

# Act 9: Programando Regresión Lineal en Python

Patricio Ricardí

March 2025

## ¿Qué es la regresión lineal?

Es un método para predecir valores numéricos usando relaciones entre variables. Por ejemplo, predecir cuántas veces se compartirá un artículo basado en sus características.

**Fórmula básica:**

$$\text{Predicción} = \text{Intercepto} + (\text{Coeficiente}_1 \times \text{Variable}_1) + \dots$$

## Pasos que seguimos

1. **Cargamos datos:** Usamos un CSV con información de artículos (palabras, enlaces, comentarios, etc.).
2. **Limpiamos datos:** Eliminamos artículos muy largos (>3000 palabras) o con muchos compartidos (>80,000).
3. **Creamos variable "suma":** Sumamos enlaces + comentarios + imágenes de cada artículo.
4. **Entrenamos modelo:** Usamos Python y scikit-learn para crear un modelo con dos variables: palabras y "suma".
5. **Evaluamos resultados:** Medimos qué tan preciso fue el modelo.

## Resultados clave

- **Coeficientes:**
  - +6.63 por palabra adicional (más palabras = más compartidos).
  - -483.41 por cada elemento interactivo (más enlaces/comentarios/imágenes = menos compartidos).
- **Precisión:**
  - Error promedio: 352,122,816.48 (diferencia vs realidad)

- $R^2 = 0.11$  (el modelo explica solo el 11% de lo que ocurre).
- **Ejemplo:** Un artículo de 2000 palabras y 20 elementos interactivos tendría:
 
$$6.63 \times 2000 - 483.41 \times 20 = 3,591.8 \text{ compartidos}$$

## Conclusiones

- El modelo es muy básico ( $R^2 = 0.11$ ) pero sirve como introducción.
- Para mejorarlo podríamos:
  - Usar más variables (ej: tema del artículo).
  - Probar métodos más avanzados (ej: redes neuronales).
  - Eliminar variables que no aportan información útil.
- ¡Es un buen primer paso para entender cómo funcionan las predicciones con Python!