



# Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

## Reporte sobre el desempeño del modelo

TC3004B.104 Inteligencia Artificial Avanzada para la Ciencia de Datos I

Profesores:

*Ivan Mauricio Amaya Contreras*

*Blanca Rosa Ruiz Hernandez*

*Antonio Carlos Bento*

*Frumencio Olivas Alvarez*

*Hugo Terashima Marín*

*Julian Lawrence Gil Soares - A00832272*

28 de Agosto de 2023

Dataset: Abalone Liga al dataset: <https://archive.ics.uci.edu/dataset/1/abalone>

Este dataset es para un problema de regresión. El objetivo es determinar la edad de los abalone a través del número de anillos que encontramos en ellos.

Modelo: Para mi solución utilizo un simple modelo de regresión lineal, este modelo no cuenta con hiper parámetros que se puedan configurar, así que la manera en la que puedo jugar con los resultados es a través de usar features distintas para la predicción.

Datos:

- Número de registros: 4177
- Número de características: 9
  - Sex
  - Length
  - Diameter
  - Height
  - Whole weight
  - Shucked weight
  - Viscera weight
  - Shell weight
  - Rings
- Clase de salida: Rings

Utilizando los valores de MSE puedo hacer una inferencia acerca del bias de mis datos.

Mis valores fueron:

- 6.98359599679207
- 6.859520586128331
- 8.820565529613267

De estos valores puede observar que hay una cercanía entre 2 de los valores que obtuve del MSE y estos terminaron relativamente bajos así que puedo inferir que mi modelo tiene un bias medio/bajo ya que los datos que estoy prediciendo están saliendo relativamente cerca de los valores reales.

Varianza:

- 2.2855593593389982
- 3.206287632630589
- 2.1742967522023076

El valor de mi varianza implica que tengo un nivel medio de varianza entre mis datos.

Para ver si mi modelo tiene underfitting, overfitting o es fit busque el MSE de mis valores de validación y los compare con el MSE de mis valores de predicción, encontré que hay una gran diferencia entre los dos, con el MSE de mis valores de validación saliéndose significativamente más alta que el MSE de mis predicciones. De estos resultados puedo inferir que mi modelo tiene overfitting así que mi modelo puede que sea muy complejo y está generalizando sobre los datos nuevos que le introduzco a la hora de generar sus predicciones.

Para mejorar el rendimiento de mi modelo decidí utilizar una transformación en base a los valores de z. Calcule el valor z de cada punto en el dataset y en caso de tener un valor z mayor o menor de un número especificado, en mi caso 2 el valor se transforma a la media de mis datos.

Después de mi transformación obtuve las siguientes métricas de desempeño:

MSE:

- 8.430369853853804
- 8.002499093269833
- 8.882133216329688

$R^2$ :

- 0.22122805761968123
- 0.26075345793831617
- 0.17949553136156227

Puedo observar que mi MSE está más estandarizado, así que puedo inferir que mis datos están más normalizados. Lo que sí es que mis valores de  $R^2$  están más cerca de 0 así que mi modelo empeoró en el sentido de cuanta de mi varianza es explicable por el modelo.