

Modelos y resultados

William Alexander Aguirre A

2024-05-26

Modelos y resultados

Para entrenar los modelos se utilizó la técnica de validación cruzada, de manera que se pudiera reducir el efecto de la autocorrelación espacial sobre el aprendizaje de cada modelo. Teniendo en cuenta que el Marco Geoestadístico Nacional del DANE provee una agrupación geográfica a nivel de manzana, sección (conjuntos de manzanas) y sectores (conjuntos de secciones), se pudo utilizar cada sección como un fold que rompiera con la dependencia espacial del precio.

```
sectores<-unique(train$cod_sector)
folds<-createFolds(sectores,k=length(sectores),list = TRUE,returnTrain = TRUE)
```

Los modelos se entrenaron en primera instancia con la información extraída de la base de datos y del censo del DANE `nbanios`, `nhabitaciones`, `piso_apartamento` ; `estrato` . Dado que los precios de las viviendas también corresponden a diferentes meses de un horizonte de tiempo de tres años, se tuvo que incluir la variable `periodo` para tratar de establecer un efecto fijo del tiempo sobre el precio. Posteriormente se fueron incluyendo las variables de seguridad y salud, encontrando que hubo una mejora significativa en la predicción al incluir este tipo de variables.

Table 1: Resumen de resultados

Algoritmo	Variable objetivo	Predictores utilizados	MAE fuera de muestra
Regresión lineal	price	nbanios, nhabitaciones, piso_apartamento, estrato	860024814
Regresión lineal	log(price)	nbanios, nhabitaciones, piso_apartamento, estrato	295197752
Regresión lineal	log(price)	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo	294830576
Elastic net	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo	283134955
Elastic net	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	268696039
Elastic net	log(price)	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	276791195
Elastic net (caret)	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	264342673
Random forest	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo	274216139
Random forest	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	244889109
Random forest	log(price)	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	252533214
Random forest (caret)	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	249855878
Random forest	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia, distancia_hospital	235325152
Xgboost	price	nbanios, nhabitaciones, piso_apartamento, estrato, Periodo, Robos_vivienda, Robos_personas, distancia_estacion_policia	277450159