Eduardo Acosta Hernández A01375206

Para medir el accuracy del modelo se utilizó el coeficiente de R2 el cual varía dependiendo del fragmento del dataset que se utiliza, por lo tanto, se realizaron varias pruebas con diversas variables del dataset, evaluando R2 con los datos de training y test.

Prueba 1

Texto

Descripción generada automáticamente

Prueba 2

Texto

Descripción generada automáticamente

Prueba 3

Texto

Descripción generada automáticamente

Como podemos observar en todos nuestros casos R2­ es mayor en el training y menor en el test, lo cual es normal, sin embargo, podemos observar que en nuestra primera prueba con datos fuertemente correlacionados R2 es mucho mayor que en las pruebas con coeficientes de correlación medios y bajos. Esto nos habla que nuestro modelo es bueno para predecir variables fuertemente correlacionadas y no tan bueno para variables débilmente correlacionadas. Lo anterior lo podemos solventar mediante técnicas de procesamiento de datos para aumentar la calidad de los datos, por ejemplo: eliminación de registros duplicados, valores faltantes, outliers, etc. Con las metodologías anteriores podríamos aumentar la correlación entre las variables y obtener mejores aproximaciones con nuestro modelo.