

Índice

Introducción

Análisis exploratorio de las tasaciones de vivienda

Modelo Automatizado de Valoración (AVM)

Aplicación empírica

Conclusiones, limitaciones y trabajo futuro

Introducción

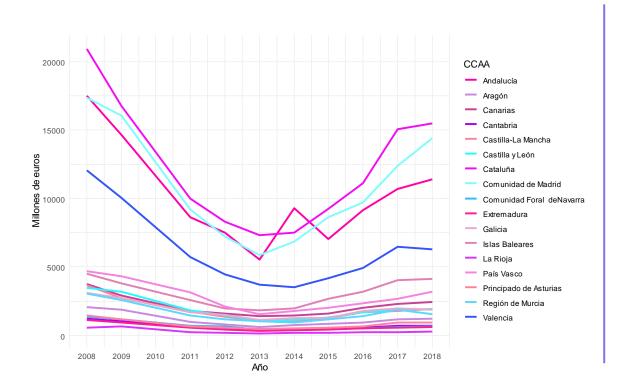




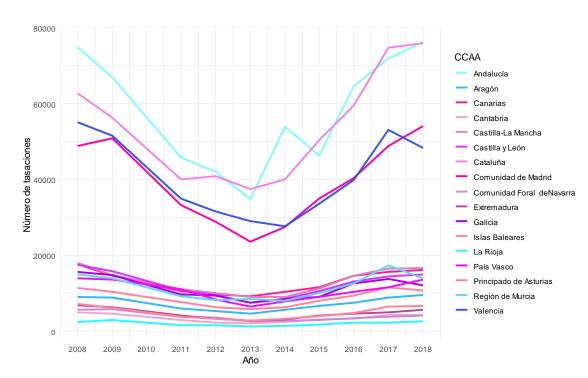


Análisis exploratorio de las tasaciones de vivienda

Importe de tasaciones de viviendas

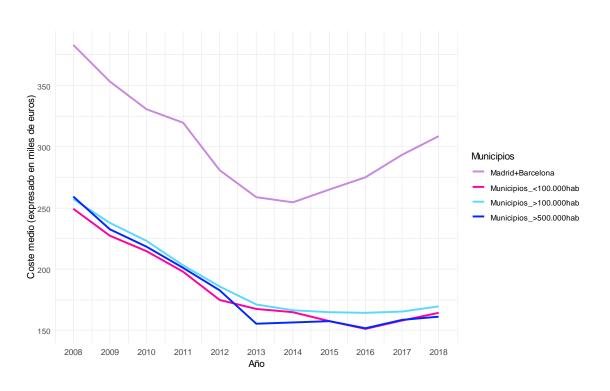


Número de tasaciones de viviendas

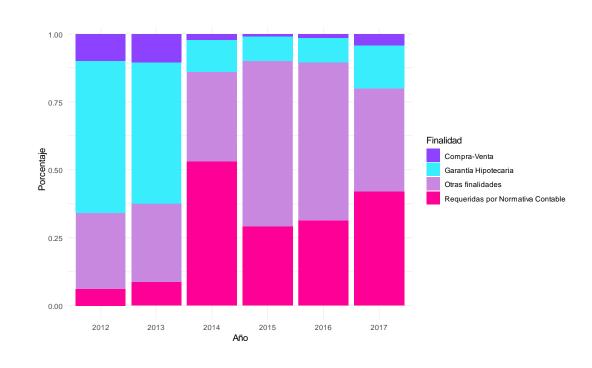


Análisis exploratorio de las tasaciones de vivienda

Estimación del coste medio de viviendas



Distribución de tasaciones de viviendas



Modelo Automatizado de Valoración AVM

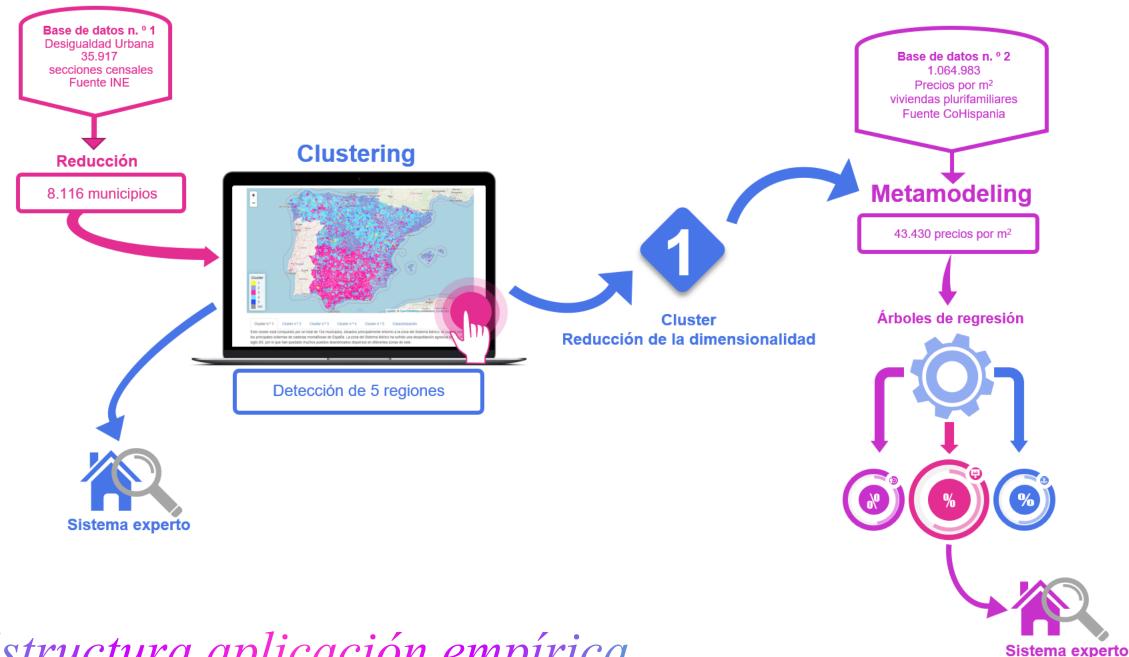


Los modelos automatizados de valoración, son métodos de valoración de inmuebles, mediante el desarrollo de modelos estadísticos, con una base matemática, realizados mediante programas informáticos, y desarrollados a partir de una amplia base de datos.

Permiten realizar valoraciones de forma masiva y a bajo coste, pero asumiendo un determinado grado de error de los resultados



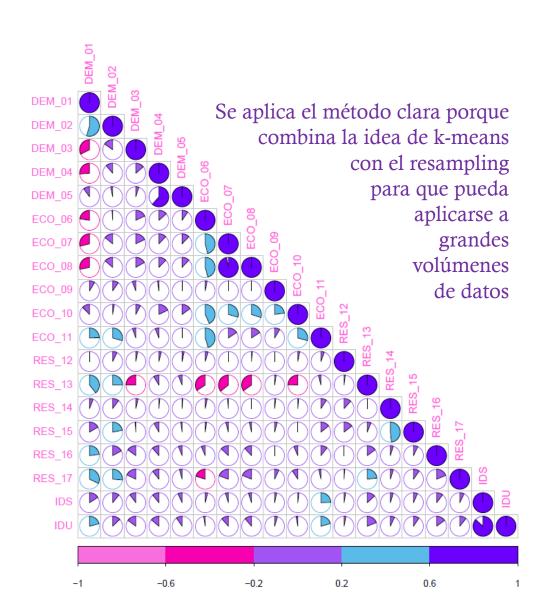




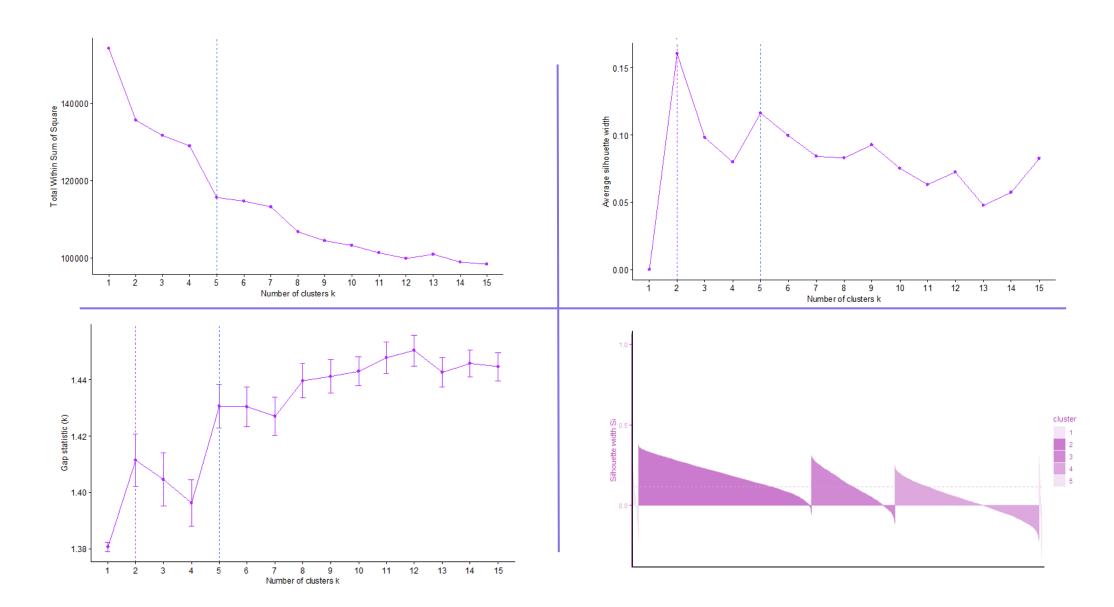
Estructura aplicación empírica

Clustering

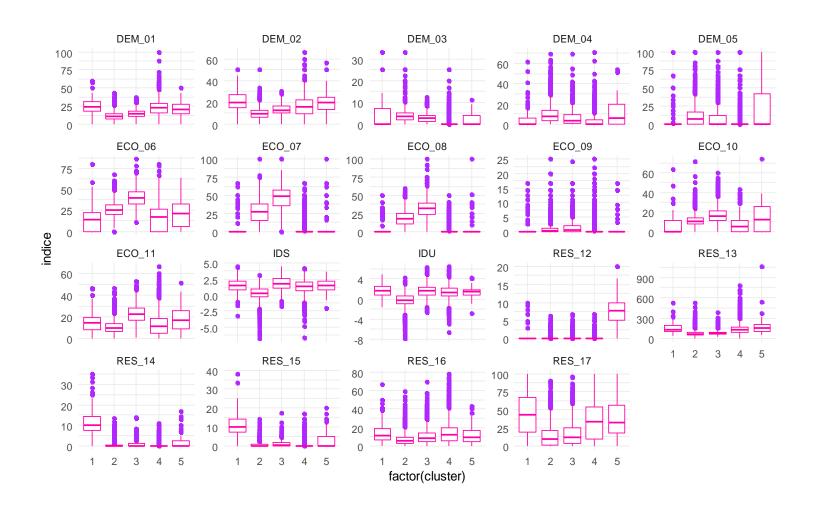


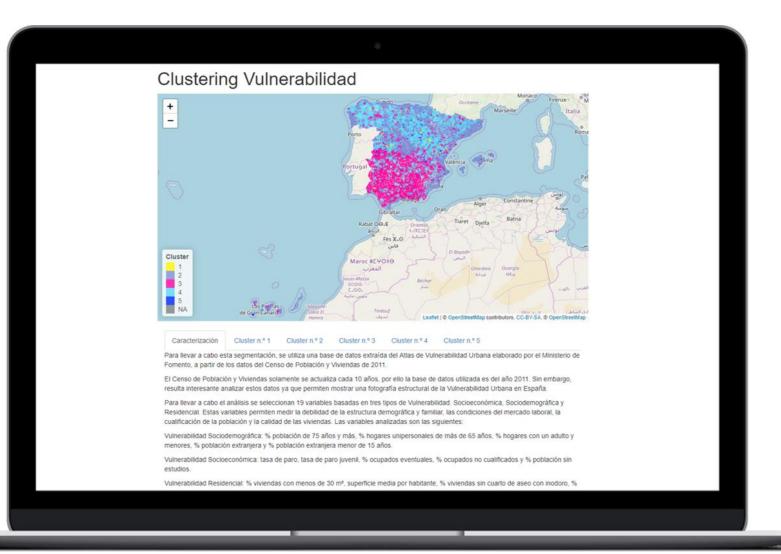


Número óptimo de clusters y validación interna



Caracterización

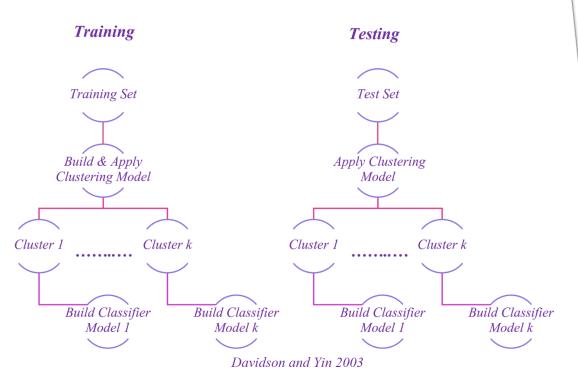


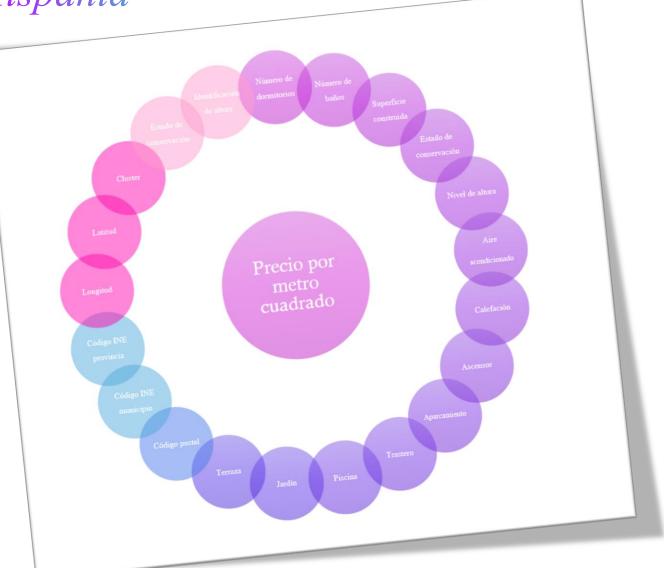


http://shiny.uclm.es/apps/tfg/rocio/

Base de datos aportada por CoHispania

Obtenida a través de web scraping de portales inmobiliarios 1.064.983 observaciones año 2018



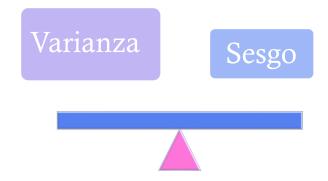


Metodología basada en Machine Learning



Sesgo hace referencia a cuánto se alejan en promedio las predicciones de un modelo respecto a los valores reales

Varianza hace referencia a cuánto varía el modelo dependiendo de la muestra empleada en el entrenamiento

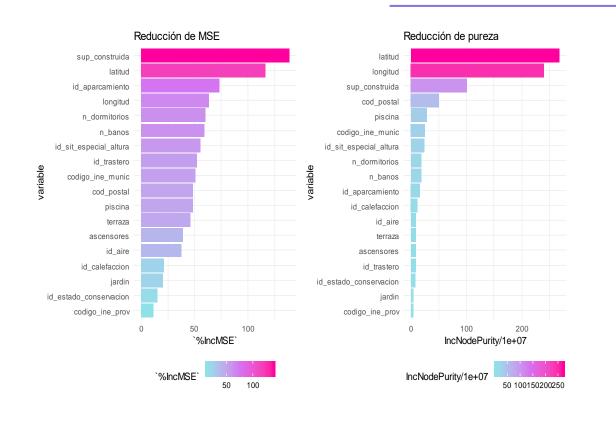


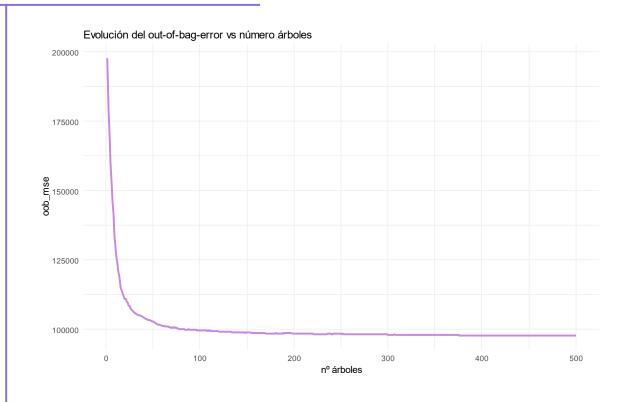
Modelo Bagging

MAPE	RMSE	MedianRE	MedianRatio
23.69	305.15	15.62	1.03

Modelo Random Forest

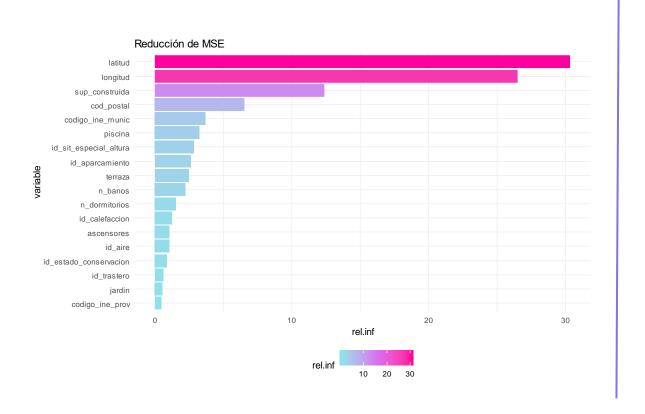
MAPE	RMSE	MedianRE	MedianRatio
23.28	299.27	15.31	1.03

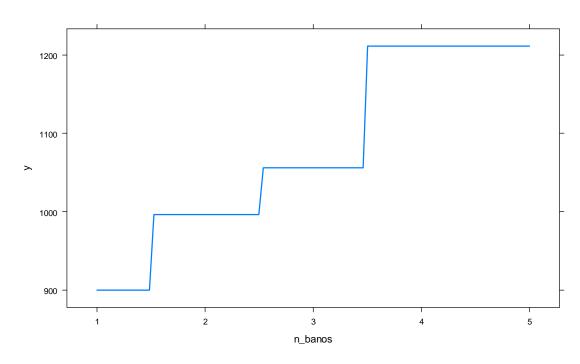




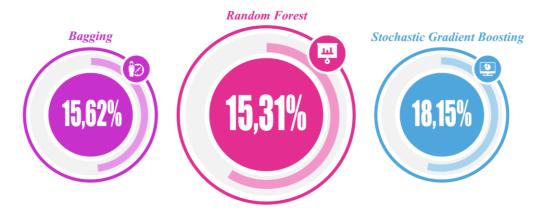
Modelo Stochastic Gradient Boosting

MAPE	RMSE	MedianRE	MedianRatio
26.20	312.71	18.15	1.03

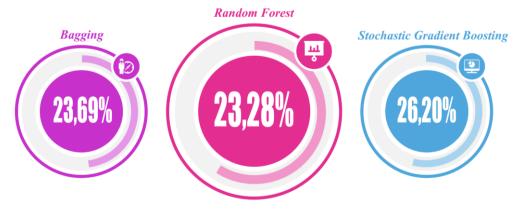




Median Absolute error



MAPE



Conclusiones



Clusters con marcada caracterización



Resultados de Random Forest similares a entornos de producción



Aplicación del algoritmo Random Forest para la creación de un AVM.

- Este algoritmo es capaz de trabajar con grandes volúmenes de datos y con entornos de gran dimensionalidad
- Controla los valores perdidos y es capaz de mantener la precisión cuando existen estos en la base de datos
- Se puede incrementar el número de árboles en el "bosque", tanto como se quiera, sin llegar a generar overfitting
- Capacidad de manejar cientos de variables predictoras sin excluir ninguna

Trabajo Futuro

- Aplicar Random Forest a los 4 clusters restantes
- Análisis de los clusters elaborados aplicados al proyecto en su estado actual
- Elaboración de artículo para revista científica
- ☑ Elaboración de una nueva segmentación con respecto al censo del año 2021

Limitaciones



Extensión de un Trabajo de Fin de Grado



Clustering en base a secciones censales



Base de datos con Outliers



Base de datos con mayor número de variables



Falta de testigos en algunas localidades

