



Curso de Posgrado: Big Data e Inteligencia Territorial

Materia: Machine Learning en Aplicaciones Espaciales

Docente: Dr. Germán Rosati
Diciembre 2020

Consignas del Trabajo Práctico Final

El dataset `ED_deptos_venta_2014.csv` contiene los resultados correspondientes al año 2014 de un relevamiento muestral de los avisos en diarios y sitios online sobre los precios de departamentos en la Ciudad Autónoma de Buenos Aires. El relevamiento se realizaba todos los años desde 2001.

El dataset contiene información acerca de los precios totales, por metro cuadrado en U\$S de los departamentos relevados. El objetivo del trabajo práctico será generar un modelo que permita valorizar los departamentos en venta de la CABA. Para ello deberá comparar las performances de una regresión lineal, de un árbol basado en CART y de un ensamble a elección (Random Forests, Gradient Boosting) y seleccionar el más adecuado al problema en cuestión.

Aspectos a tener en cuenta

Se trata de datos que probablemente tengan cierta estructura espacial. Será importante, entonces, tener en cuenta dichos patrones (tanto en los análisis descriptivos como en la estimación del modelo). Existen diversas variables en el dataset (LAT, LONG, COMUNA, BARRIO, etc.) que deberían poder ayudar en este punto.

Es importante recordar hacer la limpieza de los datos (casos perdidos, variables no relevantes, etc.)

El informe final deberá estar estructurado de la siguiente forma:

- 1) Introducción:** planteo del problema, objetivos
- 2) Breve descripción del dataset y un análisis descriptivo de algunas variables relevantes.**
- 3) Estimación y evaluación de los modelos (regresión lineal, CART y ensamble):** se deberán estimar cada uno de los tres modelos y comparar sus performances

predictivas, con una métrica adecuada. Sobre esta base, deberá seleccionar cuál es el mejor modelo, justificar dicha selección y pasar a realizar una interpretación de sus resultados

4) Interpretación de los resultados obtenidos: dependiendo del modelo seleccionado, deberá utilizar las herramientas más adecuadas de interpretación (coeficientes, variables más importantes, relaciones funcionales, etc.). Recordar que en caso de que el modelo seleccionado sea uno basado en ensamble, se encuentran a disposición las herramientas de “machine learning interpretable”: partial dependence plots, individual conditional expectation, etc.

Pautas de entrega:

Tamaño de hoja: A4 - Márgenes: 3cm. - Letra: Times New Roman - Fuente: 12 - Interlineado: 1,5

Pueden hacerlo en grupos de hasta 3 personas

Fecha de entrega:

29 de Febrero de 2021