



**Tecnológico
de Monterrey**

Momento de Retroalimentación: Módulo 2 Análisis y Reporte sobre el desempeño del modelo

Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Ángela Estefanía Aguilar Medina A01637703

13/09/2025

El modelo que se está evaluando es un modelo de regresión lineal para la predicción de emisiones de CO2 de vehículos utilizando scikit-learn.

1. Separación de datos

El dataset fue dividido en tres conjuntos:

- **Entrenamiento (60%)**
- **Validación (20%)**
- **Prueba (20%)**

Esto permitió evaluar la capacidad del modelo para generalizar sin depender únicamente de los datos de entrenamiento.

2. Resultados de evaluación

Métricas principales:

Conjunto	MSE	MAE	R ²	Varianza Explicada
Entrenamiento	22.25	2.93	0.9938	0.9938
Validación	27.70	3.09	0.9914	0.9914
Prueba	37.27	3.25	0.9897	0.9897

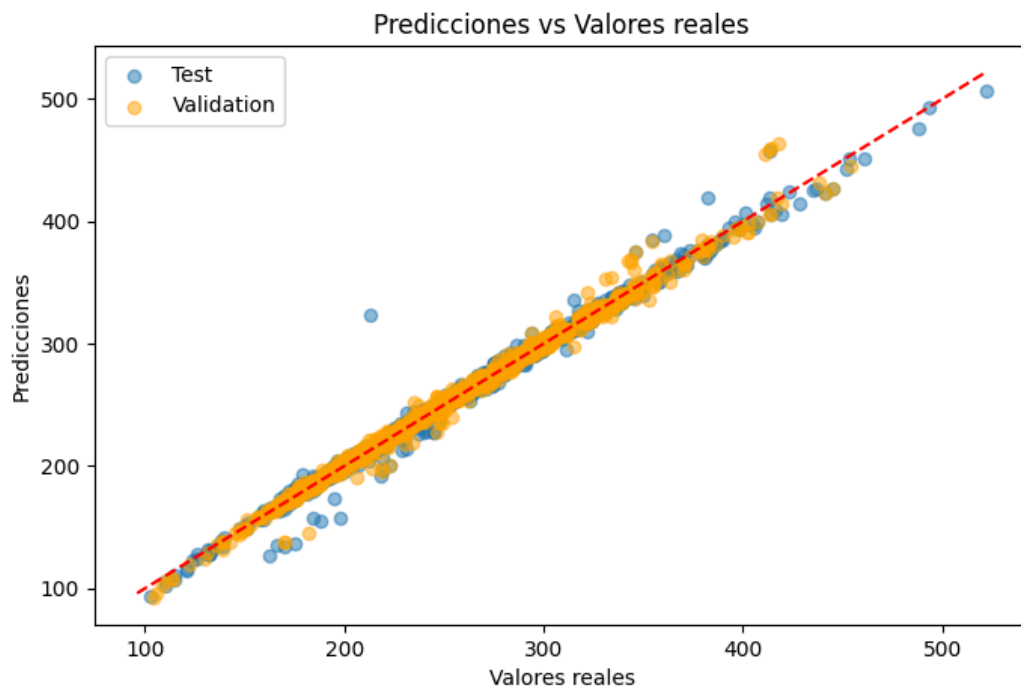
Interpretación:

- El modelo presenta un ajuste muy alto en los tres conjuntos ($R^2 > 0.98$), lo que indica que explica más del 98% de la varianza de las emisiones de CO₂.
- El error absoluto medio (MAE) es de ~3 g/km, lo que representa un nivel de error muy bajo en relación a los valores típicos de emisiones (100–500 g/km).

3. Gráficas de desempeño

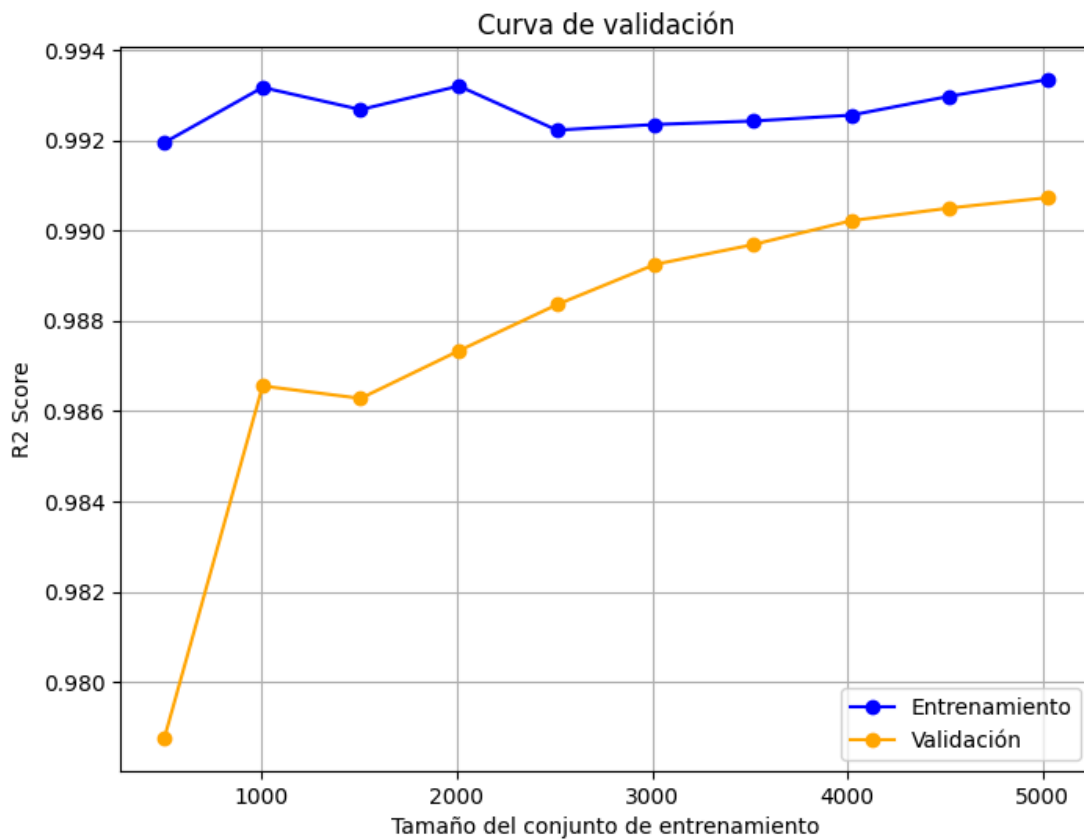
1. Predicciones vs Valores reales:

- Se observa que las predicciones siguen muy de cerca la línea de referencia. Lo que indica que el modelo logra capturar la relación entre las variables de entrada y las emisiones reales



2. Curva de validación (Learning Curve):

- La curva muestra que a medida que aumenta el tamaño del conjunto de entrenamiento, tanto el desempeño en entrenamiento como en validación convergen en valores altos ($R^2 \approx 0.99$).
- Esto sugiere que el modelo se beneficia de más datos, pero ya alcanzó un nivel de estabilidad muy alto.

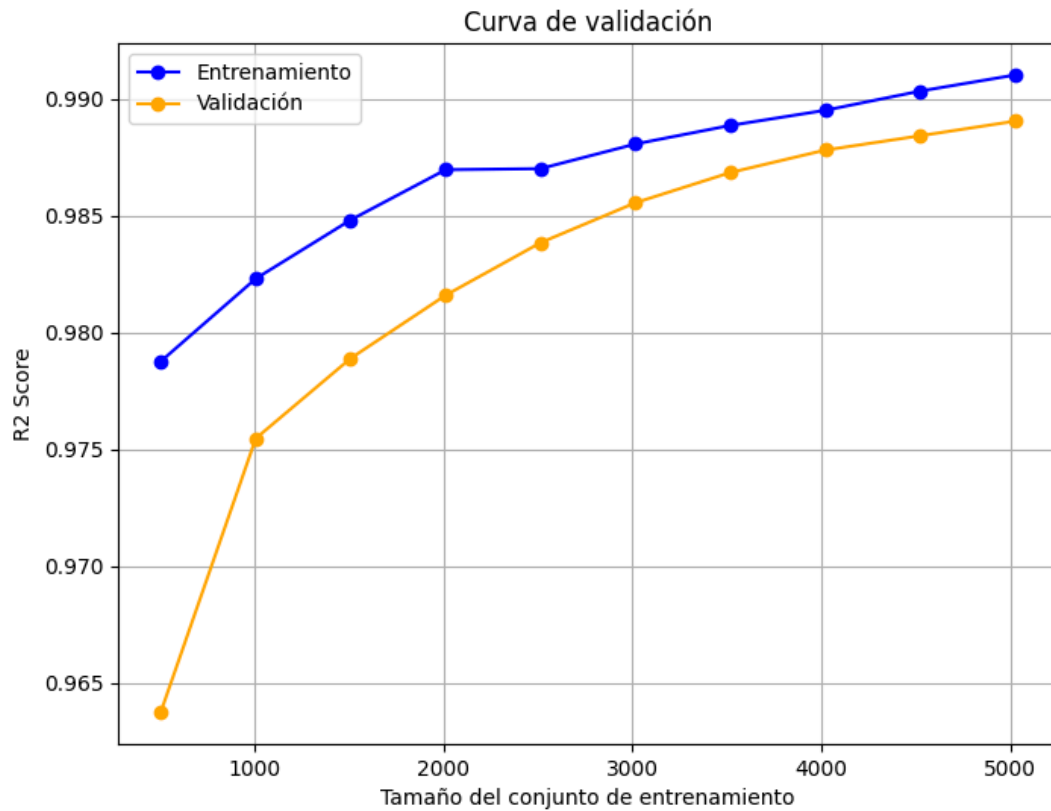


4. Diagnóstico del modelo

- **Bias (sesgo):** Bajo.
El modelo logra capturar correctamente la relación entre variables y salida, con un R² muy alto y errores bajos.
- **Varianza:** Media-Baja.
El desempeño en entrenamiento, validación y prueba es consistente, aunque se observa una ligera caída del R² en test (de 0.9938 → 0.9897), lo que indica mínima varianza.
- **Nivel de ajuste: Buen Fit (ajuste adecuado).**
El modelo no muestra señales claras de underfitting (poco aprendizaje) ni overfitting (memorizar datos), ya que mantiene un desempeño estable en los tres conjuntos.

5. Estrategia de mejora

Aunque el modelo ya tiene un muy buen desempeño, se intentó aplicar regularización Ridge, pero penalizar coeficientes muy grande y evitar el sobreajuste, aunque se pierde algo de precisión en las métricas



Modelo	Dataset	MSE	MAE	R ²	Exp. Var
Linear	Train	22.25	2.93	0.9938	0.9938
	Validation	27.70	3.09	0.9914	0.9914
	Test	37.27	3.25	0.9897	0.9897

Ridge	Train	32.66	3.71	0.9909	0.9909
	Validation	35.61	3.86	0.9889	0.9890
	Test	42.87	4.03	0.9881	0.9881

6. Conclusión

El modelo de Regresión Lineal aplicado al dataset de emisiones de CO₂ logra un desempeño sobresaliente, con un R² alto en todos los conjuntos.

El bajo error (MAE ≈ 3 g/km) lo convierte en una herramienta confiable para predecir emisiones a partir de características técnicas de vehículos.

Con el uso de técnicas de regularización, se espera mejorar aún más la generalización y obtener un modelo más robusto para futuros datos.