**Informe de Rendimiento del Modelo de Machine Learning**

Muy bien en el presente informe vamos a abarcar lo que sería la metodología aplicada para la evaluación, el rendimiento que presente y las conclusiones finales.

**Entrenamiento y Validación**

El modelo fue entrenado utilizando un conjunto de datos de especificaciones de vehículos y precios históricos. Dividimos el conjunto de datos en partes de entrenamiento y validación para asegurarnos de que el modelo generalice bien a nuevos datos.

* Datos de Entrenamiento: 80% del conjunto de datos
* Datos de Validación: 20% del conjunto de datos

**Métricas de Evaluación**

Las métricas utilizadas para evaluar el rendimiento del modelo son las siguientes:

* Error Absoluto Medio (MAE): Mide el error promedio entre los valores reales y los valores predichos por el modelo. Esta métrica es útil para comprender cuánto se desvían las predicciones del precio real en promedio.
* Coeficiente de Determinación (R²): Indica qué porcentaje de la varianza en los datos de salida puede ser explicado por el modelo. Un valor de R² cercano a 1 significa que el modelo predice con alta precisión.

**Resultados de la Evaluación**

* MAE: 12,456.58
* R² Score: 0.0312

**Conclusiones de la evaluación:**

Analizando los datos de la evaluación podemos concluir que el rendimiento del modelo está por debajo de lo esperado, en este escenario, lo primero que debemos hacer es analizar los datos que estamos utilizando en la etapa de entrenamiento y evaluación.

Si vamos al EDA, podemos observar que tenemos valores poco realistas, tenemos precios de vehículos que van desde 1 a casi 26 millones, presentando similitudes en la características, por ejemplo la marca.

Ahora bien si deseamos aumentar el rendimiento del modelo se recomienda lo siguiente:

1. Consultar con el cliente o data engineer los casos de usos atípicos en la data, por ejemplo los autos cuyo precio sería de un 1, o el que va a 26 millones.
2. Si validamos la data y continua el bajo rendimiento, recomiendo segmentar la data, y aplicar modelos en función a esa segmentación, es decir, podemos tener módulos para vehículos accesibles(baratos) y lujosos(caros).
3. Se puede analizar una fluctuación de modelos (Árboles de decisión), sin embargo, creo que es una apuesta bastante arriesgada, ya que para el caso de uso, una regresión lineal, se adapta bastante bien, y como comente en los anteriores ítems, veo como opción más factible mejorar la calidad de la data.