

UNIVERSIDAD DEL VALLE DE GUATEMALA

Minería de datos, sección 30



Informe Proyecto 2

Sebastian Garcia – 22291

Jose Angel Morales – 22689

GUATEMALA, mayo de 2025

Introducción:

El siguiente documento cuenta con la información recopilada durante el análisis a lo largo de varias semanas realizado sobre el conjunto de datos relacionados con el precio de las casas proporcionado.

Durante estas semanas el objetivo principal del análisis fue encontrar un modelo que nos permitiese predecir con precisión cual sería el precio de una casa a partir de las características que conociéramos sobre esta como pueden ser el caso de número de habitaciones, metros cuadrados de vivienda, si esta había sido renovada o no entre otras cosas.

Esta información puede ser de vital importancia para las empresas que se dedican a la construcción de casas de esta forma, dadas las características que se esperan de la casa se puede buscar un cálculo aproximado de cual podría ser el valor de la vivienda, lo cual puede luego compararse con el precio final que coloquen a la casa para evaluar si la casa pudiese estar siendo devaluada en precio o si esta estuviese siendo supra valorada.

Resumen de los resultados:

Inicialmente realizamos un análisis de los datos para reducir las variables a utilizar para predecir el precio de las casas, esto se llevó a cabo por medio de un análisis exploratorio por el cual reducimos el número de variables eliminando aquellas que no aportaban de manera significativa.

A lo largo de las semanas se tuvieron 2 enfoques distintos para intentar predecir el precio de las casas por un lado inicialmente se realizaron modelos de predicción para predecir el valor del precio de la casa, del cual el mejor fue el realizado con Redes Neuronales con una precisión de 0.83.

El problema principal que nos encontramos con estos modelos fue que el RMSE (Raíz del Error Cuadrático Medio) de todos los modelos incluido el de Redes contaban con un valor muy elevado por lo cual no consideramos que estos modelos sean buenos predictores para predecir el precio de las casas.

Por otro lado, la alternativa que usamos para predecir el precio de las casas fue por medio de una variable categórica que separase el precio de las casas en 3 categorías: Caras, Intermedias y Económicas.

Con esto realizamos distintos modelos de clasificación que no solo nos daban mejores resultados en términos de la precisión con la que se predecía la variable creada, sino

que también estos no sufrían el error que tenían los modelos predictivos de tener un indicador para no ser adecuados a la predicción del modelo.

Para clasificar los mejores modelos fueron KNN, SVM y Redes Neuronales con una precisión arriba del 90% para los 3 modelos, a pesar de que los 3 no estuvieron tan lejos en la precisión consideramos que el mejor fue Redes Neuronales.

Esto se debe a que no solo fue el modelo con mejor precisión de los 3 con una precisión del 97% sino porque este fue el que mejor logro hacerlo sin necesidad de tunear el modelo, lo cual si tuvimos que realizar con SVM y KNN para lograr resultados similares.

Otra ventaja del modelo elaborado con Redes Neuronales es que debido a la reducción realizada de variables inicialmente en el análisis exploratorio este modelo no mostro ser extremadamente más complejo que los otros 2, resultando en que este no exigiera más recursos que los otros 2 de manera significativa.

Conclusiones:

- La utilización de una variable categorica derivada de la variable numérica del precio de las casas probo ser una alternativa viable para predecir los precios.
- Las redes neuronales fueron el mejor modelo para predecir los precios de las casas sin necesitar de manipular demasiado el modelo.
- El analisis exploratorio realizado nos permite reducir la complejidad de modelos complejos como lo pueden ser Redes Neuronales.