Delta L_{it}^2 + \beta_4 \Delta L_{it}^3 + \rho \Delta \ln P_{i,t-1} + \Delta \varepsilon_{it} \quad (4)$$

Al aplicar esos rezagos sobre la “inercia” de los precios mensuales, se construye la matriz de instrumentos  $Z_{it}$  para cada unidad  $i$  y todos los periodos  $t$ :  $Z_{it} = \{\ln P_{i,t-2}, \ln P_{i,t-3}, \dots, \ln P_{i,1}\}$  estimándose los Mínimos Cuadrados de Dos Etapas (2SLS) para obtener los valores ajustados  $\Delta \ln \widehat{P}_{i,t-1}$  (procedimiento consultable en 10.3.), y así sustituirlo en la segunda etapa:

$$\Delta \ln P_{it} = \beta_1 \Delta \ln S_{it} + \beta_2 \Delta L_{it} + \beta_3 \Delta L_{it}^2 + \beta_4 \Delta L_{it}^3 + \rho \Delta \ln \widehat{P}_{i,t-1} + u_{it} \quad (5)$$

Cerrando el modelo IV-2SLS donde IV responde a las siglas en inglés de Variables Instrumentales, ofreciendo una versión clásica de Arellano-Bond. La validez de los rezagos se confirmó mediante dos pruebas estándar: la prueba AR (2) en los residuos de la ecuación en diferencias, obteniéndose un p-valor de 0,77, ayudándonos a concluir que  $\ln P_{i,t-2}$  es exógeno respecto a  $\Delta \varepsilon_{it}$ , y la prueba de Sargan/Hansen de sobre identificación indica que no se rechaza la validez conjunta de las variables instrumentales con un p-valor de 0,12.

### 10.3. Construcción de la Primera Etapa IV-2SLS:

En la primera etapa estimamos:

$$\Delta \ln P_{i,t-1} = Z_{it} \pi + \Delta X_{it} \phi + v_{it}$$

Por mínimos cuadrados, donde  $\Delta X_{it} = (\Delta \ln S_{it}, \Delta L_{it}, \dots)$  para así obtener los valores ajustados del precio y continuar con la segunda etapa especificada en el TFG.

### 10.4. Caracterización de los modelos de Machine Learning:

#### LightGBM:

Sea  $(x_i, y_i)_{i=1}^n$  tu conjunto de entrenamiento, con  $x_i \in R^p$ . En la iteración  $t$ , LightGBM ajusta un árbol  $f_t \in \mathcal{F}$  resolviendo:

$$f_t = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n [g_i(t) - f(x_i)]^2 + \Omega(f)$$

En la cual,  $\mathbf{g}_i(\mathbf{t}) = -\frac{\partial l(y_i, \hat{y})}{\partial \hat{y}} \Big|_{\hat{y}=\hat{y}_i(t-1)}$  es el gradiente negativo de la pérdida  $l$  (MSE en nuestro caso).  $\Omega(f)$  penaliza la complejidad del árbol, típicamente en función del número de hojas y de la magnitud de los pesos.

La predicción se actualiza aditivamente:

$$\hat{y}_i(\mathbf{t}) = \hat{y}_i(\mathbf{t} - 1) + \eta f_t(\mathbf{x}_i),$$

donde  $\eta$  es la tasa de aprendizaje, fijada en este TFG en un 0,5.

#### Random Forest:

La predicción de un solo árbol se denota como  $T_b(\mathbf{x}_i)$ , siendo la predicción agregada  $\hat{y}_i = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}_i)$  y la importancia de la variable se mide de la siguiente forma (Mean Decrease in Impurity):

$$\text{Imp}(X_j) = \frac{1}{B} \sum_{b=1}^B \sum_{s \in S_{b,j}} \Delta \text{MSE}_s,$$

Donde  $S_{b,j}$  son los nodos en el árbol  $b$  que utilizan la variable  $X_j$  y  $\Delta \text{MSE}_s$  es la reducción de MSE lograda en ese nodo.

#### ElasticNet:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\min} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \left( \alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right) \right\}$$

En la que a su vez  $\|\beta\|_1 = \sum_j |\beta_j|$ ,  $\|\beta\|_2^2 = \sum_j \beta_j^2$  y un  $\alpha \in [0,1]$  ajusta el balance Lasso vs Ridge: si  $\alpha = 1$ , regresa al Lasso puro, y si  $\alpha = 0$ , regresa al Ridge puro.

**Tabla 6: Top 10 variables más significativas para Eco-RETINA:**

<i>Variable</i>	<i>Coefficiente</i>	<i>Impacto (%)</i>
<i>block:1B * floor_area_sqm</i>	0.007660*** (0,0000)	0.768936
<i>block:1A * floor_area_sqm</i>	0.007221*** (9,08 e <sup>-5</sup> )	0.724685

<i>block:88 * lease_years</i>	<b>0.005982***</b> (2,69 e <sup>-6</sup> )	<b>0.600028</b>
<i>block:139B* floor_area_sqm</i>	<b>0.005075***</b> (0,006)	<b>0.508785</b>
<i>block:139A *floor_area_sqm</i>	<b>0.005049***</b> (0,000)	<b>0.506198</b>

*Nota: Impacto (%) se refiere a la variación porcentual estimada en el precio de reventa ante una unidad adicional del predictor. Se expresa en forma de elasticidad aproximada del bloque sobre el precio.*

**Tabla 7: Tabla de importancias para Random Forest y LightGBM:**

<i>Variable</i>	<i>Importancia Random Forest</i>	<i>Importancia LightGBM</i>
<i>floor_area_sqm</i>	0.548344	337.0
<i>lease_years</i>	0.118223	440.0
<i>storey_max</i>	0.072861	150.0
<i>town_BEDOK</i>	0.008488	42.0
<i>town_OTHER</i>	0.054490	128.0
<i>town_TAMPINES</i>	0.010932	41.0
<i>town_WOODLANDS</i>	0.009536	73.0
<i>town_YISHUN</i>	NaN	49.0

Destacando a nivel práctico como la flexibilidad de búsqueda automática de variables y sus transformaciones no lineales, así como interacciones, ofrece una perspectiva de análisis única con respecto a estos modelos de ML.

### 10.5. Cálculo del máximo y el mínimo locales de los años residuales del contrato:

Partiendo de la ecuación (10) que define la relación cúbica aproximada por el modelo de efectos fijos:

$$\frac{\partial \ln(P_{it})}{\partial L_{it}} = 0.1836 - 0.0054L_{it} + 0.00003855L_{it}^2.$$

Por un lado, se observa que se cumplen las condiciones necesarias y suficientes para máximo local con la segunda derivada:

$$\frac{\partial^2 \ln(P_{it})}{\partial L_{it}^2} = 2\beta_2 + 6\beta_3L_{it} = -0.0054 + 0.0000771L_{it}$$

Con esto en mano, se procede a la búsqueda de los puntos críticos partiendo de la condición de primer orden igualando la primera derivada de la función a cero:

$$0.1836 - 0.0054L_{it} + 0.00003855L_{it}^2 = 0.$$

Resolviendo mediante fórmula cuadrática obtenemos:

$$L_{it} = \frac{0.0054 \pm 0.000927}{0.0000771} \quad \rightarrow \quad L_1 \approx 58.1, \quad L_2 \approx 81.99$$

Y sustituyendo en la segunda derivada,  $f''(81.99) \approx 0.000921 > 0$  y  $f''(58.09) \approx -0.000921$ , por lo que  $L_1$  cumple condición de máximo local, mientras  $L_2$  de mínimo local por condición de segundo orden.



**Declaración Responsable sobre Autoría y Uso Ético de  
Herramientas de Inteligencia Artificial (IA)**

Yo, Sifontes Valentín,

Sebastián Con

DNI/NIE/PASAPORTE:

72295271S

declaro de manera responsable que el/la presente: **Trabajo de Fin de Grado (TFG) Titulado/a**

**Eco-RETINA: Inteligencia Artificial Verde para la Evaluación Inmobiliaria en Mercados Altamente Regulados. Evidencia desde Singapur.**

Es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a 03 de junio del 2025.

FIRMA

