



## **Inteligencia Artificial Avanzada para Ciencia de Datos**

**Instituto Tecnológico de Estudios Superiores de Monterrey**

# **Análisis y Reporte sobre el desempeño del modelo**

Rodrigo Montelongo, A00827757

17 de septiembre del 2022

En este análisis se utilizó el modelo de regresión lineal utilizando la librería Sklearn como framework.

Se separaron los datos de nuestro dataset en dos grupos: los de entrenamiento y los de prueba. El modelo se entrenó utilizando los datos de entrenamiento, y posteriormente se calculó el coeficiente de determinación para ambos grupos.

```
Coeficiente de determinación con datos de entrenamiento: 0.38854306120446513
Coeficiente de determinación con datos de prueba: 0.3856938227246892
```

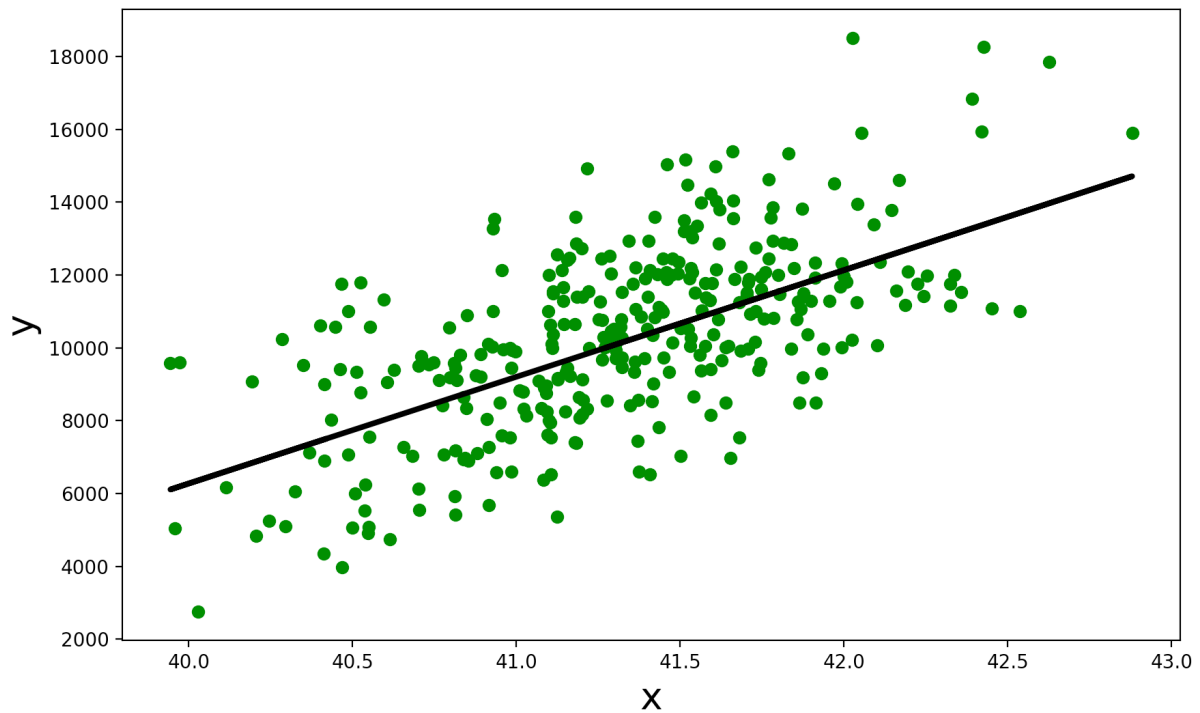
El resultado indica dos cosas. La primera es que los datos de entrenamiento tuvieron un desempeño ligeramente mejor a los datos de prueba. La segunda es que al tener un coeficiente de determinación de 0.38 podemos afirmar que el modelo es muy pobre al sólo explicar el 38 por ciento de la totalidad de los registros.

En cuanto al grado de sesgo, podemos concluir que este es alto debido a que éste se define como la diferencia entre los valores reales y los valores que se predijeron. En este caso podemos ver que el mean squared error en nuestros datos de prueba es bastante alto.

```
Mean squared error datos de prueba: 3881920.469106226
```

Por el contrario, el tener resultados similares entre los datos de entrenamiento y los datos de prueba, podemos asumir que la varianza de nuestro modelo es baja, ya que ésta se define como la variabilidad de las predicciones de nuestro modelo con distintos sets de datos.

Por último, nuestro modelo presenta under-fitting, ya que no se ajusta correctamente a la mayoría de los datos. Esto se puede ver en la siguiente gráfica con los datos reales de prueba comparados con la línea de regresión de nuestro modelo.



Para intentar mejorar el modelo se realizó una regularización de Ridge. Como se puede observar, los coeficientes de determinación tuvieron un pequeño decremento, mientras que el MSE tuvo un ligero incremento. Esto nos indica que en vez de ayudar a nuestro modelo, esta regularización lo empeoró.

Coeficiente de determinación con datos de entrenamiento: 0.38853180182449887  
Coeficiente de determinación con datos de prueba: 0.38551336273595516  
Mean squared error datos de prueba: 3883060.830948636