

## **PRIMERA ENTREGA DEL PROYECTO**

### **POR:**

Juan David Sandoval Guerrero  
Daniela Tuberquia Villa

### **MATERIA:**

Introducción a la inteligencia artificial.

### **PROFESOR:**

Raul Ramos Pollan

**UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE INGENIERÍA  
2022.**

## 1. Planteamiento del problema.

El modelo que se desarrolla basado en la competencia que tiene como título: “House price”, el cual dependiendo de las características de una vivienda tales como su superficie, localización, entre otros. Se quiere llegar a predecir el valor final de la vivienda en el mercado.

## 2. Dataset.

El dataset a utilizar se puede encontrar en el siguiente link:<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>.

El cual consta de 1459 muestras (casas) y de 79 columnas las cuales describen casi todas las características de las viviendas, algunas de estas características son las siguientes:

| Variable   | Descripción   | Valores que puede tomar la variable.   |
|------------|---|--|
| MSSubClass | Esta variable identifica el tipo de vivienda que se venderá                     | <b>20</b> , vivienda de 1946 y más nuevo todos los estilos.<br><b>30</b> , vivienda 1945 y más antiguo<br><b>40</b> , piso con ático terminado todas las edades-<br><b>45</b> , 1-½ historia- sin terminar todas las edades.<br><b>50</b> , 1-½, historia terminada de todas las edades.<br><b>60</b> , 2 pisos 1946 y más recientemente.<br><b>70</b> , 2 pisos 1945 y más antiguos<br><b>75</b> , 2-½ historia todas las edades.<br>Entre otros valores puede tomar esta variable. |
| MSZoning   | En esta variable describe la clasificación de zonificación general de la venta. | <b>A</b> , agricultura<br><b>C</b> , comercial<br><b>FV</b> , residencial de pueblo flotante<br><b>I</b> , industrial<br><b>RH</b> , residencia de alta densidad.<br><b>RL</b> , residencia de baja densidad<br><b>RP</b> ,parque residencial de baja densidad<br><b>RM</b> , residencial media  |

|              |   |  |
|--------------|---|--|
|              |   | densidad.  |
| Lot Frontage | Pies lineales de calle conectados a la propiedad. |  |
| Lot Area     | Tamaño del lote en pies cuadrados.                |  |
| Street       | Tipo de camino de acceso a la propiedad.          | <b>Grvl</b> ,grava<br><b>Pave</b> ,pavimentado   |
| Alley        | Tipo de callejón de acceso a la propiedad.        | <b>Grvl</b> ,grava<br><b>Pave</b> ,pavimentado<br><b>NA</b> ,Sin acceso a callejones   |
| Lot Shape    | forma general de la propiedad.                    | <b>Reg</b> , regular.<br><b>IR1</b> , Ligeramente irregular.<br><b>IR2</b> , moderadamente irregular.<br><b>IR3</b> , Irregular.   |
| Land Contour | Planitud de la propiedad                          | <b>Lvl</b> , cerca de Plano/Nivel.<br><b>Bnk</b> , aumento rápido y significativo de grado de calle a edificio.<br><b>HLS</b> ,pendiente significativa de lado a lado.<br><b>Low</b> , depresión.                |
| Land Contour | Tipos de servicios disponibles                    | <b>AllPub</b> , Todos los servicios públicos (E, G, W y S)<br><b>NoSewr</b> , Electricidad, Gas y Agua (Fosa Séptica)<br><b>NoSeWa</b> Solo electricidad y gas<br><b>ELO</b> , Electricidad solamente            |
| LoadConfig   | Configuración de lotes                            | <b>Inside</b> ,lote interior<br><b>Corner</b> , lote de esquina<br><b>CulDSac</b> , <b>Cul-de-sac</b> ,<br><b>FR2</b> ,fachada en dos lados de la propiedad<br><b>FR3</b> ,fachada en tres lados de la propiedad |
| Land Slope   | Pendientes de la propiedad                        | <b>Gtl</b> ,pendiente suave<br><b>Mod</b> ,pendiente moderado<br><b>Sev</b> ,pendiente grave.  |
| Neighborhood | Ubicaciones físicas dentro de los límites de      | <b>Blmngtn</b> ,Bloomington Heights  |

|              |                                 |   |
|--------------|---------------------------------|---|
|              | la ciudad de Ames               | <b>Blueste</b> , Bluestem<br><b>BrDale</b> , Briardale<br><b>BrkSide</b> , Brookside<br><b>ClearCr</b> , Clear Creek<br><b>CollgCr</b> , College Creek<br><b>Crawfor</b> , Crawford<br><b>Edwards</b> , Edwards<br><b>Gilbert</b> , Gilbert<br><b>IDOTRR</b> , Iowa DOT and Rail Road   |
| Condition 1: | Proximidad a varias condiciones | <b>Artery</b> , Adyacente a la calle principal<br><b>Feedr</b> , Adyacente a la calle alimentadora<br><b>Norm</b> , Normal<br><b>RRNn</b> , Dentro de 200' del Ferrocarril Norte-Sur<br><b>RRA</b> , <b>adyacente</b> al Ferrocarril Norte-Sur<br><b>PosN</b> , característica externa positiva cercana: parque, área verde, etc. |

tabla 1. Tabla de algunas variables del dataset.

En la tabla 1 se muestran algunas de las características de las viviendas que se tiene en el dataset.

### 3. Métrica.

La métrica de evaluación principal que se utilizara en este proyecto será el error cuadrático medio (RMSE), el cual consiste es la diferencia entre los valores predichos por un modelo y los valores observados.

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

donde N es el número total de datos del dataset,  $y_i$  es el valor real y  $\hat{y}_i$  Es el valor de la predicción.

En la métrica de negocio, se desea que las predicciones sean lo suficientemente confiables, para saber el valor de las viviendas y así hacer un análisis más detallado en este negocio inmobiliario.

### 4. Desempeño.

Lo que se desea de este modelo es predecir los precios de las ventas de las casas, según los requerimientos de cada cliente para así darle un precio estimado de la vivienda en el primer momento en el cual el cliente consulta en la empresa por su vivienda, el modelo no sería fiable si tiene un error mayor al 25% ya que se perdería credibilidad.

## BIBLIOGRAFÍA

- Kaggle, 2 de julio del 2022.  
<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/overview/tutorials>