



Universidad Autónoma de Nuevo León

Facultad de ciencias fisico matematicas

**Maestría en ciencia de datos**

# Aprendizaje Automatizado

## Práctica 2 Regresión

---

Catedrático: Jose Anastacio Hernandez Saldaña

Equipo : Cynthia Selene Martinez Espinoza

Matrícula : 1011238

San nicolás de los Garza, Nuevo León, a 24 de Julio del 2024

# Datos

## Exploración de datos

**Datos tomados de:** [Repositorio Datos ML UCI](#)

En la actualidad, en muchas ciudades se utilizan bicicletas de alquiler para mejorar la comodidad de la movilidad. Es importante que la bicicleta de alquiler esté disponible y sea accesible para el público en el momento adecuado, ya que esto reduce el tiempo de espera.

Finalmente, proporcionar a la ciudad un suministro estable de bicicletas de alquiler se convierte en una preocupación importante. La parte crucial es la predicción del número de bicicletas necesarias a cada hora para el suministro estable de bicicletas de alquiler.

El conjunto de datos contiene información meteorológica (temperatura, humedad, velocidad del viento, visibilidad, punto de rocío, radiación solar, nevadas, precipitaciones), el número de bicicletas alquiladas por hora e información sobre la fecha.

## Exploración de datos

	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)
0	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0
1	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0
2	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0
3	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0
4	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0

## Mejor Modelo con Validación cruzada

Se aplica la validación cruzada encontrando el mejor modelo de árbol de decisión.

```
DecisionTreeClassifier
DecisionTreeClassifier()
```

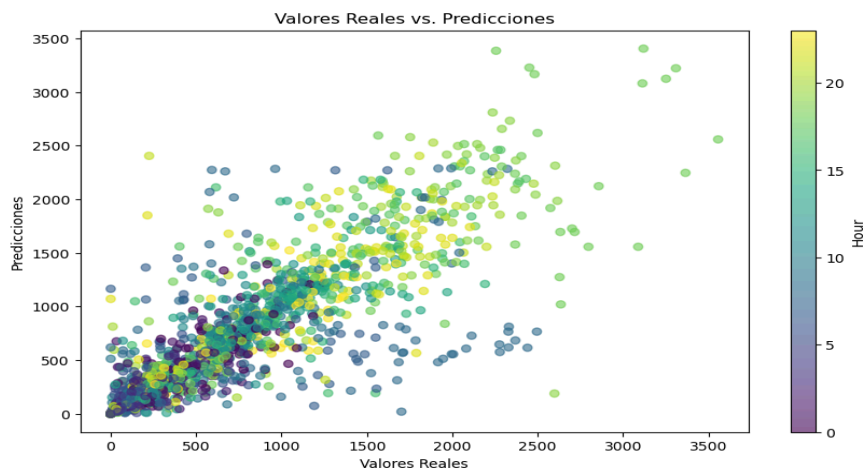
## Entrenar el mejor modelo de Árbol de decisión

Entrenar el mejor modelo nos dio los resultados de:

```
Resultados de la Validación Cruzada (R^2): [0.73164226 0.73095454 0.7676254 0.76933193 0.74267066]  
R^2 promedio: 0.7484  
R^2 en el conjunto de prueba: 0.6985
```

Mean Squared Error: 125606.54623287672

R<sup>2</sup> Score: 0.6985295080936884



## Análisis de resultados

Error Cuadrático Medio (MSE): Un MSE de 125606.54 indica que, en promedio, las predicciones del modelo tienen un error cuadrático bastante alto. Esto sugiere que las predicciones no son muy precisas.

Coefficiente de Determinación ( $R^2$ ): Un  $R^2$  de 0.698 Este valor indica que el modelo de árbol de decisión explica aproximadamente el 69.8% de la variabilidad en la renta de bicicletas. Esto sugiere que el modelo tiene un rendimiento razonablemente bueno, pero hay espacio para mejoras. La hora del día, la temperatura, la humedad y otros factores meteorológicos (como la lluvia y la nieve) son determinantes importantes en la renta de bicicletas. El modelo identifica estos factores como los más influyentes. Impacto de las condiciones meteorológicas: Las condiciones meteorológicas adversas, como altas temperaturas, lluvia y nieve, tienden a reducir la demanda de bicicletas.