**Predicción Precios de viviendas en Chapinero (Bogotá) y el Poblado (Medellín)**

**Introducción**

La industria de la finca raíz es una de las más importantes en la estabilidad y crecimiento económico de los países por la creación de trabajo que supone y por el flujo de dinero en las transacciones. De acuerdo con lo dicho con anterioridad, para las compañías en esta industria es esencial poder realizar predicciones de los precios de las viviendas en busca de generar unas mejores ventas y de no infravalorar estos precios. Unas predicciones acertadas de los precios de las viviendas implican unas mejores ventas y un crecimiento de la industria más acelerado. En este ejercicio se buscó predecir de la mejor manera los precios de viviendas en Chapinero (Bogotá) y el Poblado (Medellín) encontrándose que el modelo Superlearner (glmnet, ranger, Lm y mean) es el mejor modelo de predicción al tener la varianza más baja de todos los modelos construidos.

**Análisis de los datos**

**Modelos y resultados**

Para poder realizar la predicción de los precios de las viviendas el primer procedimiento realizado fue **entrenar nuestro modelo con** **la siguiente forma funcional:**

(***Esta misma forma funcional se utilizó para los 5 modelos realizados en el procedimiento).***

Se decidió realizar **5 diferentes modelos de entrenamiento** para así encontrar el mejor modelo predictivo, estos cinco modelos fueron: **OLS, Ridge, Lasso, Elastic Net y Superlearner**. Los **modelos Ridge y Lasso** se construyeron a partir del **Lambda mínimo óptimo** y el **modelo Superlearner** se decidió construirlo en base a **Random Forest, Elastic Net**.

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelos** | | **MSE** | **Lambda Óptimo** |
| Modelo 1 | Estimación con OLS | 3.014304e+17 | No aplica |
| Modelo 2 | Estimación con Ridge – Alpha=0 | 3.023e+17 | 45016048 |
| Modelo 3 | Estimación con Lasso – Alpha=1 | 3.014e+17 | 1407074 |
| Modelo 4 | Estimación con Elastic Net | 3.011238e+17 | 900321 |
| Modelo 5 | Superlearner (inluye glmnet, ranger, Lm, mean) | 1.612468e+17 | No aplica |

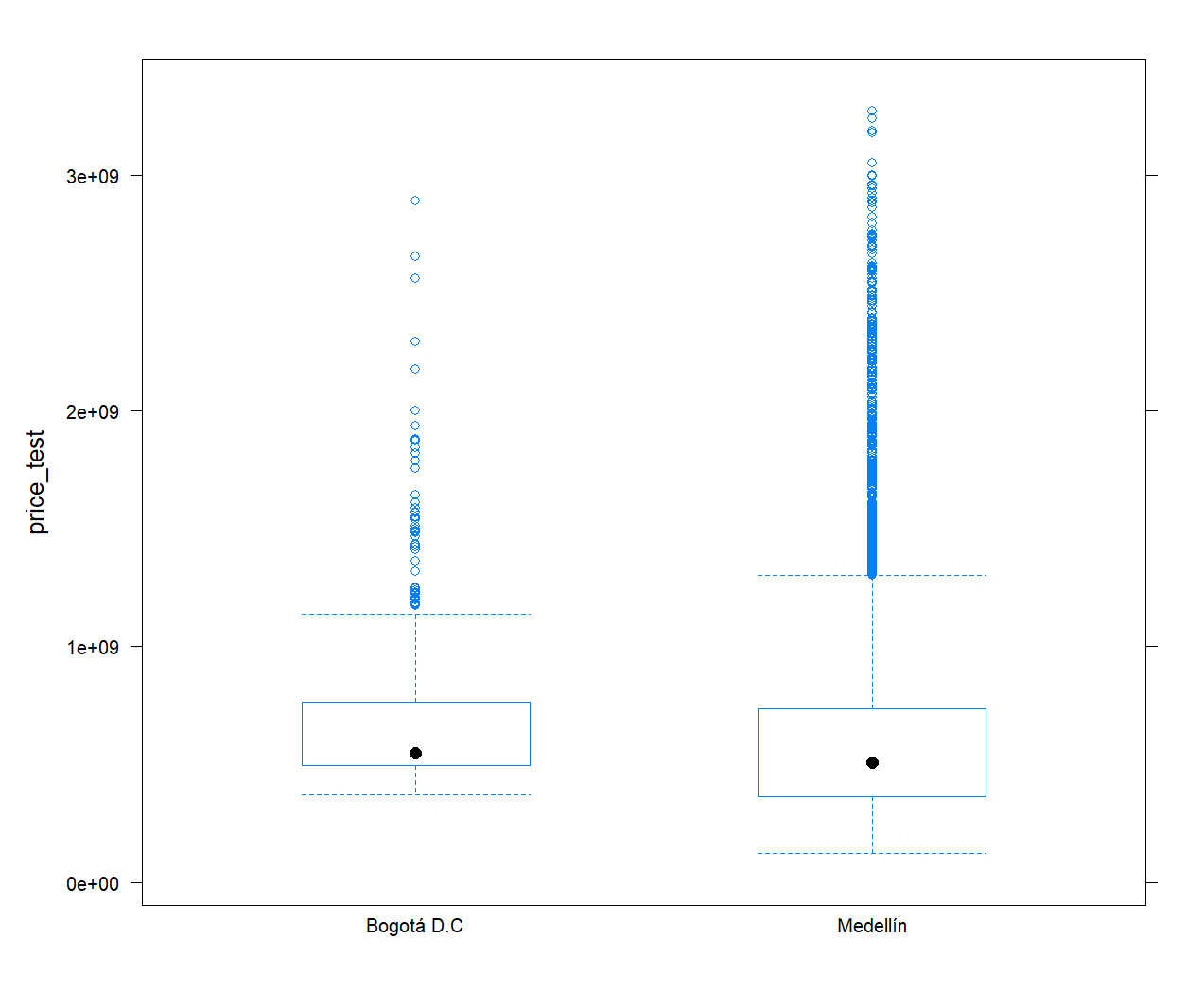
Tabla 1. Modelos utilizados en el problema de clasificación.

Como se observa en la Tabla 1, el mejor modelo de predicción del precio de las viviendas en Chapinero (Bogotá) y el Poblado (Medellín) es a partir del **Modelo 5 – Superlearner (incluye glmnet, ranger, Lm, mean**). Este es el mejor modelo ya que, **es el modelo con menor MSE (1.612468e+17)** lo cual se puede interpretar como el modelo que mejor predice el precio y el que menor variación presenta dentro de los datos (menor varianza).

|  |  |  |
| --- | --- | --- |
| **Superlearner** | **Riesgo** | **Coeficiente** |
| SL.mean | 6.142570e+17 | 0 |
| SL.lm | 3.014159e+17 | 0 |
| SL.ranger | 1.620275e+17 | 1 |
| SL.glmnet | 3.014292e+17 | 0 |

Tabla 2. Modelo Superlearner

En la Tabla 2 podemos apreciar los resultados del modelo Superlearner, encontrando que el modelo ranger (Random Forest) es el modelo que mejor predice los precios de las viviendas con un coeficiente agregado de 1 mientras que los demás modelos (mean, lm y glmnet) obtuvieron un coeficiente de 0, es decir, de poca importancia en el modelo Superlearner. Un análisis valioso al realizar este modelo es que, al ponderar los algoritmos individuales, se encontró que el único algoritmo relevante para esta predicción del precio es el proveniente de Random Forest.



Gráfica 1. Boxplot de las predicciones por ciudad Gráfica 2. Coeficientes del Superlearner

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mínimo** | **1st cuantil** | **Mediana** | **Promedio** | **3rd cuantil** | **Máximo** |
| 1.219e+08 | 3.738e+08 | 5.152e+08 | 6.309e+08 | 7.413e+08 | 3.272e+09 |

Tabla 3. Estadísticas descriptivas predicciones del precio

La Gráfica 1 nos muestra un boxplot de las predicciones del modelo por ciudad analizándose que los precios de las viviendas en Chapinero (Bogotá) y las del Poblado (Medellín) son muy similares pues el precio de la mayoría de estas está por debajo de los 1,000,000,000 COP, pero siendo la mediana de los precios mayor en Chapinero (Bogotá). Ahora bien, este gráfico nos ayuda a determinar que en el Poblado (Medellín) es donde se encuentran las viviendas más baratas y caras.

La Tabla 3 nos muestra las estadísticas descriptivas de las predicciones de los precios de las viviendas pudiéndose determinar de que el precio mínimo predicho es de 121,900,000 COP mientras que el precio máximo predicho es de 3,272,000,000 COP. El promedio de los precios predichos es de 630,900,000 COP lo cual se encuentra acorde con el precio promedio de las viviendas en nuestra base train (763,300,000 COP).

**Conclusiones y recomendaciones**