



Tecnológico de Monterrey

**Momento de Retroalimentación: Módulo 2 Uso de framework o biblioteca de
aprendizaje máquina para la implementación de una solución**

Diego Armando Ulibarri Hernandez

A01636875

ITC

Fecha: 12/09/22

Instituto Tecnológico y de Estudios Superiores de Monterrey

TC3006C.101

Análisis de resultados utilizando LinearRegression de sklearn

Para hacer las predicciones del dataset seleccionado, decidí utilizar regresión lineal con multivariable y entrenar el modelo con la ayuda de sklearn.

Los datos del dataset fueron divididos en una parte train y un test, donde test son valores que el modelo nunca ha visto y con ello y ayuda de la función predict() poder medir el accuracy que tiene el modelo. Para poder medir la precisión de mi modelo utilice el coeficiente de determinación, que da un resultado de 0.68, o bien, un 68% de valores bien estimados (en referencia a y_test), así mismo, muestro los coeficientes y el error cuadrado.

‘El error cuadrático medio (MSE) mide la cantidad de error en los modelos estadísticos. Evalúa la diferencia cuadrática promedio entre los valores observados y predichos. Cuando un modelo no tiene error, el MSE es igual a cero. A medida que aumenta el error del modelo, aumenta su valor. El error cuadrático medio también se conoce como desviación cuadrática media (MSD)’(Frost, 2021)

Finalmente, hago la comparación de la Y_test esperada y la Y_predict para ver que tan exacto es el modelo, podemos observar que si bien no son estimaciones exactas si son cercanas a lo que estamos esperando (en la mayoría de los casos).

```
Coeficiente de determinación: 0.681058055095769
Exactitud del modelo (train): 0.5581197274606959
Coeficientes: [ 5.44074186e+00 -2.70791490e-01 -4.75863892e-03  1.
09142527e+00
 2.29043054e+02 -2.94925908e+01]
Mean squared error: 53.50561912450505
  Actual_Price  Price_Prediction
358           45.1           47.886254
350           42.3           41.164046
373           52.2           44.273014
399           37.3           40.197615
369           22.8           27.513265
..           ...           ...
31            25.0           40.368699
283            23.5           33.091396
157            35.6           42.327834
203            46.1           43.353985
208            26.2           28.439798
[83 rows x 2 columns]
```

Como conclusión:

El score obtenido está por arriba de 0.65, lo que nos indica que el modelo es medianamente aceptable (al menos en este caso que no es muy significativo o perjudicial si da una mala predicción), sin embargo, en este ejemplo los datos no han sido tratados ni analizados de manera detenida y concreta, por lo que podríamos esperar que si limpiáramos el Dataset y aplicáramos técnicas de normalización, nuestro resultado sería más exacto ya que el modelo sería capaz de trabajar mejor con los datos brindados.