

# Informe de Pruebas de Modelos de Regresión para Predicción de Demoras en Transporte Masivo

## Introducción

Este informe detalla los resultados de las pruebas realizadas a un componente de software diseñado para predecir el tiempo de demora en sistemas de transporte público. Se emplearon dos modelos de aprendizaje automático: regresión lineal y árbol de decisión, y se evaluaron sus desempeños utilizando métricas estándar.

## Objetivo

El objetivo principal fue evaluar la precisión y eficacia de los modelos de regresión lineal y árbol de decisión en la predicción de demoras en el transporte público, considerando diversas variables como hora, origen, destino, tipo de transporte y condiciones climáticas.

## Conjunto de Datos

Se utilizó un conjunto de datos que incluía información sobre un total de 1999 viajes, con las siguientes características:

- **Hora:** Hora del día en formato 24 horas.
- **Origen:** Estación de origen.
- **Destino:** Estación de destino.
- **Transporte:** Medio de transporte (bus, metro, etc.).
- **Clima:** Condiciones climáticas (soleado, lluvioso, etc.).
- **Demora:** Tiempo de demora en minutos (variable objetivo).

## Metodología

1. **Preprocesamiento:**
  - Codificación de variables categóricas (origen, destino, transporte, clima) utilizando LabelEncoder.
2. **División de datos:**
  - División del conjunto de datos en entrenamiento (80%) y prueba (20%) utilizando train\_test\_split.
3. **Entrenamiento de modelos:**
  - **Regresión lineal:** Entrenamiento del modelo utilizando las características mencionadas.
  - **Árbol de decisión:** Entrenamiento del modelo con los mismos datos.
4. **Evaluación:**
  - Cálculo de las métricas MSE (Error Cuadrático Medio) y  $R^2$  (Coeficiente de

Determinación) para ambos modelos.

## Resultados

Modelo	MSE	R <sup>2</sup>
Regresión Lineal	8.522	-0.0022
Árbol de Decisión	19.881	-1.338

- **MSE:** Mide el promedio de los errores al cuadrado entre los valores reales y los predichos. Valores más bajos indican mejor ajuste.
- **R<sup>2</sup>:** Indica la proporción de la varianza de la variable objetivo explicada por el modelo. Valores más cercanos a 1 indican mejor ajuste.

## Conclusiones

- **Regresión Lineal:** Aunque tiene un MSE relativamente bajo, su R al cuadrado casi nulo nos indica que no está capturando bien las relaciones en los datos. En pocas palabras, no podemos confiar en que este modelo esté haciendo un buen trabajo.
- **Árbol de Decisión:** Aquí la situación es más clara: un MSE alto y un R al cuadrado tan negativo nos muestran que este modelo no está funcionando bien en absoluto. No solo no mejora respecto al promedio, sino que lo empeora.