

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	20 , vivienda de 1946 y más nuevo todos los estilos. 30 , vivienda 1945 y más antiguo 40 , piso con ático terminado todas las edades- 45 , 1-½ historia- sin terminar todas las edades. 50 , 1-½, historia terminada de todas las edades. 60 , 2 pisos 1946 y más recientemente. 70 , 2 pisos 1945 y más antiguos 75 , 2-½ historia todas las edades. Entre otros valores puede tomar esta variable.
MSZoning	En esta variable describe la clasificación de zonificación general de la venta.	A , agricultura C , comercial FV , residencial de pueblo flotante I , industrial RH , residencia de alta densidad. RL , residencia de baja densidad RP , parque residencial de baja densidad RM , 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.	Grvl ,grava Pave ,pavimentado
Alley	Tipo de callejón de acceso a la propiedad.	Grvl ,grava Pave ,pavimentado NA ,Sin acceso a callejones
Lot Shape	forma general de la propiedad.	Reg , regular. IR1 , Ligeramente irregular. IR2 , moderadamente irregular. IR3 , Irregular.
Land Contour	Planitud de la propiedad	Lvl , cerca de Plano/Nivel. Bnk , aumento rápido y significativo de grado de calle a edificio. HLS ,pendiente significativa de lado a lado. Low , depresión.
Land Contour	Tipos de servicios disponibles	AllPub , Todos los servicios públicos (E, G, W y S) NoSewr , Electricidad, Gas y Agua (Fosa Séptica) NoSeWa Solo electricidad y gas ELO , Electricidad solamente
LoadConfig	Configuración de lotes	Inside ,lote interior Corner , lote de esquina CulDSac , Cul-de-sac , FR2 ,fachada en dos lados de la propiedad FR3 ,fachada en tres lados de la propiedad
Land Slope	Pendientes de la propiedad	Gtl ,pendiente suave Mod ,pendiente moderado Sev ,pendiente grave.
Neighborhood	Ubicaciones físicas dentro de los límites de	Blmngtn ,Bloomington Heights

	la ciudad de Ames	Blueste , Bluestem BrDale , Briardale BrkSide , Brookside ClearCr , Clear Creek CollgCr , College Creek Crawfor , Crawford Edwards , Edwards Gilbert , Gilbert IDOTRR , Iowa DOT and Rail Road
Condition 1:	Proximidad a varias condiciones	Artery , Adyacente a la calle principal Feedr , Adyacente a la calle alimentadora Norm , Normal RRNn , Dentro de 200' del Ferrocarril Norte-Sur RRA , adyacente al Ferrocarril Norte-Sur PosN , 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 la empresa, 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>