

Aprendizaje Automático Supervisado

Regresión para predicción del precio de vivienda de alquiler en Barcelona

Reinel Méndez Reina
14 abril 2024

Objetivo General: Analizar y comprender, las variables que afectan el precio de la vivienda de alquiler en Barcelona.

Objetivos específicos:

1. Recomendar barrios según las preferencias de los usuarios.
2. Dar una predicción de la vivienda en Barcelona según las preferencias del usuario.
3. Entender que variables afectan el precio de la vivienda de alquiler
4. Entender como las dinámicas de la ciudad afectan el precio del alquiler de viviendas en Barcelona

Introducción:

El **aprendizaje supervisado** es un tipo de aprendizaje automático (machine learning) en el que se entrena un modelo para entender y aprender de la relación entre un conjunto de características (variables independientes) y una variable objetivo (variable dependiente) a partir de datos reales.

Regresión: En problemas de regresión, la variable objetivo es continua. El objetivo es predecir un valor numérico, como el precio de una casa, en función de ciertas características, como el número de habitaciones, el tamaño del terreno, etc.

Regresión con árboles de Decisión: Utiliza árboles de decisión para modelar la relación entre las variables predictoras y la variable objetivo. Divide iterativamente el espacio de características en regiones más pequeñas y más homogéneas para hacer predicciones.

Regresión Lineal: Es uno de los modelos de regresión más simples y utilizados. Busca ajustar una línea recta a los datos para minimizar la suma de los errores cuadráticos entre las predicciones del modelo y los valores reales. Puede ser un modelo simple o múltiple, dependiendo del número de variables predictoras.

Descripción del proyecto:

El proyecto desde un principio se concibió, como la forma de relacionar variables que, según mi propio criterio, considere eran importantes a la hora de buscar y determinar el barrio y el precio de alquiler, que se debe pagar por vivir en los barrios de Barcelona, según las preferencias del usuario. Y si estas variables eran importantes o no a la hora de buscar un alquiler en la ciudad.

Por esto se realizó un estudio detallado y procesamiento de múltiples conjuntos de datos, que permitieran obtener información valiosa y completa de variables que a criterio propio podría influenciar la variable target 'Precio anual',

para luego obtener el siguiente conjunto de datos:

Título: Barcelona Datos de Alquiler

Fuentes: (a) Origen: Todos los datos fueron extraídos de la página oficial del gobierno de Cataluña "OPEN DATA". Este dataset es una recopilación de diferentes datasets, que fueron procesados y unificados un solo dataset. (b) Creador: Reinel Mendez Reina (c)

Fecha: 18 Marzo, 2024

Past Usage:

Información Relevante:

En que barrio es mejor vivir dependiendo de las preferencias de cada persona.

Numero de registros: 73

Numero de variables: 17 Variables.

Información de Variables:

- Any: 2023
- Codi_Districte: Código que se le asigna desde el ayuntamiento a cada distrito
- Nom_Districte: Nombre del distrito
- Codi_Barri: Código que se le asigna desde el ayuntamiento a cada barrio
- Nom_Barri: nombre del barrio
- Lloguer_mitja: Descripción
- Precio anual: Precio promedio del alquiler por barrios de Barcelona
- Camaras_seguridad: Cámaras de seguridad controladas por la policía local
- Total contratos: Promedio de contratos por barrio en Barcelona
- m2: Promedio de m2 para viviendas de alquiler en Barcelona
- Med_Hab: Promedio de habitaciones por vivienda de alquiler en Barcelona
- Med_baños: Promedio de baños por viviendas de alquiler en Barcelona
- Med_Pers: Promedio de personas que viven en las viviendas de alquiler en Barcelona
- Transporte: Transporte público (estaciones de bus día, nocturno, estaciones de bus)
- Areas verdes: Parques, jardines, zonas de juego e interiores de manzana
- Equip social: zonas wifi, centros cívicos, universidades, instalaciones deportivas, hospitales, CAPS, CUAPS
- Equip ocio: Restaurantes, ocio nocturno, Museos, zoológicos

Valores de variables faltantes: None.

Con este conjunto de datos se realizaron las regresiones y es el sustento de todo el proyecto.

Sustentación:

En primer lugar: Los datos son el recurso mas importante de mi proyecto: por esto fueron datos que se extrajeron de fuentes oficiales del ayuntamiento de Barcelona, portal Open Data.

Se realizó un procesamiento individual para cada conjunto, para obtener las variables más importantes que a mi criterio podría estar relacionadas con el precio de la vivienda de alquiler en Barcelona.

Como resultado de estas limpiezas de datos por variables, se obtuvo el conjunto de datos unificado con unas 17.

Para empezar el procesamiento de los datos unificados, se volvió a hacer limpieza de datos y luego se procede a hacer un análisis exploratorio para tener conocimientos a profundidad y con sentido de los datos, esto permitió conocer la naturaleza de mis datos y entender estadísticamente el comportamiento de los mismos.

En segundo lugar, se procedió con el procesamiento de los datos y creación de modelos, sin embargo, en mi proyecto logro evidenciar que mis modelos iniciales no fueron lo suficientemente capaces de generar predicciones que se ajustaran a los datos reales, se evaluaron diferentes modelos, se trato de mejorar los modelos, se implementaron nuevos modelos, se hizo ingeniería de características, y las métricas no eran las mejores, por tanto se decidió que se escogerían los dos modelos con mejores métricas y se procedería a hacer un stacking.

El **stacking** es una **técnica de ensamblaje de modelos en el aprendizaje automático** que combina múltiples modelos base para mejorar la precisión de las predicciones. En el stacking, en lugar de hacer predicciones directamente con los modelos base, se utiliza un meta-modelo (también llamado "meta-regresor" o "meta-clasificador") para combinar las predicciones de los modelos base y generar una predicción final.

Se logro que el **Meta-modelo** tuviera unos mejores rendimientos en cuanto a métricas, por tanto, se demuestra un mejor ajuste a los datos reales y mas confianza. Que los modelos individuales.

Streamlit: Streamlit es una biblioteca de Python que se utiliza para crear aplicaciones web interactivas de manera rápida y sencilla. Con Streamlit, puedes convertir rápidamente tu código Python en una aplicación web interactiva que se ejecuta en tu navegador web.

Como forma producto final de este proyecto, se implementó un Streamlit, con una interfaz sencilla, el usuario puede escoger las variables de su preferencia para que le sea recomendado el barrio donde debería vivir y además, le dará el precio (predicción) que deberá pagar por vivir en este barrio.

Conclusión: Pretendo que el meta-modelo comprenda las relaciones complejas entre las variables suministradas y logre predecir lo mas exacto posible el precio del alquiler de vivienda en Barcelona y así que la recomendación sea más exacta para los usuarios.