

Propuesta

Elizabeth Viveros Vergara y Juan B. Martínez Parente

24 de noviembre de 2018

Propuesta de proyecto

Para el proyecto final de Aprendizaje de Máquina consideraremos dos bases de datos:

- Listado de propiedades en Airbnb para Berlín, París, Ámsterdam y Barcelona (descargadas en Inside Airbnb (<http://insideairbnb.com/get-the-data.html>)) con variables descriptivas de cada propiedad (como número de cuartos, número de baños, ubicación, tamaño, etc.), además de información de los *hosts*, *ratings*, entre otros.
- Información recopilada de fuentes del Coneval (<http://coneval.org.mx>), INEGI (<http://sc.inegi.org.mx/cobdem/contenido.jsp?rf=false&solicitud=>), Conapo (<http://www.conapo.gob.mx/es/CONAPO/Publicaciones>) e INAFED (http://www.inafed.gob.mx/en/inafed/Socioeconomico_Municipal) a nivel municipio provenientes de diferentes encuestas de 2010 y relacionadas con el índice de marginación (rubros de salud, educación, vivienda, seguridad, economía).

Nuestro objetivo es probar varios modelos para predecir, en el primer caso, el precio por una noche en una propiedad de Airbnb con determinadas características, y, en el segundo caso, qué grado de marginación cuenta cada municipio. Esto con el fin de comparar el desempeño de distintas familias de modelos.

Utilizaremos R y Python para programar los siguientes modelos:

- regresión lineal con regularización,
- redes neuronales,
- bosques aleatorios,
- árboles extremadamente aleatorizados (*extremely randomized trees*).