

Primera entrega de proyecto

POR:

Luisa Maria Castro Cortez

MATERIA:

Introducción a la inteligencia artificial

PROFESOR:

Raul Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA FACULTAD DE INGENIERÍA
MEDELLÍN 2023

1. Planteamiento del problema

Si nos ponemos en la tarea de preguntar a un comprador de vivienda que nos describa la casa de sus sueños probablemente no empezará por la altura del techo o la proximidad a una estación del metro. Pero existen muchas variables que influyen en las negociaciones sobre el precio de venta de una propiedad. Es por esto por lo que se desea desarrollar un modelo que permita estimar el precio de venta de las viviendas, basado en las variables que pueden presentar las viviendas, al tener mejores estimaciones, se podrá generar un mayor interés por parte de los compradores de vivienda al momento de buscar una nueva propiedad.

2. Dataset

El dataset a utilizar proviene de una competencia de kaggle en la cual se proporcionan datos con 79 variables explicativas que describen aspectos de las viviendas residenciales en Ames, Iowa. El dataset este compuesto por un conjunto de archivos .csv y .txt que proporcionan la información requerida.

El archivo que contiene los datos de las viviendas es nombrado data_description y contiene la siguiente información:

- SalePrice - the property's sale price in dollars. This is the target variable that you're trying to predict.
- MSSubClass: The building class
- MSZoning: The general zoning classification
- LotFrontage: Linear feet of street connected to property
- LotArea: Lot size in square feet
- Street: Type of road access
- Alley: Type of alley access
- LotShape: General shape of property
- LandContour: Flatness of the property
- Utilities: Type of utilities available
- LotConfig: Lot configuration
- LandSlope: Slope of property
- Neighborhood: Physical locations within Ames city limits
- Condition1: Proximity to main road or railroad
- Condition2: Proximity to main road or railroad (if a second is present)
- BldgType: Type of dwelling
- HouseStyle: Style of dwelling
- OverallQual: Overall material and finish quality
- OverallCond: Overall condition rating

- YearBuilt: Original construction date
- YearRemodAdd: Remodel date
- RoofStyle: Type of roof
- RoofMatl: Roof material
- Exterior1st: Exterior covering on house
- Exterior2nd: Exterior covering on house (if more than one material)
- MasVnrType: Masonry veneer type
- MasVnrArea: Masonry veneer area in square feet
- ExterQual: Exterior material quality
- ExterCond: Present condition of the material on the exterior
- Foundation: Type of foundation
- BsmtQual: Height of the basement
- BsmtCond: General condition of the basement
- BsmtExposure: Walkout or garden level basement walls
- BsmtFinType1: Quality of basement finished area
- BsmtFinSF1: Type 1 finished square feet
- BsmtFinType2: Quality of second finished area (if present)
- BsmtFinSF2: Type 2 finished square feet
- BsmtUnfSF: Unfinished square feet of basement area
- TotalBsmtSF: Total square feet of basement area
- Heating: Type of heating
- HeatingQC: Heating quality and condition
- CentralAir: Central air conditioning
- Electrical: Electrical system
- 1stFlrSF: First Floor square feet
- 2ndFlrSF: Second floor square feet
- LowQualFinSF: Low quality finished square feet (all floors)
- GrLivArea: Above grade (ground) living area square feet
- BsmtFullBath: Basement full bathrooms
- BsmtHalfBath: Basement half bathrooms
- FullBath: Full bathrooms above grade
- HalfBath: Half baths above grade
- Bedroom: Number of bedrooms above basement level
- Kitchen: Number of kitchens
- KitchenQual: Kitchen quality
- TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)
- Functional: Home functionality rating
- Fireplaces: Number of fireplaces
- FireplaceQu: Fireplace quality
- GarageType: Garage location
- GarageYrBlt: Year garage was built
- GarageFinish: Interior finish of the garage
- GarageCars: Size of garage in car capacity
- GarageArea: Size of garage in square feet
- GarageQual: Garage quality

- GarageCond: Garage condition
- PavedDrive: Paved driveway
- WoodDeckSF: Wood deck area in square feet
- OpenPorchSF: Open porch area in square feet
- EnclosedPorch: Enclosed porch area in square feet
- 3SsnPorch: Three season porch area in square feet
- ScreenPorch: Screen porch area in square feet
- PoolArea: Pool area in square feet
- PoolQC: Pool quality
- Fence: Fence quality
- MiscFeature: Miscellaneous feature not covered in other categories
- MiscVal: \$Value of miscellaneous feature
- MoSold: Month Sold
- YrSold: Year Sold
- SaleType: Type of sale
- SaleCondition: Condition of sale

train.csv - el conjunto de entrenamiento

test.csv - el conjunto de prueba

sample_submission.csv - un envío de referencia de una regresión lineal sobre el año y el mes de la venta, los metros cuadrados del lote y el número de dormitorios.

3. Métricas

La métrica de evaluación principal para el modelo será error cuadrático medio (RMSE) entre el logaritmo del valor predicho y el logaritmo del precio de venta observado. (Tomar los logaritmos significa que los errores en la predicción de casas caras y casas baratas afectarán por igual al resultado). el cual se calcula mediante la siguiente expresión:

$$\epsilon = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}$$

Donde ϵ es el valor del RMSE, n es el número total de observaciones en el dataset, p_i es la predicción de la variable objetivo y O_i es el valor real de la variable objetivo.

Por otra parte, en cuanto a la métrica de negocio, se tiene interés en que las predicciones sean lo suficientemente confiables para saber precio de venta de las viviendas. Con esta información se pueden realizar análisis financieros para determinar qué tan viables pueden ser comprar ciertas viviendas en particular.

4. Desempeño

Lo que se esperaría de un modelo de este tipo es obtener la predicción del precio de venta de cada vivienda. Con esta información se desearía obtener negociaciones más acertadas de acuerdo con los diferentes tipos de vivienda.

5. Bibliografía

- Anna Montoya, DataCanary. (2016). House Prices - Advanced Regression Techniques. Kaggle. <https://kaggle.com/competitions/house-prices-advanced-regression-techniques>