

*Predicción del Tiempo entre la
Radicación de Licencias y el Inicio
de Obra en Proyectos
Residenciales: Aplicación de Redes
Neuronales, Modelo de Cox y
Random Forest*

Subsecretaría de Planeación y Política

Redy Adolfo Lopez Lopez

Subsecretario

Subdirección de Información Sectorial

Adriana Ivonne Cárdenas

Subdirectora

Equipo técnico - Subdirección de Información Sectorial

Daniela Sedano Saenz

Juan Camilo Peña

Cristian Torres¹

¹ cristian.torres@habitatbogota.gov.co, daniela.sedano@habitatbogota.gov.co y juan.pena@habitatbogota.gov.co

Contenido

Objetivo 4

Introducción..... 5

Revisión de la Literatura..... 6

Datos y variables..... 8

Resultados..... 8

Pronóstico de iniciaciones con base en las unidades licenciadas en 2024..... 15

Conclusiones y recomendaciones..... 17

Bibliografía 18

Objetivo

Este estudio tiene como objetivo fortalecer las capacidades institucionales de monitoreo de la Secretaría Distrital del Hábitat (SDHT) respecto al tiempo transcurrido entre la expedición de la licencia de construcción y el inicio de obra en proyectos inmobiliarios residenciales. Con ello, se busca optimizar la identificación y la priorización de intervenciones sobre proyectos que, por su perfil, podrían requerir acompañamiento técnico o jurídico para viabilizar su ejecución.

La estrategia metodológica se fundamenta en el uso de herramientas como: las redes neuronales artificiales, algoritmos de clasificación tipo Random Forest y el modelo semiparamétrico de riesgos proporcionales de Cox. Estas técnicas permiten estimar el tiempo esperado hasta el inicio de obra a partir de una combinación de características intrínsecas del proyecto (como tipo de promotor, área licenciada o uso del suelo) y determinantes contextuales de orden macroeconómico, tales como la evolución de las tasas de interés o los ciclos del mercado inmobiliario.

El enfoque adoptado permite capturar la complejidad dinámica del proceso edificador en contextos urbanos como el de Bogotá, contribuyendo a una mejor gestión institucional y a una asignación más eficiente de recursos de seguimiento y asistencia técnica.

Introducción

En el contexto urbano de ciudades en crecimiento como Bogotá, el tiempo que transcurre entre la radicación de una licencia de construcción y el inicio efectivo de obra constituye un indicador clave tanto del dinamismo del sector inmobiliario como de la eficiencia institucional en la gestión del desarrollo urbano. Este lapso refleja múltiples dimensiones del proceso edificador: desde la viabilidad financiera del proyecto y las condiciones del entorno macroeconómico, hasta la agilidad administrativa y los factores propios de los promotores inmobiliarios. Como lo advierten estudios como Gyourko et al. (2008) y Quigley & Raphael (2005), los retrasos en la ejecución de proyectos no solo afectan la oferta de vivienda y su asequibilidad, sino que también introducen ineficiencias en el uso del suelo urbano y en la ejecución de políticas públicas sectoriales.

En este estudio se aplica y compara un conjunto de metodologías estadísticas y de aprendizaje automático con el fin de predecir el tiempo transcurrido entre la aprobación de la licencia de construcción y el inicio de obra en proyectos residenciales. Específicamente, se implementan tres enfoques: redes neuronales artificiales (RNA), el modelo de riesgos proporcionales de Cox y el algoritmo de Random Forest. Estas técnicas se entrenan a partir de una base de datos consolidada que integra características técnicas del proyecto (como número de unidades, área del lote, tratamiento urbanístico y tamaño de la empresa constructora) con variables del entorno macroeconómico (como la tasa de interés vigente y la rotación del inventario de vivienda), evaluando su capacidad predictiva y sus aportes diferenciales para el diseño de sistemas de monitoreo y alerta.

El desarrollo de este tipo de herramientas predictivas permite a entidades públicas como la Secretaría Distrital del Hábitat (SDHT) anticipar cuellos de botella, priorizar proyectos con alto riesgo de retraso y diseñar intervenciones focalizadas de acompañamiento técnico y jurídico.

Revisión de la Literatura

Diversos estudios han evidenciado que los proyectos de construcción residencial enfrentan significativas fuentes de incertidumbre incluso después de obtenida la licencia de construcción. Estas se originan principalmente en la volatilidad de las condiciones de mercado, los ajustes en el marco regulatorio y las restricciones en el acceso al financiamiento (Glaeser & Gyourko, 2003). Tales factores pueden traducirse en demoras sustanciales en el inicio efectivo de las obras, afectando tanto la eficiencia del proceso edificador como la capacidad de planeación y seguimiento por parte de las autoridades urbanas.

En este contexto, estimar el tiempo transcurrido entre el licenciamiento y el inicio de obra se vuelve una herramienta clave para la gestión anticipada de riesgos y la planificación territorial. Dicha estimación puede abordarse desde dos enfoques analíticos complementarios: como un problema de análisis de duración o tiempo hasta evento (Cox, 1972), o bien como una tarea de predicción no lineal, altamente sensible a interacciones complejas entre variables observadas (Hastie, Tibshirani & Friedman, 2009). Antes de presentar los resultados, se introduce brevemente las metodologías empleadas:

Redes Neuronales Artificiales (RNA) constituyen una clase de modelos computacionales inspirados en el funcionamiento del sistema nervioso humano, diseñados para aprender patrones a partir de los datos mediante procesos iterativos. Los teoremas de aproximación universal (Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989) demuestran su capacidad para aproximar cualquier función continua con un nivel arbitrario de precisión, lo que elimina la necesidad de asumir ex ante una forma funcional específica. Esta flexibilidad, sin embargo, exige datos de entrenamiento representativos y de alta calidad para garantizar un buen desempeño predictivo.

Estas redes modelan el tiempo esperado hasta el inicio de obra como una función no paramétrica de múltiples variables, incluyendo tanto características del proyecto (como área licenciada, tipo de promotor o localización) como determinantes macroeconómicos (como tasas de interés o condiciones del mercado).

$$y_i = \alpha_0 + \sum_{j=1}^h \alpha_j G \left[\beta_{0j} + \sum_{i=1}^p \beta_{ij} x_i \right] + \epsilon_i$$

Donde:

- y_i valor observado de la variable *tiempo transcurrido* entre la licencia y el inicio de obra.
- x_i son los valores de las variables propias del proyecto y algunas variables macroeconómicas empleadas.
- α_j Representan los pesos que conectan la capa oculta con la capa de salida, incluyendo su respectivo término de sesgo definido como α_0 .
- β_{ij} son los coeficientes que representan los pesos de la capa de entrada a la capa oculta. Siendo el respectivo sesgo.

- G Corresponde a la función de activación o transferencia aplicada a las salidas de la capa de entrada, la cual determina la activación de las neuronas en la capa oculta. Comúnmente, se utiliza la función logística (sigmoide) como activación en la capa oculta, mientras que en la capa de salida se emplea la función identidad, especialmente en modelos con variable dependiente continua.
- h número de neuronas en la capa oculta.
- p número de neuronas en la capa de entrada. Su valor determina el número de rezagos con que se analizará la serie de tiempo.
- ϵ_i representa los errores aleatorios de modelo, los cuales se asumen que son independientes e idénticamente distribuidos con media cero y variancia constante.

De manera complementaria a los modelos de redes neuronales, se implementó un enfoque de aprendizaje automático basado en **Random Forest** para la predicción del tiempo transcurrido entre la radicación de licencias y el inicio de obra en proyectos inmobiliarios residenciales. Esta técnica, reconocida por su robustez y flexibilidad, construye múltiples árboles de decisión sobre subconjuntos aleatorios del conjunto de datos, combinando sus predicciones individuales para generar una estimación agregada más estable y precisa. Su estructura de ensamble le permite capturar relaciones no lineales y complejas entre las variables explicativas, sin requerir supuestos paramétricos estrictos ni condiciones de normalidad en los datos.

La variable dependiente del modelo corresponde al número de días entre la fecha de radicación de la licencia y el inicio efectivo de obra. Como predictores, se incorporaron tanto atributos técnicos del proyecto —como área aprobada, número de unidades, área del lote, tipo de tratamiento urbanístico y clasificación de la empresa constructora— como variables de entorno macroeconómico, incluyendo tasas de interés y rotación del inventario residencial.

Adicionalmente, y con el propósito de incorporar uno de los enfoques más reconocidos para el análisis del tiempo hasta la ocurrencia de un evento y para incluir una herramienta interpretativa, se estimó un **modelo de riesgos proporcionales de Cox**. Este enfoque, ampliamente utilizado en análisis de supervivencia, resulta particularmente adecuado en contextos donde se presentan observaciones censuradas, como es el caso de proyectos que, al momento del corte, no han iniciado obra pese a contar con licencia aprobada.

El modelo de Cox permite estimar la razón de riesgo de inicio de obra para cada proyecto, condicionada a no haber ocurrido previamente, en función de un conjunto de covariables. Las variables incluidas fueron las mismas empleadas en los modelos anteriores, lo cual permite una comparación coherente entre metodologías. Una de las principales ventajas de este enfoque es su carácter semi-paramétrico, que facilita la interpretación directa del efecto de cada variable sobre el tiempo de espera, manteniendo al mismo tiempo una flexibilidad considerable en la forma funcional del modelo.

Datos y variables

El análisis se realiza con una tabla de 392 proyectos inmobiliarios residenciales radicados en Bogotá entre 2018 y 2024, provenientes del registro de licencias de la Secretaría Distrital del Hábitat (SDHT). Para cada proyecto, se logró identificar y empatar la dirección reportada con la información contenida en el Censo Inmobiliario de la Galería Inmobiliaria (GI), lo que permite acceder a variables relacionadas con las características físicas del proyecto, su tipología constructiva y su evolución comercial en el mercado formal.

La variable dependiente del modelo es el número de días transcurridos entre la fecha de radicación de la solicitud de licencia y la fecha efectiva de inicio de obra. Esta variable presenta una distribución altamente asimétrica, con presencia de valores extremos (outliers) y observaciones censuradas, es decir, proyectos que al momento del corte no habían iniciado obra.

Las variables explicativas utilizadas fueron:

- *Tratamiento urbanístico*: categórica, indica las condiciones de edificabilidad permitidas.
- *Tasa de interés*: variable continua mensual al momento de la radicación.
- *Rotación del inventario*: ventas mensuales / oferta de unidades en el mes y zona.
- *Tamaño del proyecto*: número total de unidades previstas, valor reportado en la primera visita censal.
- *Tamaño de la empresa constructora*: clasificación en micro, pequeña, mediana y gran empresa.
- *Área del lote*: extensión en metros cuadrados del terreno.

Resultados

Antes de presentar los resultados, es fundamental precisar el proceso de selección y evaluación implementado. Este consistió en analizar el desempeño de diversas especificaciones bajo cada una de las metodologías propuestas, con el fin de identificar aquellas configuraciones que ofrecieran una mejor capacidad predictiva. A partir de este proceso, se definió un subconjunto de modelos por metodología, permitiendo obtener múltiples estimaciones del tiempo esperado de inicio de obra por proyecto.

Este conjunto de predicciones posibilita la construcción de límites de tolerancia para la generación de alertas tempranas dentro del sistema de monitoreo. En términos prácticos, el funcionamiento de la herramienta puede ilustrarse con el siguiente ejemplo: supóngase que ingresa un nuevo caso, denominado “Proyecto 1”, al sistema. La herramienta generará un conjunto de n estimaciones del tiempo esperado hasta el inicio de obra, cada una proveniente de una de las especificaciones previamente seleccionadas, de la siguiente forma:

$$\hat{t}_1, \hat{t}_2, \hat{t}_3, \hat{t}_4 \dots \dots \hat{t}_n$$

Estas estimaciones permiten construir un intervalo de predicción, a partir del cual es posible establecer umbrales de alerta que. Como propuesta metodológica, se plantea la siguiente regla de decisión: Un proyecto será clasificado como candidato a asesoría prioritaria si el tiempo transcurrido entre la aprobación de la licencia y el inicio efectivo de obra supera el límite de tolerancia, definido como la media de las estimaciones generadas por las especificaciones más una desviación estándar.

A modo ilustrativo, en el ejemplo presentado en la **Tabla 1**, el umbral se estima en 148 días, calculado como la suma de la media de los tiempos estimados y su desviación estándar. Esto implica que, si un proyecto alcanza los 148 días sin haber iniciado obra tras la aprobación de la licencia, será considerado como señal de alerta. Es importante resaltar que esto solo un sistema de alerta que permite identificar proyectos, lo cual no implica ningún tipo de acompañamiento a este proceso de construcción.

Tabla 1 Ejemplo propuesta de funcionamiento para un proyecto dadas las estimaciones de las especificaciones mejor ranqueadas.

Especificación	Tiempo estimado
Modelo 1	120
Modelo 2	140
Modelo 3	150
Modelo n	130
Promedio (μ)	135
Desviación estándar (π)	13
Rango tolerancia ($\mu+\pi$)	148

Fuente: SDHT-SIS.

Redes Neuronales: Desde el enfoque de redes neuronales, se implementó un proceso de validación cruzada para evaluar el desempeño de múltiples arquitecturas. La selección del modelo óptimo se basó en el Error Cuadrático Medio (MSE), por su capacidad para capturar con sensibilidad la magnitud de los errores y penalizar desviaciones significativas en las predicciones.

Los resultados obtenidos se presentan a continuación, destacando las configuraciones con mejor desempeño en función de la base de datos disponible. La arquitectura con el menor MSE correspondió a una red neuronal con dos capas ocultas y 11 nodos, lo que sugiere una adecuada capacidad de modelación no lineal. No obstante, dada su solidez predictiva, también se incluyen dentro del conjunto de especificaciones recomendadas las configuraciones (10,1), (6,5) y (8,1), que mostraron resultados competitivos.

Gráfica 1 RMS por arquitectura

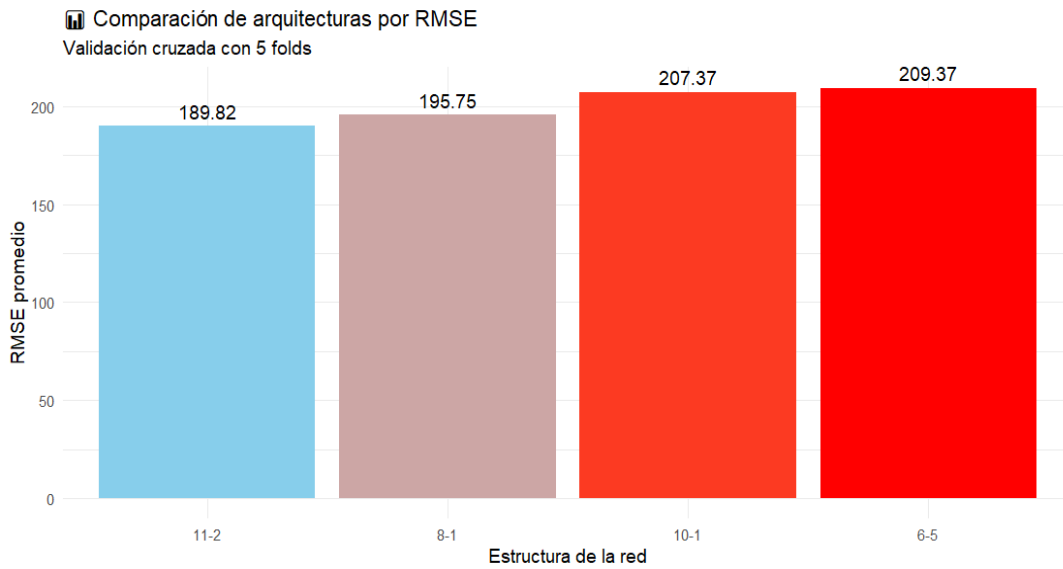
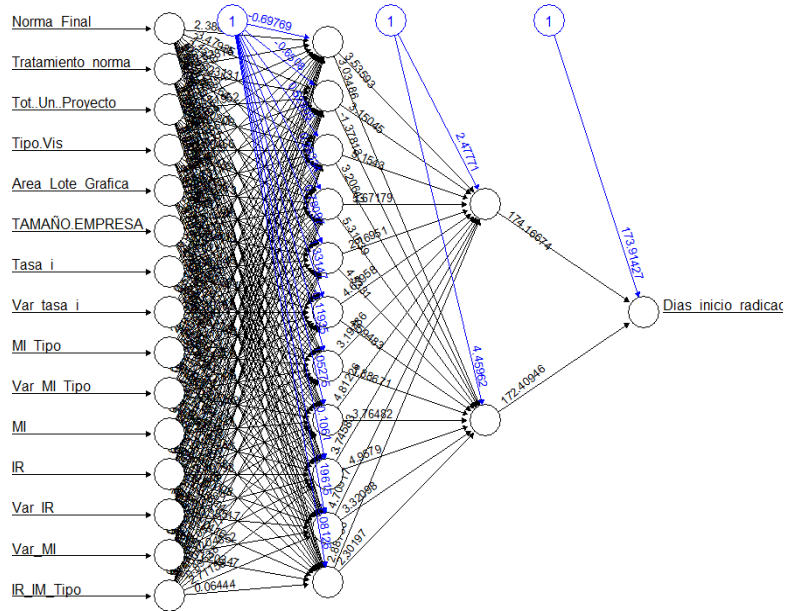


Tabla 2 Resultados mejores especificaciones.

MSE por cada arquitectura

Arquitectura (nodos, capas ocultas)	MSE (días)
(10, 1)	207
(6, 5)	209
(11, 2)	189
(8, 1)	195

Resultados mejor arquitectura (11,2)

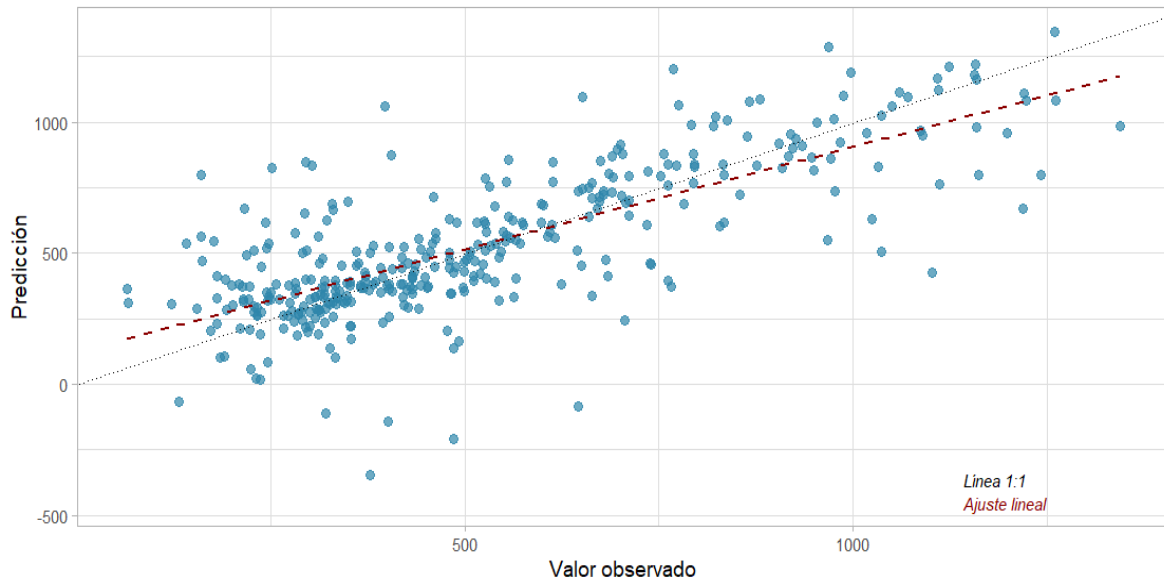


Fuente: SDHT-SIS.

Gráfica 2 Resultados mejor especificación encontrada para redes neuronales (11,2).

🔗 Observado vs. Predicho (mejor red: 11-2)

RMSE: 189.82



Fuente: SDHT-SIS

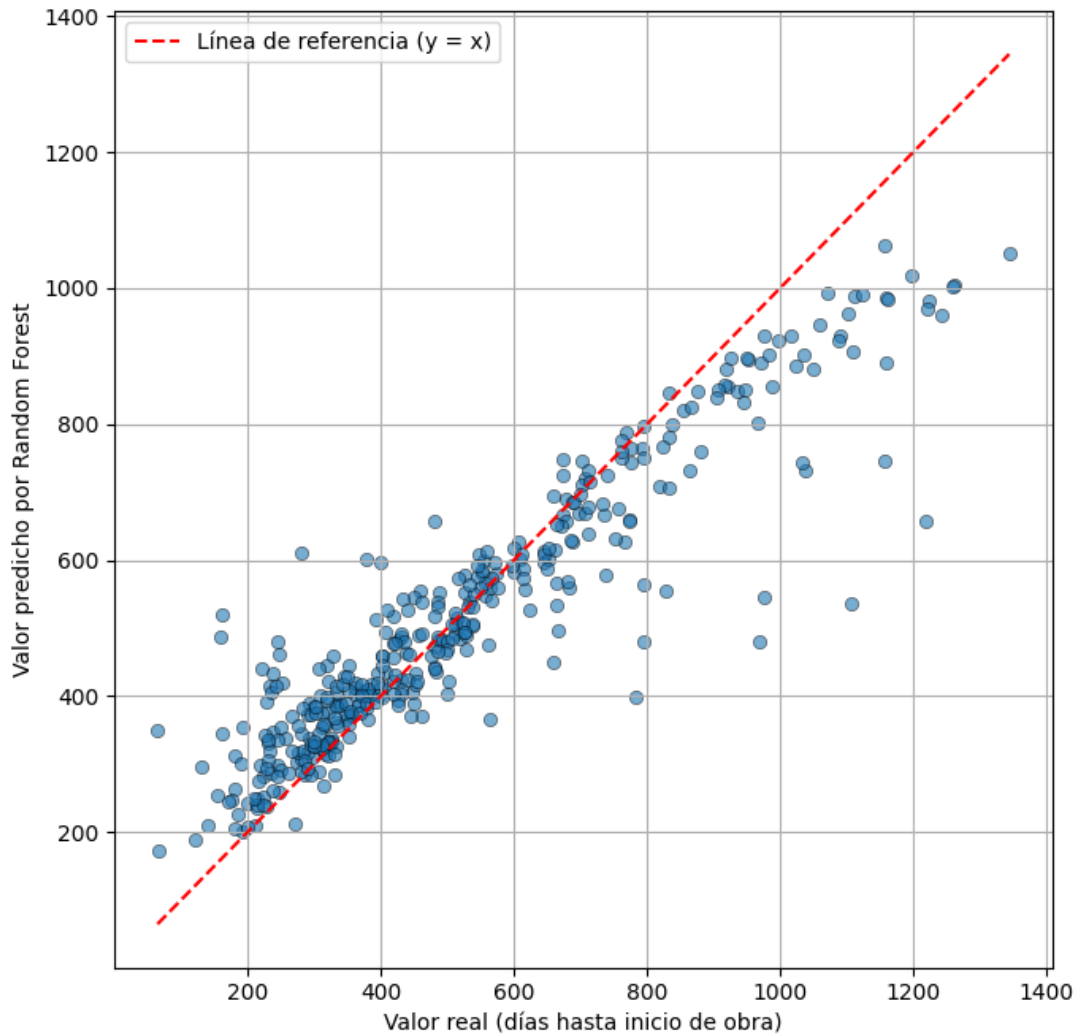
Random Forest: Para la metodología de Random Forest, se implementó un proceso de optimización de hiperparámetros mediante búsqueda aleatoria (*RandomizedSearchCV*), combinada con validación cruzada. Esta estrategia permitió explorar 60 combinaciones distintas de parámetros, extraídas de distribuciones definidas para cada hiperparámetro relevante (como el número de árboles, la profundidad máxima, el número mínimo de muestras por nodo, entre otros). La validación se llevó a cabo mediante un esquema K-Fold con 5 particiones, lo que aseguró una evaluación robusta del desempeño del modelo y mitigó el riesgo de sobreajuste

Los hiperparámetros considerados en la búsqueda incluyeron:

- Número de árboles: [200, 300, 400, 500, 600]
- Profundidad máxima de los árboles: [6, 8, 10, 12, 15, 20, None]
- Mínimo número de muestras para dividir un nodo: [2, 4, 6, 8, 10]
- Mínimo número de muestras por hoja: [1, 2, 4, 6, 8, 10]
- Proporción de características consideradas por árbol: ["sqrt", "log2", 0.5]
- Valor de complejidad para la poda (*ccp_alpha*): valores lineales entre 0 y 0.02 (cinco puntos)

El criterio de selección fue el Error Absoluto Medio (MAE) negativo, que permite identificar el modelo con menor error absoluto promedio en las predicciones. El proceso fue paralelizado utilizando todos los núcleos disponibles del sistema, optimizando así el tiempo de cómputo. A continuación, se presentan los resultados del Random Forest con mejor desempeño, según la base de datos disponible.

Gráfica 3 Resultados mejor especificación encontrada para Random Forest



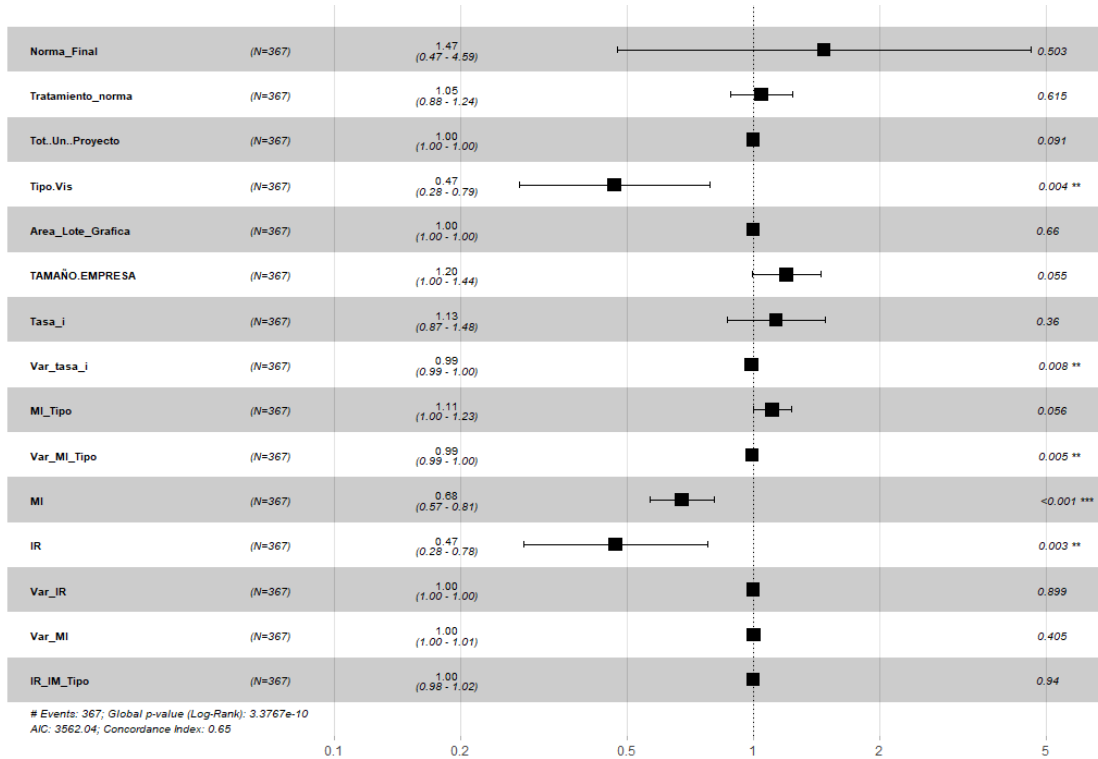
Fuente: SDHT-SIS

Modelo de riesgos proporcionales de Cox: Las estimaciones obtenidas mediante el modelo de riesgos proporcionales de Cox indican que las variables con mayor incidencia sobre el tiempo transcurrido entre el licenciamiento y el inicio de obra son las siguientes: la tipología del proyecto (en particular, los proyectos de tipo VIS), la variación de la tasa de interés entre el momento de licenciamiento y el inicio de obra (*Var_tasa_i*), el índice de rotación de inventario (*MI*), y la tasa de reemplazo (relación entre iniciaciones y lanzamientos).

En primer lugar, una mayor variación positiva en la tasa de interés se asocia con un incremento en el tiempo de inicio de obra, lo que sugiere una mayor cautela por parte de los promotores frente a entornos financieros más volátiles. En segundo lugar, un mayor valor del índice de rotación de inventario —lo que indica una menor velocidad de comercialización de las unidades— también se relaciona con demoras en el inicio de obra, reflejando posibles señales de sobreoferta o baja absorción del mercado. Por último, una tasa de reemplazo reducida (es decir, una menor proporción de iniciaciones respecto a los lanzamientos) tiende a aumentar el tiempo de espera, lo que puede

interpretarse como un ajuste por parte de los constructores ante expectativas de mercado menos favorables.

Gráfica 4 Efecto y significancia de las covariables utilizadas

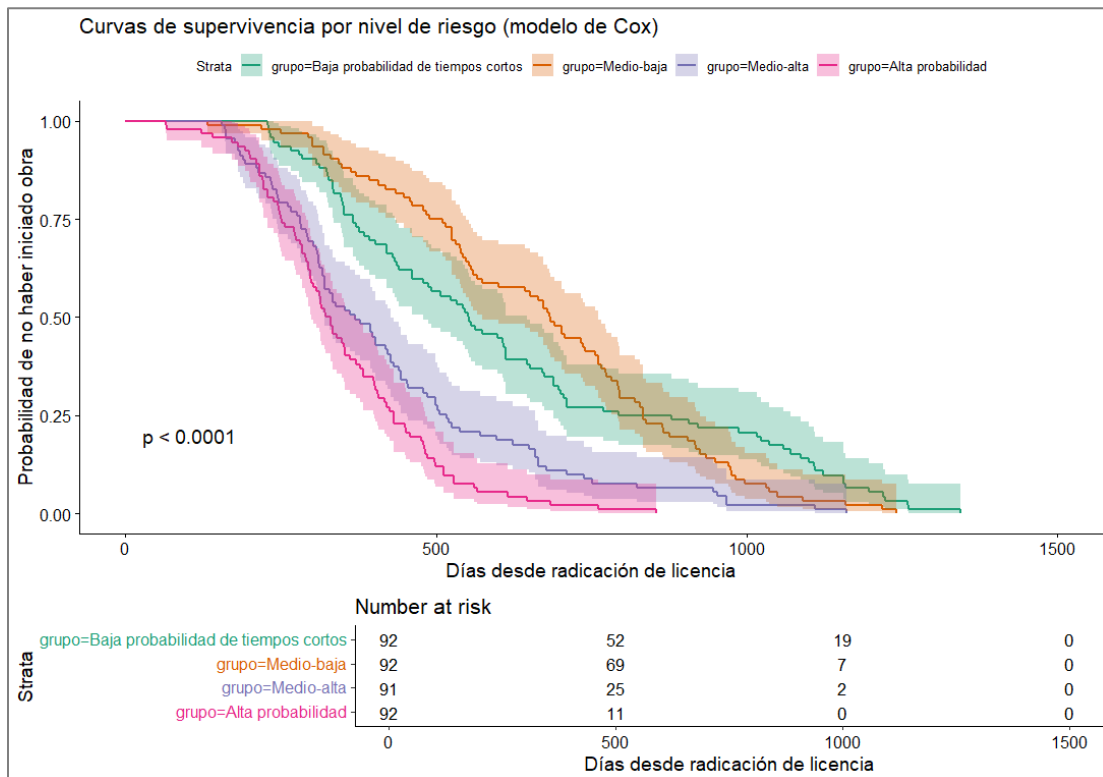
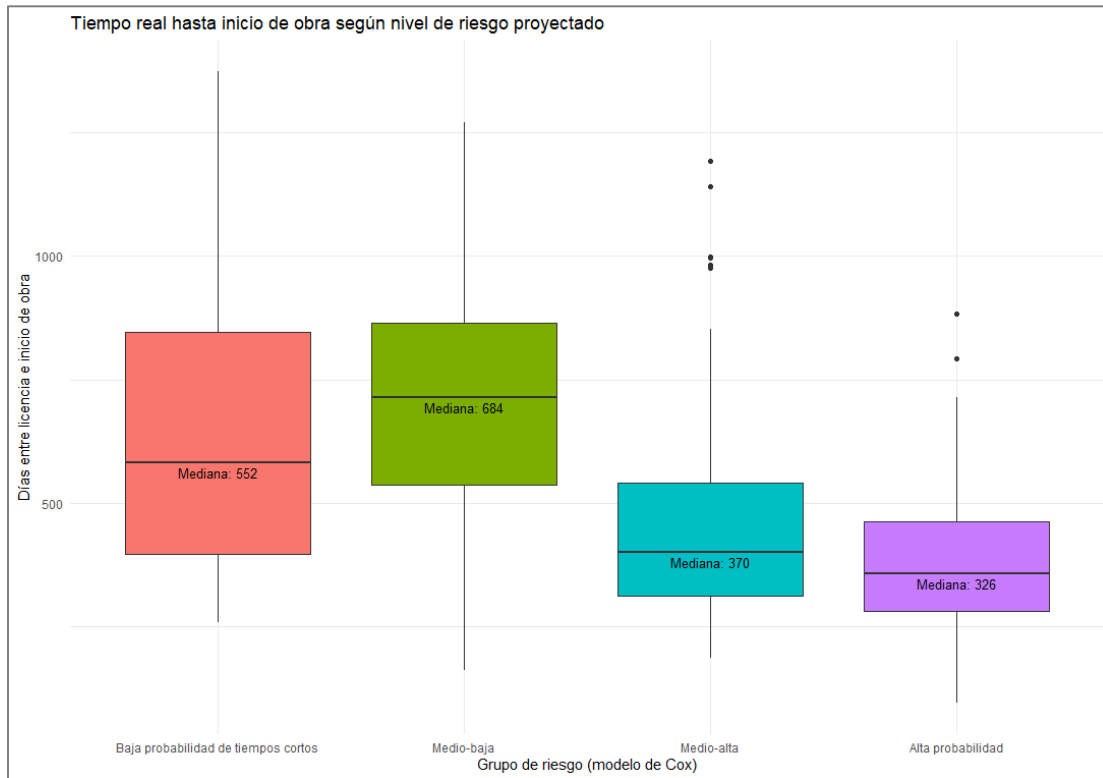


Fuente: SDHT-SIS

El modelo de riesgos proporcionales permite clasificar los proyectos según su perfil de riesgo en cuatro categorías. Entre ellas, destacan dos grupos de especial atención institucional: el primero agrupa proyectos con baja probabilidad de iniciar obra en un tiempo corto, los cuales presentan una media de 552 días desde el licenciamiento; el segundo corresponde a proyectos con probabilidad media-baja, cuya mediana se sitúa en 684 días. Aunque este último grupo muestra una mediana más elevada, el primero presenta una mayor dispersión y presencia de tiempos extremadamente largos, lo que incrementa su nivel de riesgo para efectos de monitoreo y gestión.

En contraste, los proyectos clasificados con probabilidad media-alta y alta de iniciar obra en tiempos cortos presentan medianas de 370 y 326 días, respectivamente, reflejando perfiles operativos más ágiles y predecibles desde la perspectiva institucional.

Gráfica 5 Agrupación de proyectos según el modelo de Cox



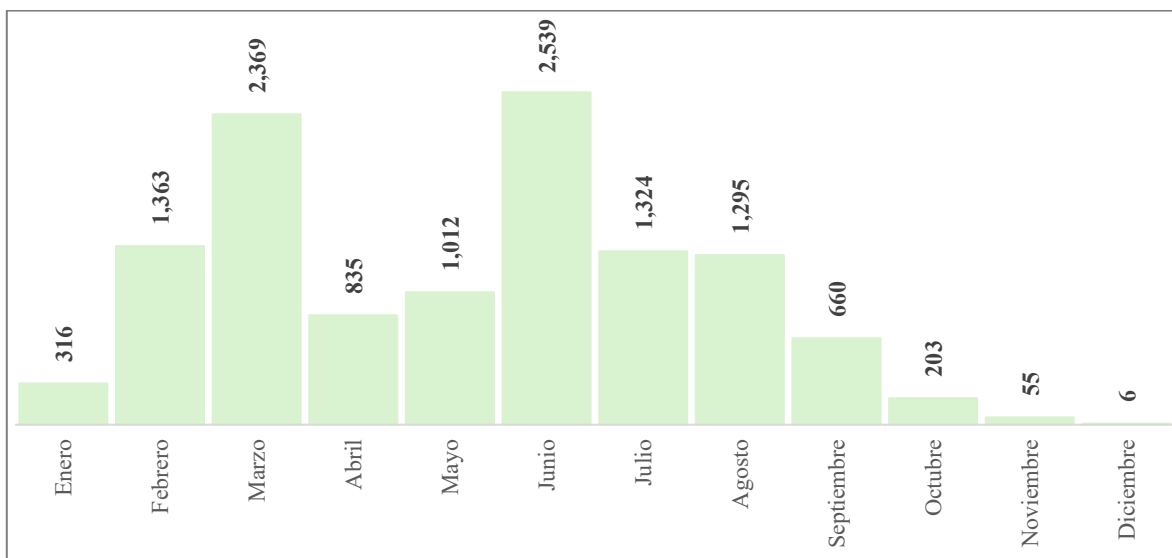
Fuente: SDHT-SIS

Pronóstico de iniciaciones con base en las unidades licenciadas en 2024.

Durante 2024 se radicaron 1.267 licencias en la modalidad de obra nueva, que corresponden a un total de 11.977 unidades residenciales aprobadas en Bogotá. De este volumen, cerca del 70% de las unidades se concentraron en el primer semestre del año, lo que evidencia un fuerte dinamismo en las etapas iniciales del ciclo constructivo.

La distribución mensual muestra picos relevantes en marzo (2.369 unidades) y junio (2.539 unidades), mientras que a partir de septiembre se observa una desaceleración progresiva en el número de unidades licenciadas, cerrando con valores marginales en noviembre (55) y diciembre (6).

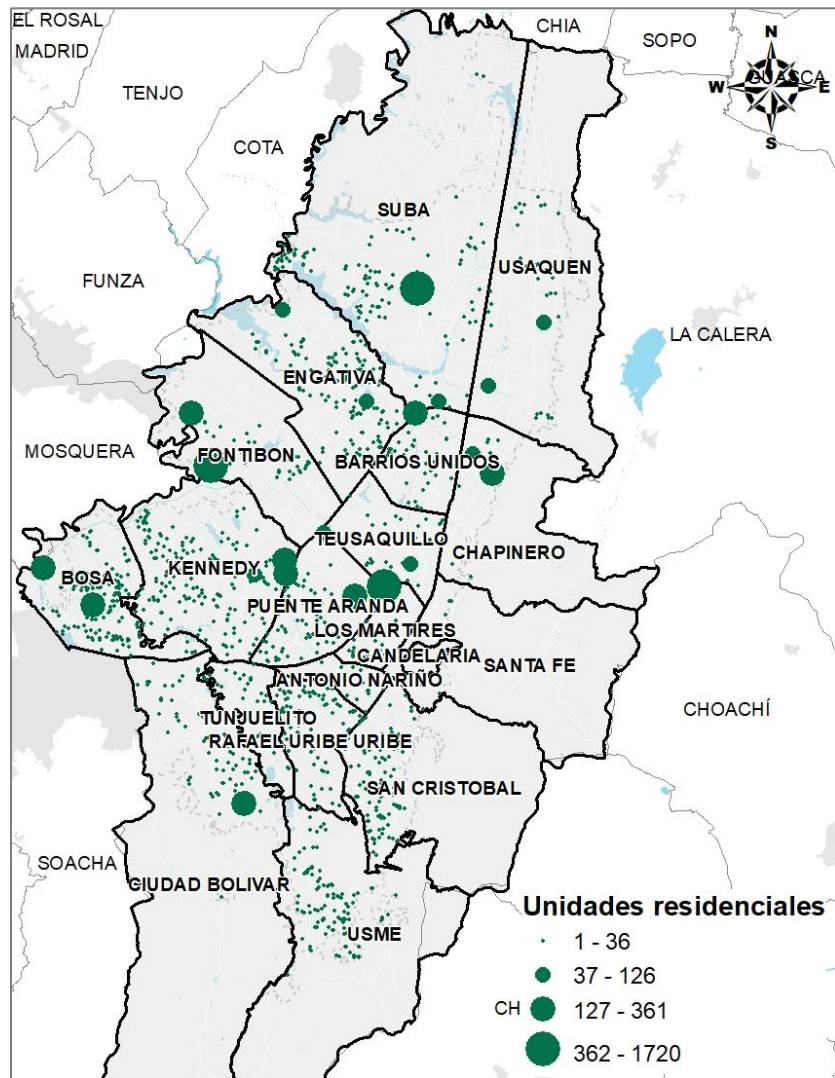
Gráfica 6 Unidades licenciadas en 2024 según mes



Fuente: SDHT-SIS, con datos de licencias (SDP)

Durante 2024, las unidades licenciadas se concentraron principalmente en Puente Aranda con 2.406 unidades (22%), seguida por Suba con 1.984 unidades (18%) y Fontibón con 1.394 unidades (13%). A estas se suman Kennedy con 974 unidades (9%) y Bosa con 844 unidades (8%), de manera que estas cinco localidades reunieron en conjunto más del 70% del total de unidades licenciadas en la ciudad.

Mapa 1. Licencias radicadas en 2024 según número de unidades licenciadas.



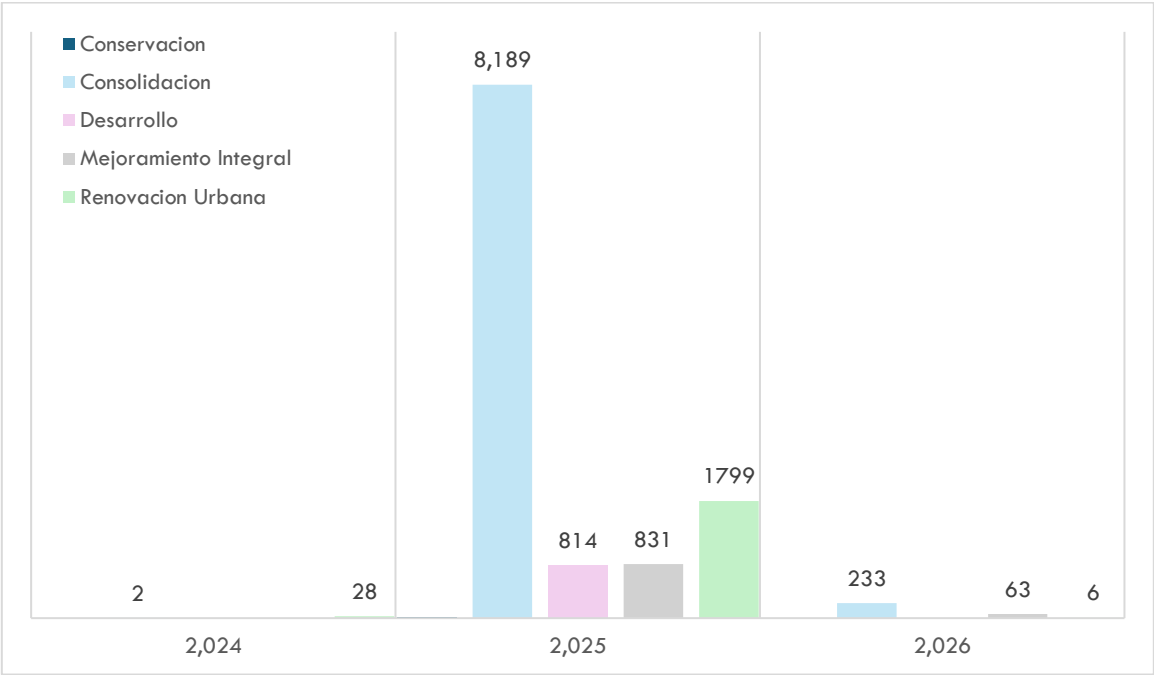
Fuente: SDHT-SIS, con datos de licencias (SDP)

En cuanto a los tratamientos urbanísticos, el tratamiento de consolidación fue predominante, concentrando 70,3% de las unidades (8.425). Le siguieron renovación urbana con 15,3% (1.833 unidades) y, en menor medida, los tratamientos de mejoramiento integral (7,5%) y desarrollo (6,8%). La conservación tuvo una participación marginal, con apenas 0,1% (9 unidades).

Tratamiento	Unidades	Porcentaje
Conservación	9	0.1%
Desarrollo	814	6.8%
Mejoramiento Integral	896	7.5%
Renovación	1.833	15.3%
Consolidación	8.425	70.3%
Total	11.977	100.0%

Este panorama de licencias constituye la base para el pronóstico de iniciaciones de obra en 2025. De acuerdo con las estimaciones del tiempo transcurrido entre la aprobación de la licencia y el inicio de obra, para las 11.947 unidades licenciadas en 2024 se prevé que el 97% (11.642 unidades) inicien construcción en 2025, con un pico de 3.675 unidades en mayo. Al igual que en las radicaciones, se espera que el tratamiento de consolidación concentre cerca del 70% de las iniciaciones (8.189 unidades), confirmando su rol central en la dinámica edificadora de la ciudad.

Gráfica 6 Estimación de las unidades licenciadas mediante el tiempo estimado de inicio de las licencias aprobadas en 2024.



Fuente: SDHT-SIS.

Conclusiones y recomendaciones

Es importante destacar que el valor práctico de esta herramienta depende de su actualización continua. Dado que el entorno del mercado inmobiliario, las condiciones macroeconómicas y las dinámicas regulatorias pueden cambiar con el tiempo, resulta fundamental reentrenar periódicamente los modelos con datos recientes. Esta actualización sistemática no solo mejora la capacidad predictiva de las distintas arquitecturas evaluadas, sino que también garantiza la relevancia y precisión del sistema de alertas para la toma de decisiones por parte de la SDHT.

Bibliografía

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*.

Glaeser, E. L., & Gyourko, J. (2003). The impact of building restrictions on housing affordability.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*.

Gyourko, J., Saiz, A., & Summers, A. A. (2008). *A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index*. Urban Studies.

Quigley, J. M., & Raphael, S. (2005). *Regulation and the high cost of housing in California*. American Economic Review.

Moreno Monroy, A. I., et al. (2020). *Measuring urban expansion and densification: A global and comparative perspective*. World Bank Policy Research.

UN-Habitat. (2022). *The State of the World's Cities 2022: Urbanization and emerging challenges*.