

Supuestos de regresión después del modelado

Informe resumido ejecutivo para la Comisión de Taxis y Limusinas de la Ciudad de

Nueva York preparado por **Automatidata**

ISSUE / PROBLEM

La Comisión de Taxis y Limusinas de la Ciudad de Nueva York contrató a Automatidata para predecir las tarifas de los taxis. En esta parte del proyecto, el equipo de datos de Automatidata creó el entregable para la solicitud original de su cliente: un modelo de regresión.

RESPONSE

El equipo de datos de Automatidata optó por crear un modelo de regresión lineal múltiple (MLR) basado en el tipo y la distribución de los datos proporcionados. El modelo MLR mostró un modelo exitoso que estima las tarifas de los taxis antes del viaje.

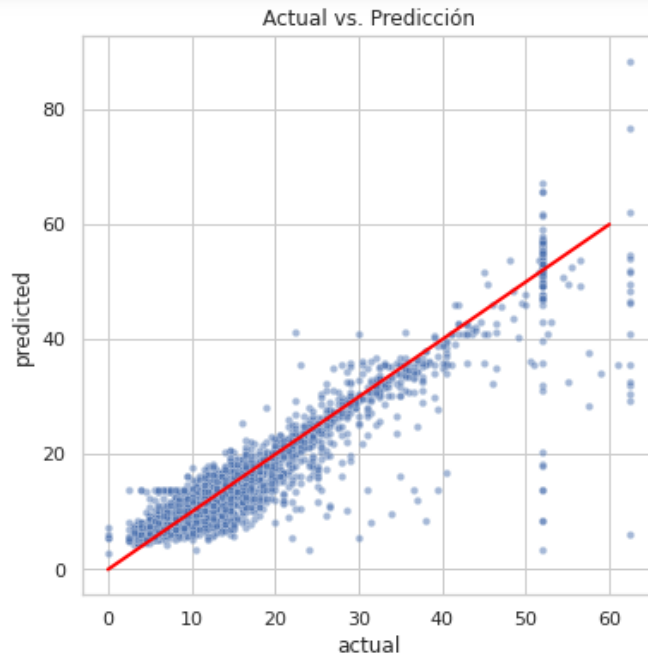
El rendimiento del modelo es alto tanto en los conjuntos de entrenamiento como en los de prueba, lo que sugiere que el modelo no está demasiado sesgado y que el modelo no está sobreajustado. El modelo funcionó mejor con los datos de prueba.

IMPACT

La imputación de valores atípicos optimizó el modelo, específicamente en lo que respecta a las variables de: monto de la tarifa y duración.

El modelo de regresión lineal proporciona un marco sólido para predecir el importe estimado de la tarifa de los viajes en taxi.

Con el fin de mostrar la eficacia del modelo de regresión lineal, el equipo de datos de Automatidata incluyó un diagrama de dispersión que compara el importe de la tarifa prevista y la real. Este modelo se puede utilizar para predecir la cantidad de tarifas de los viajes en taxi con una confianza razonable. El cuaderno suministrado muestra un análisis más detallado de los residuos del modelo.



Texto alternativo: El diagrama de dispersión muestra un gráfico de modelo de regresión lineal que ilustra el importe de la tarifa prevista y real para los viajes en taxi.

Mettricas del modelo:

- Ajuste neto del modelo
 - ✓ R^2 0,87, lo que significa que el 86,8% de la varianza está descrita por el modelo.
 - ✓ MAE 2.1
 - ✓ MSE: 14.36
 - ✓ RMSE 3.8

KEY INSIGHTS

- La característica con el mayor efecto en el monto de la tarifa fue la duración del viaje, lo cual no fue inesperado. El modelo reveló un aumento medio de 7 dólares por cada minuto adicional, sin embargo, este no es un punto de referencia fiable debido a la alta correlación entre algunas características.
- Solicite datos adicionales de itinerarios subrepresentados.
- La Comisión de Taxis y Limusinas de la Ciudad de Nueva York puede usar estos hallazgos para crear una aplicación que permita a los usuarios (pasajeros de TLC) ver la tarifa estimada antes de que comience su viaje.
- El modelo proporciona una predicción de tarifas generalmente sólida y confiable que se puede utilizar en los esfuerzos de modelado posteriores.