



Universidad Autónoma de Nuevo León

Facultad de ciencias fisico matematicas

**Maestría en ciencia de datos**

# Aprendizaje Automatizado

## Práctica 3 Clasificación

---

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:

Evaluación en el conjunto de prueba:

Accuracy: 0.9144

Precision: 0.9183

Recall: 0.9077

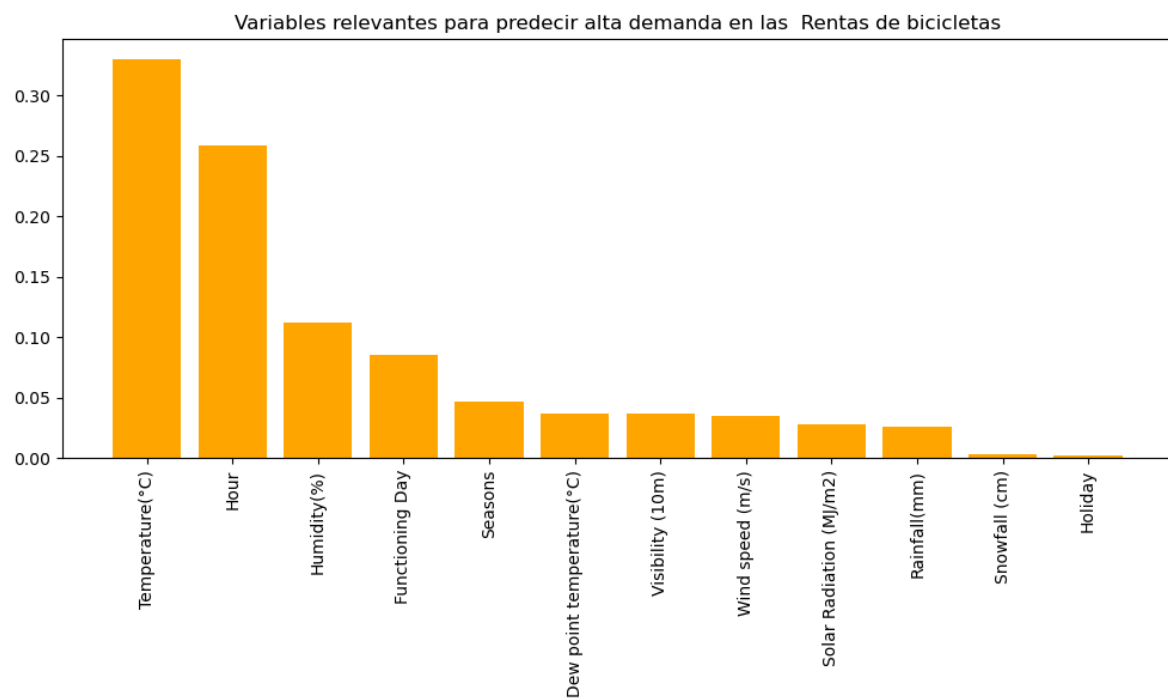
F1 Score: 0.9130

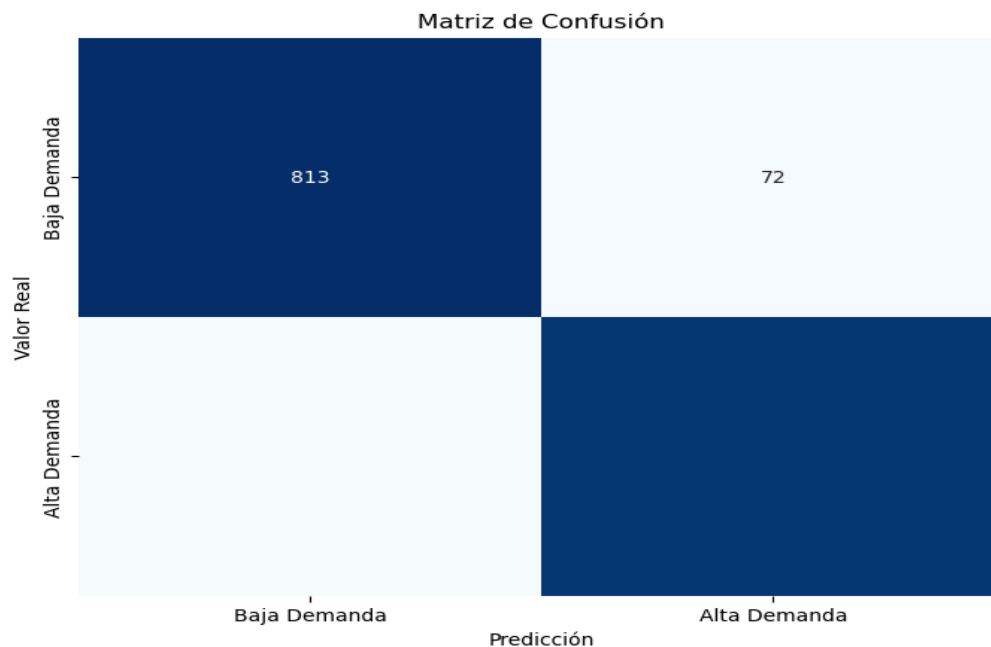
ROC AUC: 0.9143

---

Mean Squared Error: 0.08618721461187215

R<sup>2</sup> Score: 0.6552147479131234





## Análisis de resultados

El modelo de árbol de decisión ha mostrado ser el mejor modelo de clasificación según la métrica de accuracy tanto en la validación cruzada como en el conjunto de prueba. Un accuracy de 0.9144 indica que el modelo predice correctamente si habrá alta demanda de bicicletas en aproximadamente el 91.44% de los casos

La precisión de 0.9183 y el recall de 0.9077 sugieren que el modelo es muy bueno tanto para identificar correctamente los períodos de alta demanda (recall) como para asegurar que las predicciones de alta demanda son correctas (precisión).

Las variables como la hora del día, la temperatura, la humedad, y las condiciones meteorológicas tienen una influencia significativa en la demanda de bicicletas.

Con un accuracy superior al 90%, y métricas de precisión, recall, F1 y ROC AUC que reflejan un rendimiento robusto, el modelo proporciona una base sólida para la toma de decisiones operativas y estratégicas.