

TALLER 2

Objetivo: evaluar las capacidades del estudiante para aplicar técnicas de machine learning, las cuales permita descubrir insights, sugerir accionables al negocio y calcular el valor ganado.

Contexto del negocio: Apoyo a una empresa de servicios inmobiliarios.

El mercado inmobiliario Colombiano representa más del 20% del PIB del país, siendo parte integral de la economía, proporcionando empleos a través de la construcción y moviendo cantidades de dinero de órdenes significativo en compraventa de propiedades y servicios de alquiler.

HabitAlpes es una startup Colombiana que está armando un portafolio de servicios de consultoría para los propietarios de vivienda y para el gobierno distrital. Entre sus diferentes proyectos se encuentran el soporte para avalúo de los precios de vivienda, la compra, remodelación y reventa de propiedades en la capital del país, consultoría y elaboración de contratos de arrendamiento, correcciones de avalúos catastrales e informes de sugerencia de reclasificación de estratos en barrios de diferentes localidades de la ciudad. Adicionalmente, la empresa busca ser efectiva al momento de ofrecer a sus clientes el portafolios de servicios más conveniente según sus necesidades.

Es por esto por lo que HabitAlpes le ha contratado para que, usando herramientas de Machine Learning e Inteligencia de Negocios, usted participe en la implementación de uno de sus servicios. En particular, HabitAlpes quiere comenzar con la construcción de una herramienta de cálculo de precio de venta para apartamentos de la ciudad, de tal manera que sus clientes puedan ingresar la información de sus respectivos apartamentos y obtener un valor estimado sugerido que esté de acuerdo con las características y ubicación de este, y en línea con los precios de venta de propiedades similares en el sector que le permitan ser competitivos. Para comenzar, la empresa ha recopilado una muestra de la información de apartamentos en el mercado de los últimos dos meses, que servirá como su insumo para la construcción de una prueba de concepto de la herramienta.

Adicionalmente, usted sabe que el proceso de compra requiere la inversión de 6 horas de un perito (salario por hora: \$9.500) que revise los datos y genere un precio. Con un algoritmo de Machine Learning capaz de estimar de forma precisa los montos, se estima que se reducirá el tiempo del experto a 1 hora únicamente. HabitAlpes es capaz de revisar hasta 500

apartamentos al mes. Según datos históricos, se prevé que, por cada estimación por debajo de 20 millones del valor del apartamento, el cliente solicitará un avalúo presencial, mientras que las estimaciones que sobreestiman el valor no son reportadas.

ACTIVIDADES Y ENTREGABLES

1. **[10%] Entendimiento y preparación de los datos:** Genere un breve reporte de entendimiento de datos en donde destaque las dimensiones del dataset, las características de las imágenes e indicadores que considere importante.
2. **[20%] Entrenamiento del modelo de Machine Learning:** Utilizando el conjunto de datos proporcionado, construya un modelo que sea capaz de predecir el precio de una vivienda. Pruebe al menos dos modelos diferentes. Recuerde utilizar de la siguiente manera cada conjunto:
 - a. El conjunto de entrenamiento (*train*) para ajustar los hiperparámetros de los modelos.
 - b. El conjunto de prueba (*test*) para seleccionar el mejor modelo.
 - c. El conjunto de validación (*validation*) para realizar el análisis de resultados del modelo seleccionado.
3. **[20%] Análisis cuantitativo de resultados del modelo:** Calcule las métricas de evaluación para sus modelos, y explique su significado y su valor para la empresa. Justifique la calidad de su modelo, qué elementos impactaron positivamente dichas métricas, y qué oportunidades de mejora encuentra.
4. **[20%] Análisis cualitativo de resultados del modelo:** Utilizando valores SHAPLEY y Lime, realice un análisis cualitativo de sus modelos. Examine tanto el comportamiento global como el comportamiento de una muestra de predicciones individuales. Construya una interpretación de cómo está funcionando su modelo.
5. **[20%] Generación de valor:** Determine la ganancia esperada para HabitAlpes con la implementación del modelo por cada estimación, así como el punto a partir del cual la inversión en el modelo empezará a dejar dividendos (*break-even point*). Realice una estimación paramétrica, haciendo explícitos todos los supuestos usando datos, incluyendo fuentes aplicables y estadísticas recientes. Es posible que necesite clasificar las estimaciones usando una matriz, para estimar los valores asociados a los diferentes tipos de predicciones. Para este punto, se le recomienda considerar:
 - a. Los costos de tiempo asociado a peritos.
 - b. El ahorro de tiempo teórico de su modelo.
 - c. El costo asociado a los errores de su modelo.
 - d. El ahorro real por estimación acertada.

- e. El ROI tomando en cuenta los costos de desarrollo y despliegue de su modelo.
6. **[10%] Insights:** Redacte un informe ejecutivo o una presentación corta en donde muestre los hallazgos o insights más importantes en términos del modelo y los resultados. Ofrezca una recomendación final para el negocio.

FORMATO DE ENTREGA

- El taller debe desarrollarse en parejas.
- Todos los recursos generados deben entregarse mediante un repositorio de GitHub con su estructura apropiadamente documentada mediante el archivo Readme.
- El repositorio de GitHub debe ser público.
- El o los notebooks generados deben poderse ejecutar secuencialmente sin ningún error. En el caso de haber creado varios notebooks, se debe clarificar el orden en que deben ser ejecutados. No olvide también incluir el dataset utilizado.
- Se penalizará a los estudiantes que no cumplan con este formato de entrega