

PRIMERA ENTREGA DE PROYECTO

POR:

Danilo Tovar Arias

Daniel Rosas Mendoza

MATERIA:

Introducción a la inteligencia artificial

PROFESOR:

Raúl Ramos Pollán



UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERIA

MEDELLIN 2023

1. PLANTEAMIENTO DE PROBLEMA:

Un problema común que surge en el ámbito de la compra de alojamientos de una vivienda es la toma de decisiones sobre qué opción es la más adecuada en función de las circunstancias y las necesidades individuales de cada persona, así mismo, el precio del alojamiento en una propiedad puede estar influenciado por una variedad de factores, como la ubicación, el tamaño de la propiedad, el número de habitaciones, el estado de la propiedad, entre otros.

Con el fin de dar una respuesta a este problema, se tiene como objetivo la creación de un modelo de regresión que tenga en cuenta todas las características relevantes de la propiedad, utilizando el dataset de precios de vivienda de París como conjunto de datos de entrenamiento.

2. DATASET

El dataset seleccionado para trabajar en la solución de este problema, es un conjunto de archivos que podemos encontrar en una competencia de la página Kaggle, en su categoría de playground, temporada tres, episodio seis. Este dataset contiene tres archivos.

- **train.csv:** El conjunto de datos de entrenamiento; Obtener 'price' es el objetivo
- **test.csv:** El conjunto de datos de prueba; el objetivo es predecir acordemente 'price' en contraste de estos datos.
- **sample_submission.csv:** muestra de un archivo de envío en el formato correcto

El conjunto de datos para esta competencia (tanto de entrenamiento como de prueba) se generó a partir de un modelo de aprendizaje profundo entrenado en la Predicción del precio de la vivienda de París . Las distribuciones de características son similares, pero no exactamente iguales, a las del original. Los datos contenidos dentro de estos archivos son:

- **squareMeters** – Metros Cuadrados totales

- **numberOfRooms** – Número de habitaciones
- **hasYard** – Tiene patio?
- **hasPool** – Tiene piscina?
- **floors** – Número de pisos
- **cityCode** – Código Zip
- **cityPartRange** – Entre mayor el rango, más exclusive es el vecindario.
- **numPrevOwners** – Número de propietarios anteriores
- **made** – Año de construcción
- **isNewBuilt** – es Nueva?
- **hasStormProtector** – Tiene protección ante tormentas?
- **basement** – Metros cuadrados de sótano
- **attic** - Metros cuadrados del ático
- **garage** – Tamaño del garaje
- **hasStorageRoom** – Tiene almacén?
- **hasGuestRoom** – Tiene habitación de invitados?
- **price** – valor predicho

3. METRICA

La métrica de evaluación principal para el modelo será la raíz del error cuadrático medio (RMSE, por sus siglas en inglés) es una medida comúnmente utilizada en estadística y en el campo del aprendizaje automático para evaluar la precisión de un modelo de predicción.

El RMSE se calcula como la raíz cuadrada de la media de los errores al cuadrado entre los valores reales y los valores predichos. En otras palabras, el RMSE mide la diferencia promedio entre los valores reales y los valores predichos y representa la cantidad de error en las predicciones del modelo.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

RMSE expresa el valor del error, N es la cantidad total de observaciones en el dataset, finalmente, y_i es el precio real de la vivienda, y \hat{y}_i representa el valor predicho, en la instancia i .

El uso de la raíz cuadrada en el RMSE se realiza para que la medida esté en la misma escala que las unidades de la variable de respuesta, lo que facilita la interpretación y comparación de diferentes modelos. En resumen, la raíz del error cuadrático medio sirve para evaluar la precisión de un modelo de predicción.

Por otro lado, para la métrica de negocio, en que las predicciones sean suficientemente confiables para saber la evolución de los precios de acuerdo a las características de las viviendas. Con esta información se pueden realizar análisis financieros para la promoción de alojamientos específicos a las necesidades del cliente.

4. DESEMPEÑO

Lo que se esperaría de este modelo es obtener la predicción del precio de una propiedad en función de sus características y cómo estos factores afectan el precio de la propiedad, con el objetivo de utilizar este conocimiento para promover el mercado de alojamiento y arrendamiento de vivienda, así mismo, puede ser utilizado para ayudar a las personas a tomar decisiones informadas sobre la compra del alojamiento en una propiedad en París, así como para ayudar a los profesionales del mercado inmobiliario.

5. BIBLIOGRAFÍA

- Playground Series - Season 3, Episode 6. Kaggle. (2023).
<https://www.kaggle.com/competitions/playground-series-s3e6/data>
- Root-mean-square deviation. Wikipedia. https://en.wikipedia.org/wiki/Root-mean-square_deviation