

# Taller 5 AI: Modelos de Regresión (*Challenge*)

Facultad de Ingeniería  
Departamento de Electrónica

**Nota:** fecha máxima de envío de los entregables (ver punto 1.c.) **sábado 23 de abril de 2022 a las 11:59 a.m.** al correo [warenas@javeriana.edu.co](mailto:warenas@javeriana.edu.co) con el asunto “Taller 5 AI\_32700”. Por cada minuto de retraso en la entrega se descontará una (1) décima.

## Objetivos:

- Utilizar modelos de regresión para la estimación de datos.
- Diseñar una estrategia que garantice la correcta selección del modelo de regresión, teniendo como referencia el posible *underfitting/overfitting* sobre los datos de entrenamiento.

1. El archivo “*data.npy*” es un diccionario, el cual contiene dos matrices: *training\_set* y *testing\_set*. El conjunto de entrenamiento, “*training\_set*”, está compuesto por 500 observaciones del número de *Scooters* eléctricos alquilados en una determinada ciudad. Cada dato está representado por diez atributos, definidos de la siguiente manera:

Tabla 1: Atributos de una observación.

1	2	3	4	5	6	7	8	9	10
Instant	Season	Year	Month	Holiday	Weekday	Temperature	Casual*	Registered*	Count*

\*El conteo total de alquileres diarios corresponde a “Count”, mientras que “Casual” y “Registered” indican si los alquileres corresponden a usuarios casuales o registrados, respectivamente.

## Challenge:

- Para un día de fiesta ( $Holiday = 1$ ), realice una regresión polinomial de “Casual” en función de “Temperature”. Es decir,  $x = \text{“Temperature”}$  y  $y = \text{“Casual”}$ . Diseñe una estrategia para determinar el orden del polinomio que realiza la mejor estimación.
- Con base en el mejor modelo obtenido según su criterio, realice las predicciones de  $y$  sobre el conjunto de prueba “*testing\_set*”.

\*Observación: recuerde que la predicción, al ser un conteo de alquileres, deberá ser un número entero.

- Para facilitar la evaluación del *Challenge*, envíe:

- Descripción breve de la estrategia empleada (no más de 3 líneas).

- Orden del polinomio.
- Los parámetros del modelo,  $\beta$ 's, con al menos cinco (5) cifras significativas.
- Las predicciones obtenidas en el punto b.