

24 - 26 | Noviembre 2021 | Madrid
XLVI Reunión de Estudios Regionales

International Conference on Regional Science

Ciudades llenas, territorios vacíos

Universidad Autónoma de Madrid



RESUMEN

Título: Predicción del precio de la vivienda mediante algoritmos de aprendizaje automático

Autores y e-mails: Mora-Garcia, Raul-Tomas; Cespedes-Lopez, Maria-Francisca; Perez-Sanchez, Vicente-Raul; Perez-Sanchez, Juan-Carlos.
rtmg@ua.es, paqui.cespedes@ua.es, raul.perez@ua.es, jc.perez@ua.es

Departamento: Edificación y Urbanismo

Universidad: Universidad de Alicante

Área Temática: *Big Data y Algoritmos de Machine Learning en Ciencia Regional*

Resumen: *Los algoritmos de aprendizaje automático (Machine Learning) se están utilizando para múltiples aplicaciones de la vida real, así como en investigación. En la actualidad y como consecuencia de la tecnología digital, existe una mayor disponibilidad de grandes conjuntos de datos estructurados y georreferenciados, facilitando el uso de estos algoritmos que permiten analizar e identificar patrones, así como realizar predicciones que faciliten la toma de decisiones a los usuarios. Un amplio campo de trabajo se focaliza en el estudio del precio de la vivienda y la valoración masiva de éstas. En esta investigación se utilizan varios algoritmos de aprendizaje automático para desarrollar un modelo de predicción del precio de la vivienda en un caso de estudio de la ciudad de Alicante. Se utiliza una base de datos de precios de oferta georreferenciados extraídos de portales inmobiliarios entre el año 2019 y el 2021, formada por más de 38.000 inmuebles. Entre los algoritmos utilizados, destacan los basados en boosting (Light Gradient Boosting Machine, Gradient Boosting Regressor, Extreme Gradient Boosting y AdaBoost Regressor), los basados en bagging (Random Forest Regressor y Extra Trees Regressor), entre otros algoritmos tradicionales (Linear Regression, Ridge Regression, K Neighbors Regressor y Decision Tree Regressor). Se han comparado los rendimientos de 10 algoritmos de aprendizaje automático, identificando aquellos con los que se obtiene un mayor poder explicativo y menor error de estimación. Para ajustar los hiperparámetros de cada algoritmo de aprendizaje se ha utilizado un subconjunto de datos para el entrenamiento, empleando la técnica de validación cruzada (cross validation), y evaluando las medidas de ajuste y error en los subconjuntos de validación y prueba (no usados en el entrenamiento) para evaluar el sobreajuste (overfitting). Se abordan las siguientes fases: Preparación de los datos, Ingeniería de características, Selección del modelo, Optimización de hiperparámetros, y Evaluación del modelo.*

Palabras Clave: *Precio de la vivienda, Valoración masiva, Machine Learning, Hiperparámetros, Evaluación del modelo.*

Clasificación JEL: R31 - Oferta y mercados de la vivienda